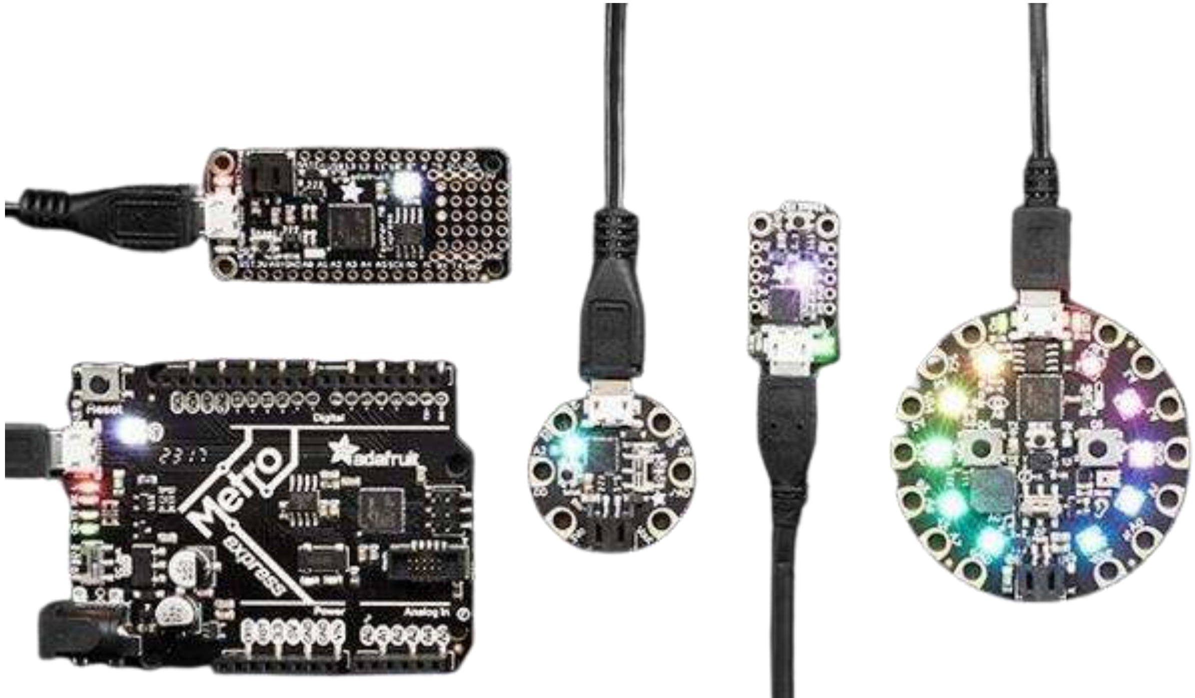
SETS

Society for Electronic Transactions and Security

Artificial Intelligence and Hardware Security



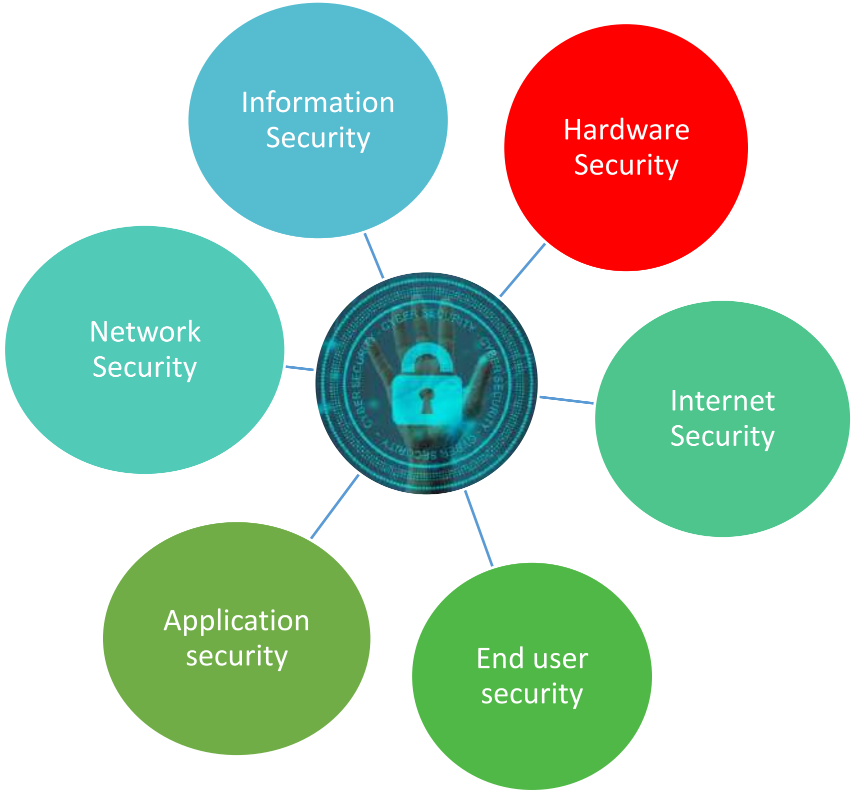
Eswari Devi N
Hardware Security Research Group
SETS, Chennai

Outline

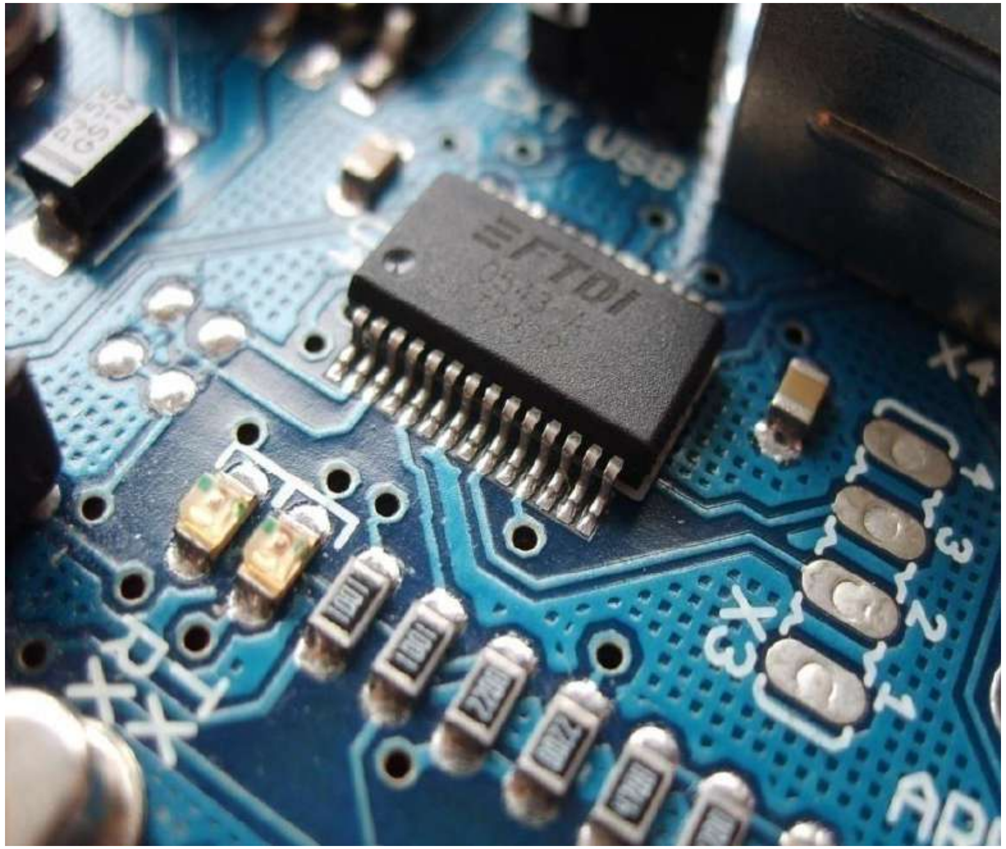
- Hardware Security
- AI for Hardware Security
  - Side Channel Analysis
  - Hardware Trojans
  - Reverse Engineering
  - AI in secure physical design flow
- AI in IoT security
- AI in Digital Twin for Hardware Security
- AI against Hardware Security
  - Poisoning attacks
  - Side Channel Analysis
  - Physical Unclonable Functions

Cybersecurity and AI

- Protecting the Hardware, Software and Data from cyber threats
- AI in cybersecurity can automatically detect the presence of threat and can defend even without the involvement of humans
- AI acts as a Powerful weapon against cyberthreats
- Its applications include classification algorithms used for
  - Zero day attacks
  - Identifying and prioritizing attacks
  - Malware
  - Spam detection
  - Anomaly detection algorithms
  - Detect malicious traffic or user behaviors



Why Hardware Security ?

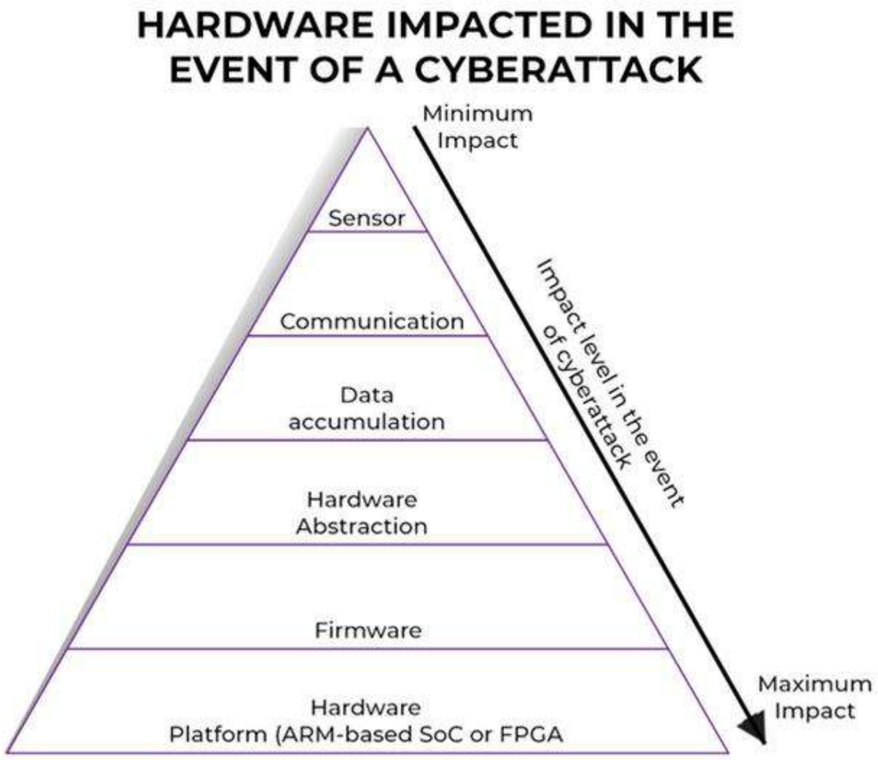


- Provides platform for software execution and support for implementing cryptography
- The encryption algorithms are implemented in hardware – cryptographic co-processors (speed and security)
- Secure system = Secure + Trusted hardware support
- Root of Trust – more trusted than software

Source: https://www.toolbox.com/it-security/vulnerability-management/articles/what-is-hardware-security/#_001

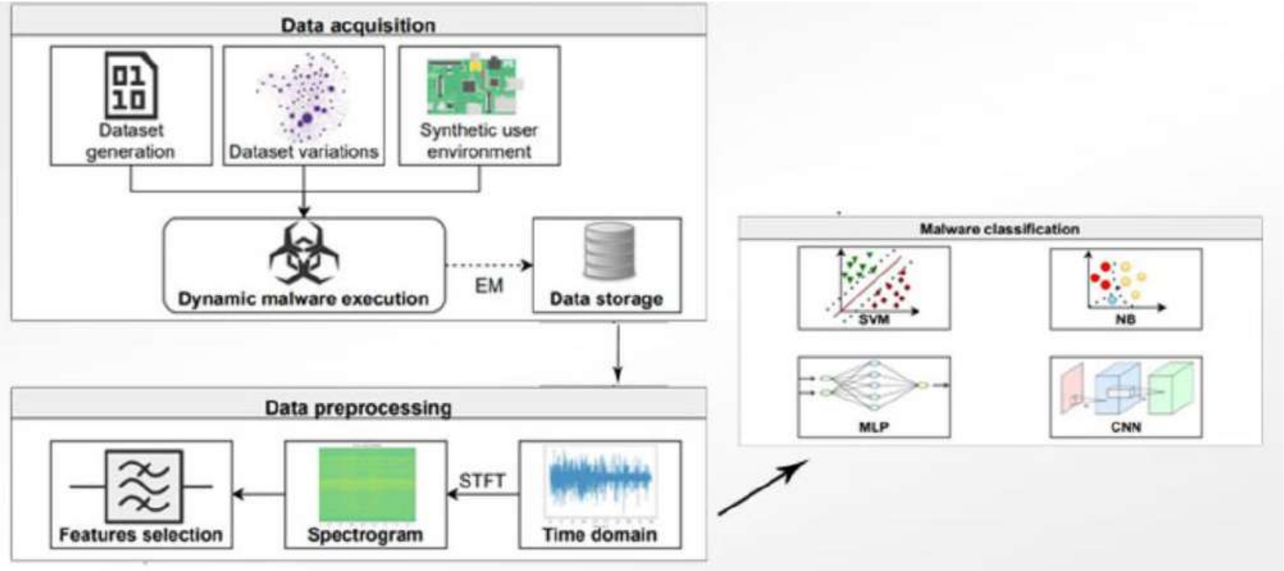
Hardware Security

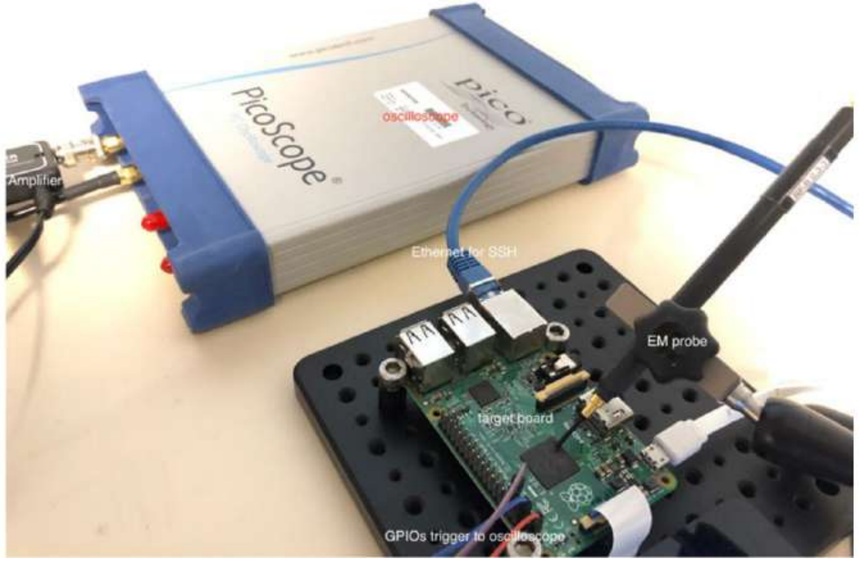
- Supply chain
  - a malicious foundry may insert Hardware Trojans into fabricated chips
  - The delivered IP cores may contain malicious logic and/or design flaws which could be exploited by attackers after the IP cores are integrated into SoC platforms
- Side Channel Attacks
- Fault Injection
- Probing attacks
- Reverse Engineering


HARDWARE IMPACTED IN THE EVENT OF A CYBERATTACK
Minimum Impact
Sensor
Communication
Data accumulation
Hardware Abstraction
Firmware
Hardware Platform (ARM-based SoC or FPGA
Impact level in the event of cyberattack
Maximum Impact

Protecting the physical devices from threats that would facilitate unauthorized access to the data

Detect evasive Malware on IoT Devices using Electromagnetic emanations

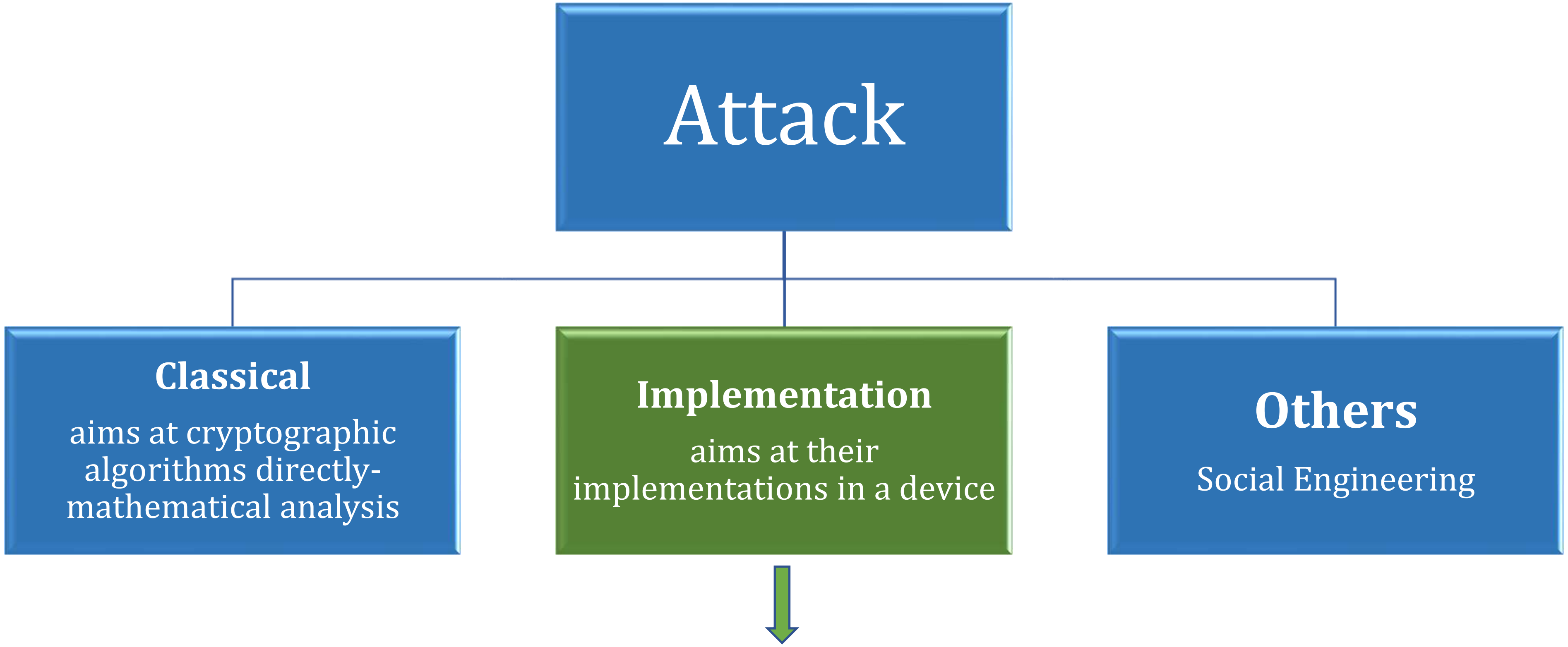




- The findings were presented by a group of academics from the Research Institute of Computer Science and Random Systems (IRISA)
- Electromagnetic field emanations from the Internet of Things (IoT) devices as a side-channel to glean precise knowledge about the different kinds of malware targeting the embedded systems

Source: https://thehackernews.com/2022/01/detecting-evasive-malware-on-iot.html

Cryptanalysis



Attack

Classical

aims at cryptographic algorithms directly- mathematical analysis

Implementation

aims at their implementations in a device

Others

Social Engineering

Implementation Attack – Do not aim at the weakness of the algorithm, but on its implementation

“The action or process of observing something in order to gain information"

Cryptanalysis

Classical attack

Known/Chosen plain text attack

- Known/Chosen plain text and corresponding cipher text

Cipher text only attack

- set of cipher texts

Adaptive plain text attack

- can request the ciphertexts of additional plaintexts

Related key attack

- some mathematical relationship connecting the keys is known to the attacker

Implementation attack

Side Channel Attack

- emit information during operation through side channel

Fault analysis

- aims to induce faults in a cryptographic circuit so that it behaves abnormally and/or delivers incorrect results, which in turn reveals information about the secret key

Probing attacks

- wires between logic cells,etc. are electrically contacted to read out their state while the cryptographic circuit operates normally

Reverse engineering

- the software or hardware implementation of an unknown cryptographic algorithm is analyzed to find out how it actually works

Other attack

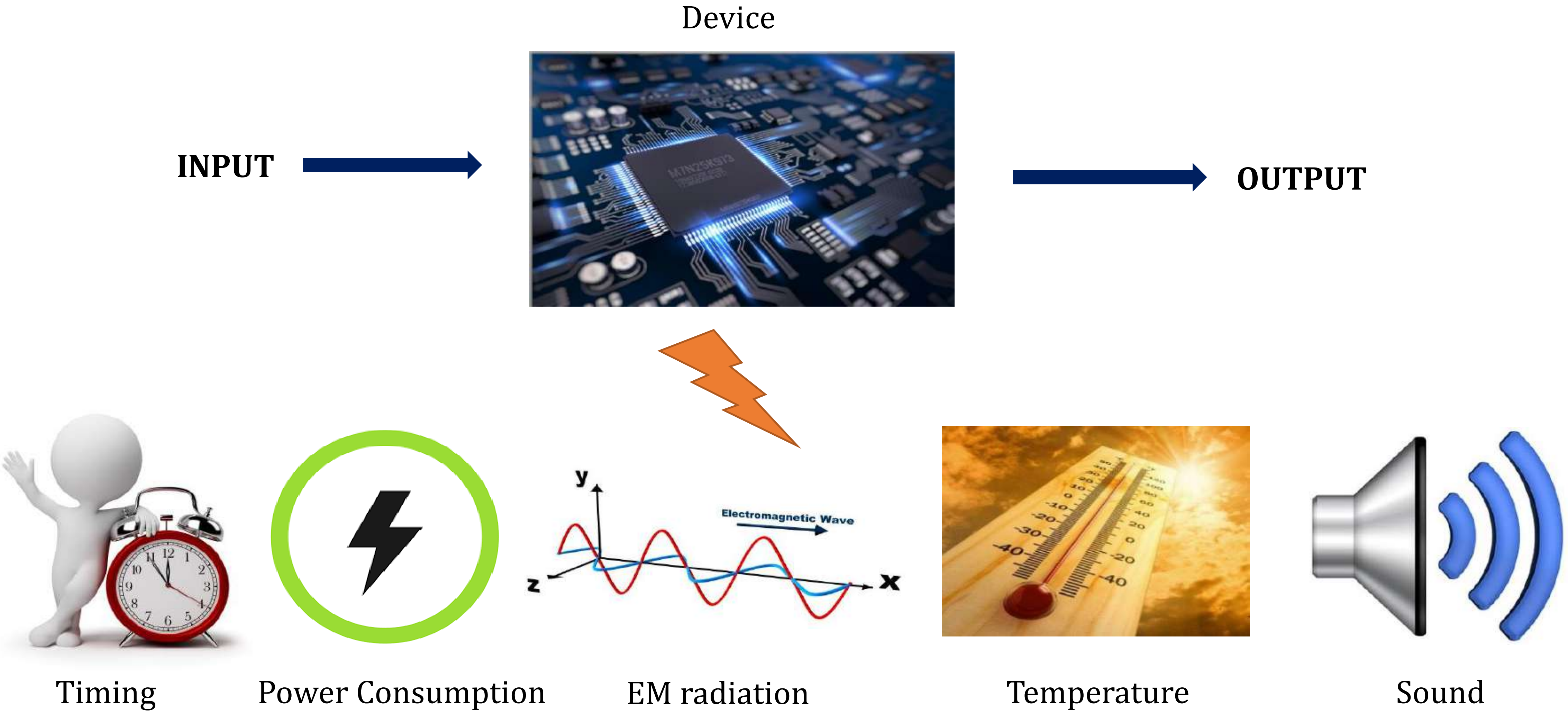
Brute Force Attack

- guess all the possible logical combinations

Dictionary Attack

- uses a wordlist in order to find a match of either the plaintext or key

Side Channel Information



Device

INPUT

OUTPUT

Timing

Power Consumption

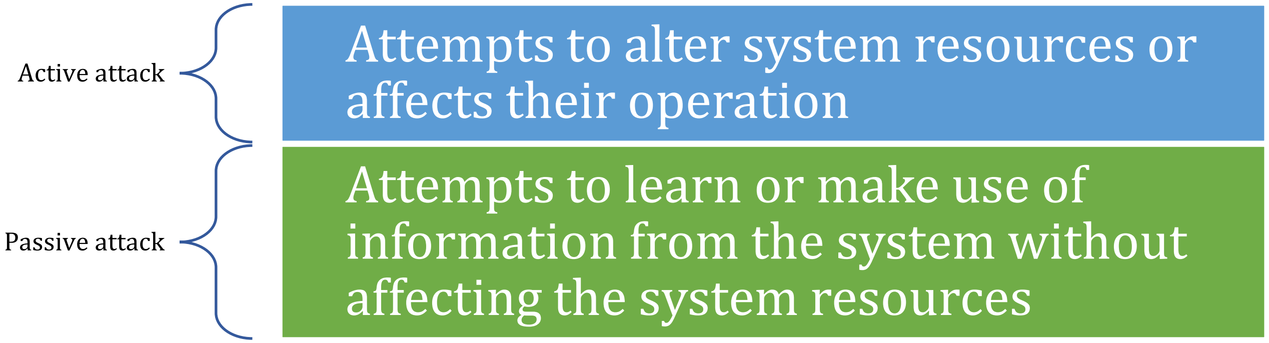
EM radiation

Temperature

Sound

Side Channel Attacks

- 1995 Paul Kocher
- To exploit the weakness in physical implementation of the cryptosystem
- Side channel leakages are gathered to analyze the information used in the algorithm implemented in hardware
- Most powerful attack type

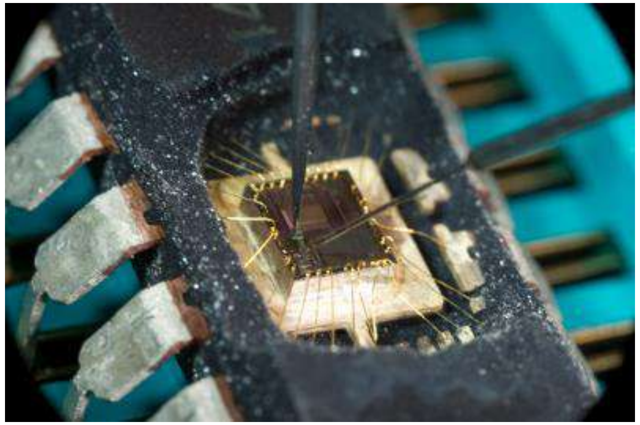


Active attack

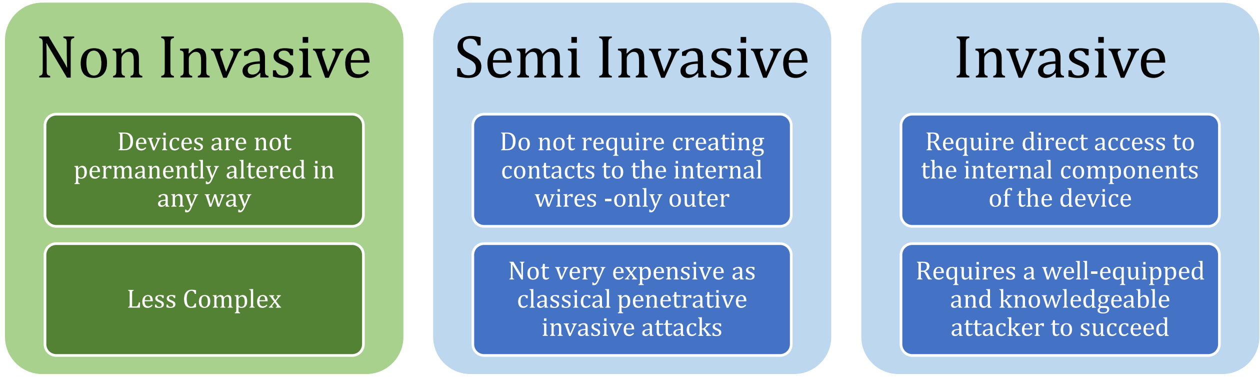
Attempts to alter system resources or affects their operation

Passive attack

Attempts to learn or make use of information from the system without affecting the system resources



Tampering techniques



Non Invasive

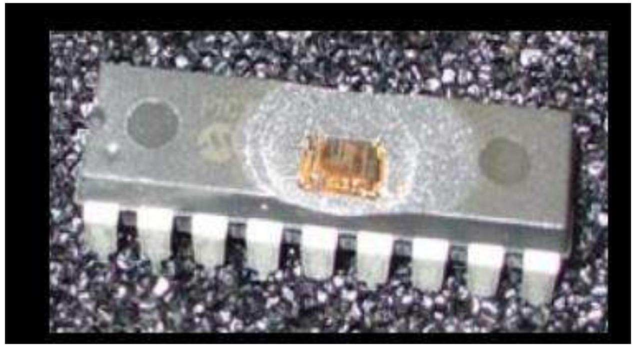
- Devices are not permanently altered in any way
- Less Complex

Semi Invasive

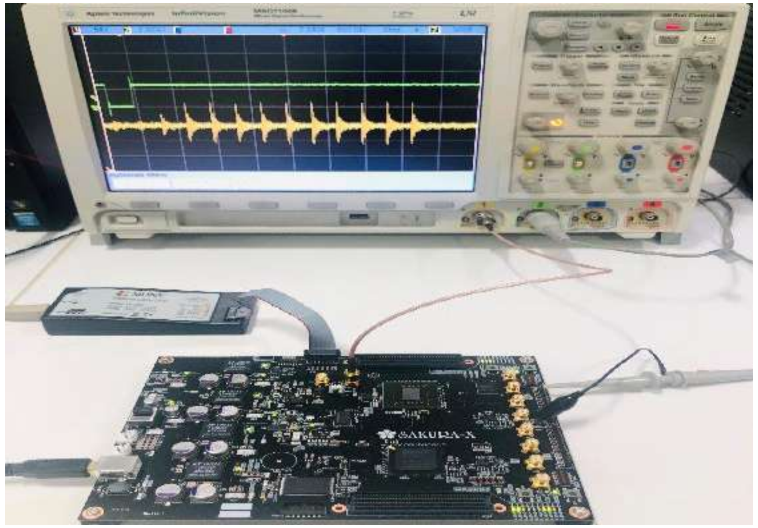
- Do not require creating contacts to the internal wires -only outer
- Not very expensive as classical penetrative invasive attacks

Invasive

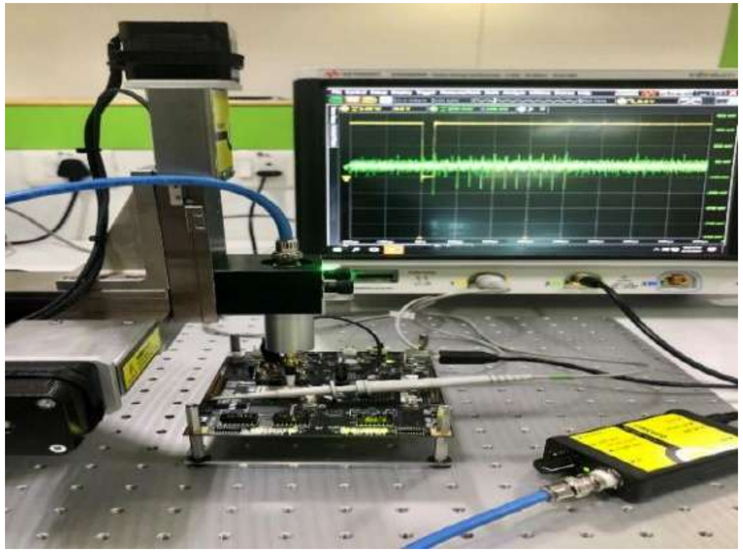
- Require direct access to the internal components of the device
- Requires a well-equipped and knowledgeable attacker to succeed



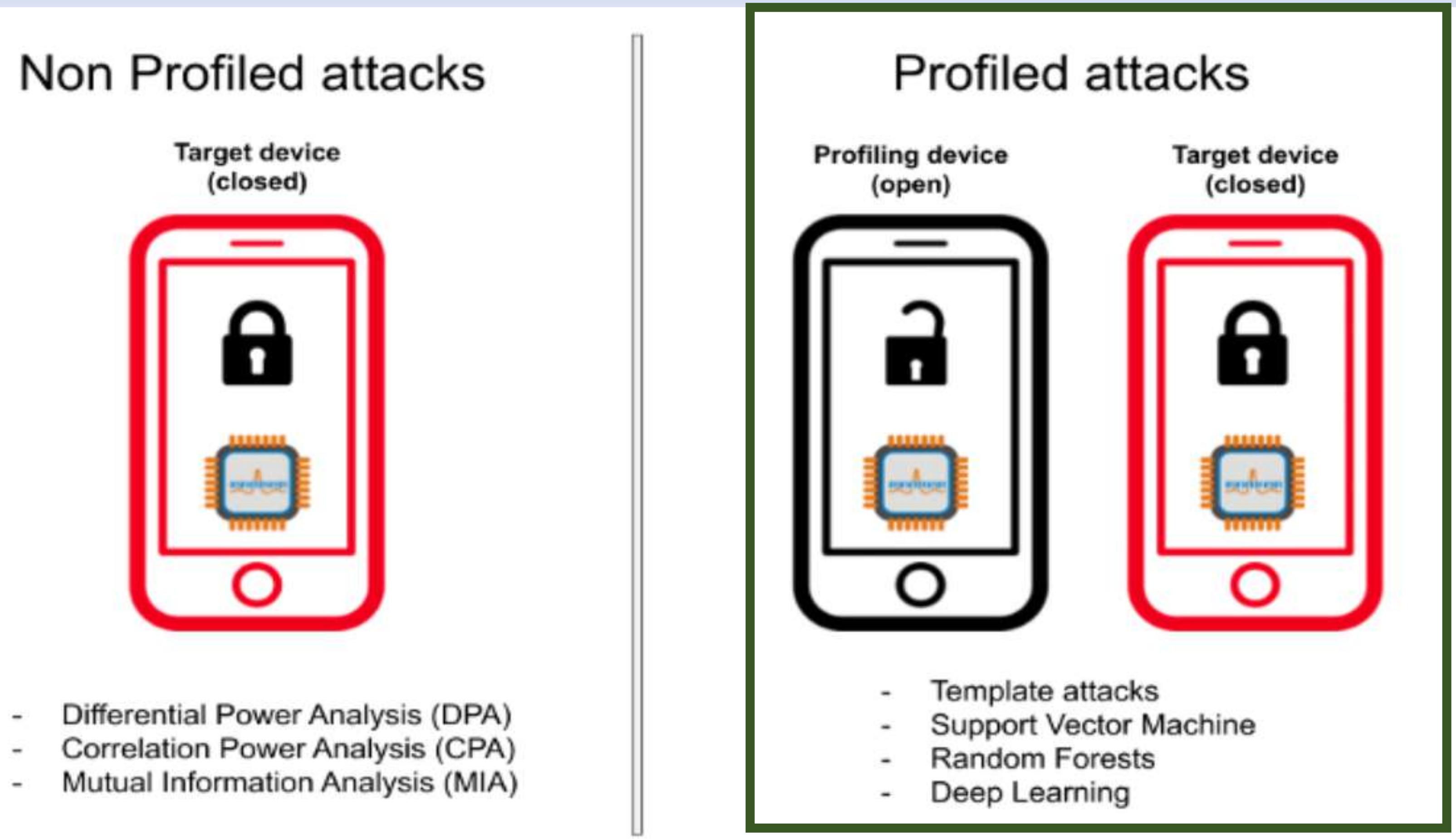
Power measurement setup at SETS



EM measurement setup at SETS



Template Attacks



Non Profiled attacks

- Differential Power Analysis (DPA)
- Correlation Power Analysis (CPA)
- Mutual Information Analysis (MIA)

Profiled attacks

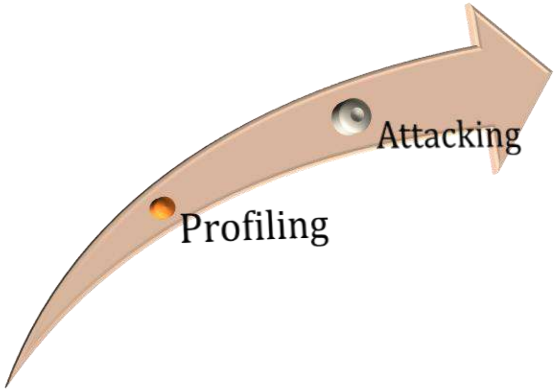
- Template attacks
- Support Vector Machine
- Random Forests
- Deep Learning

Closed Target; data cannot be changed

Open copy; any data can be changed

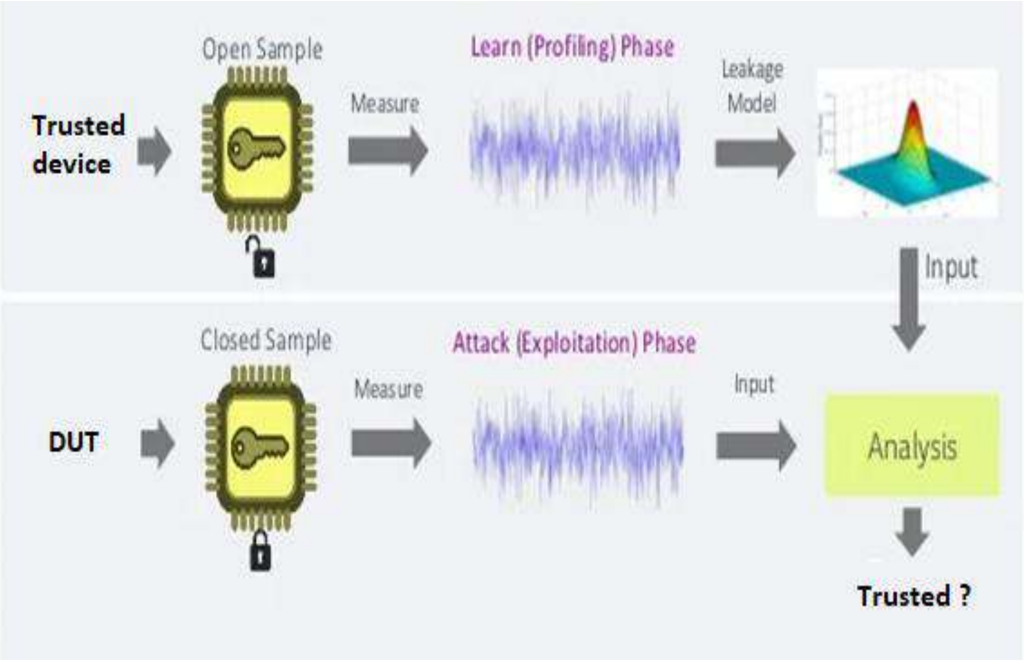
DL based Template Attacks

- The most powerful attack from the information theoretic point of view
- An attacker creates a "profile" of a sensitive device and applies this profile to quickly find a victim is trusted or not
- Raise an alert if there is a suspicious behavior



1. Record a large number of power/EM traces using many different inputs
2. Train the model
3. Record a small number of traces on target device
4. Apply the attack traces to the trained model

DL based Template Attacks

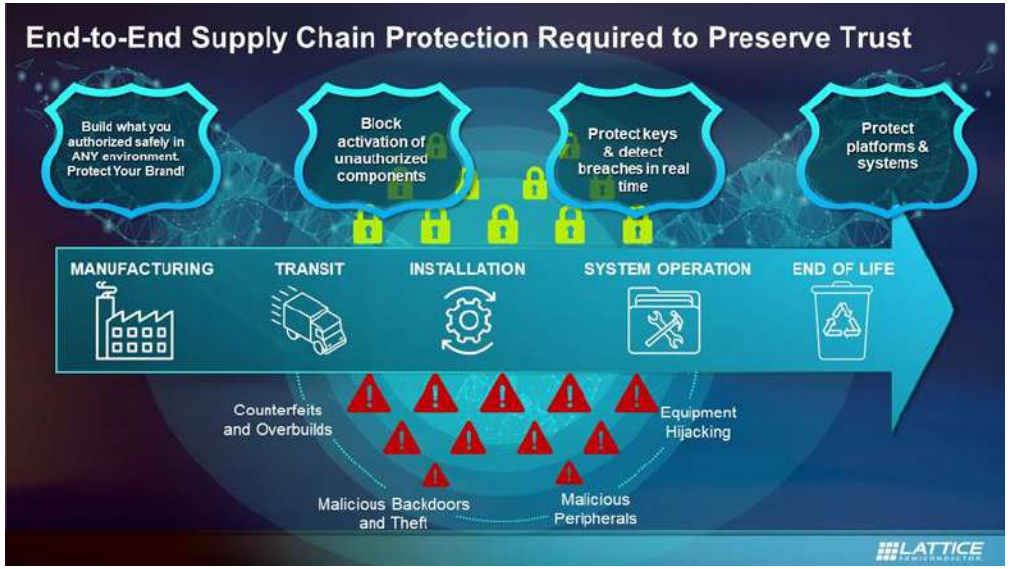


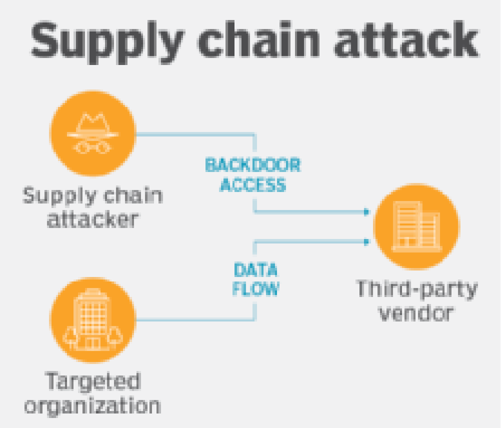
- The goal is to take advantage of the side channel information to detect anomalies in emanations when they deviate from previously observed patterns and raise an alert when suspicious behavior emulating the malware is recorded in comparison to the system's normal state
- Template Attack becomes difficult to apply when measurement dimension exceeds 100
- Retains the spatial property and the functions
- CNN outperforms Multi Layer Perceptron (MLP) in the presence of desynchronization/jittering (translation problem)

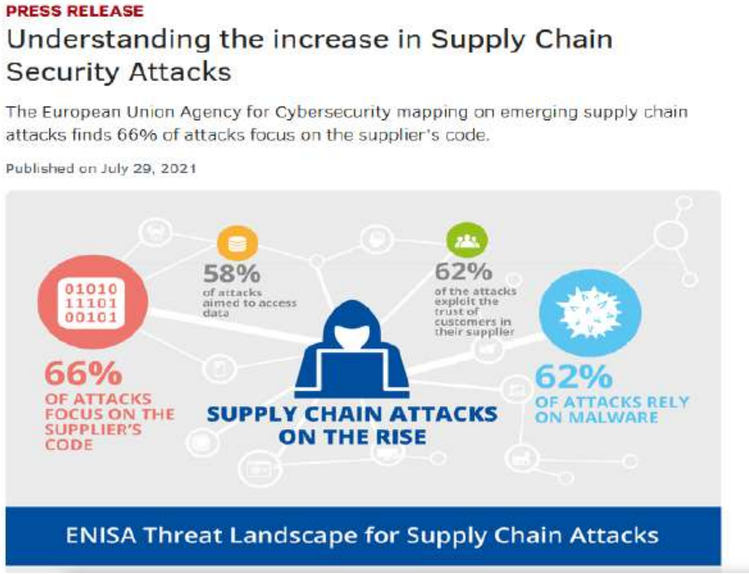
Benadjila R., Prouff E., Strullu R., Cagli E., Dumas C., (2018). Study of deep learning techniques for side-channel analysis and introduction to ASCAD database. ANSSI, France & CEA, LETI, MINATEC Campus, France.

Supply Chain Vulnerability

- Understanding supply chain is important in establishing security in hardware level and software level


End-to-End Supply Chain Protection Required to Preserve Trust
Build what you authorized safely in ANY environment. Protect Your Brand!
Block activation of unauthorized components
Protect keys & detect breaches in real time
Protect platforms & systems
MANUFACTURING
TRANSIT
INSTALLATION
SYSTEM OPERATION
END OF LIFE
Counterfeits and Overbuilds
Equipment Hijacking
Malicious Backdoors and Theft
Malicious Peripherals
LATTICE SEMICONDUCTOR




PRESS RELEASE

Understanding the increase in Supply Chain Security Attacks

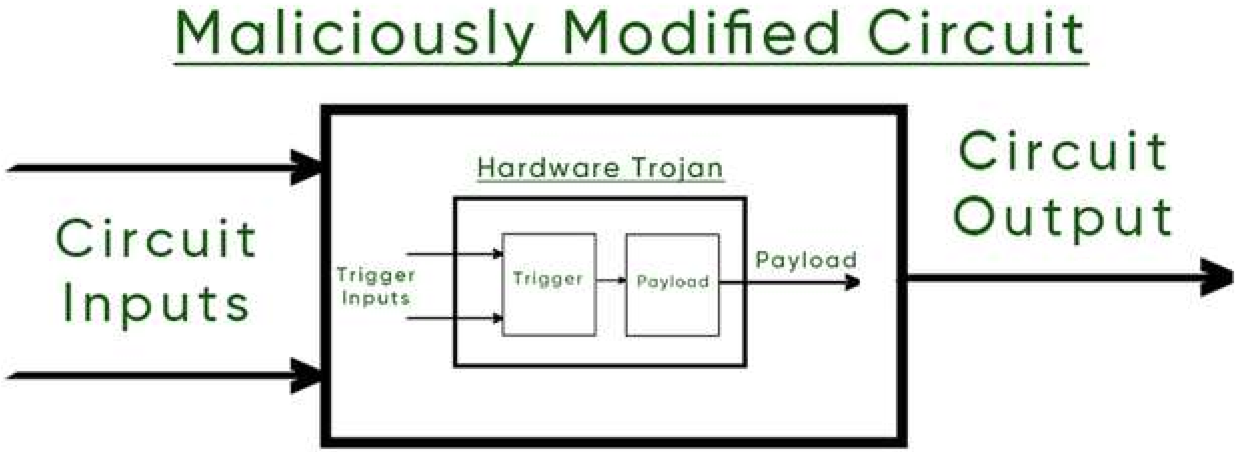
The European Union Agency for Cybersecurity mapping on emerging supply chain attacks finds 66% of attacks focus on the supplier's code.

Published on July 29, 2021

01010
11101
00101
66%
OF ATTACKS FOCUS ON THE SUPPLIER'S CODE
58%
of attacks aimed to access data
62%
of the attacks exploit the trust of customers in their supplier
62%
OF ATTACKS RELY ON MALWARE
SUPPLY CHAIN ATTACKS ON THE RISE
ENISA Threat Landscape for Supply Chain Attacks

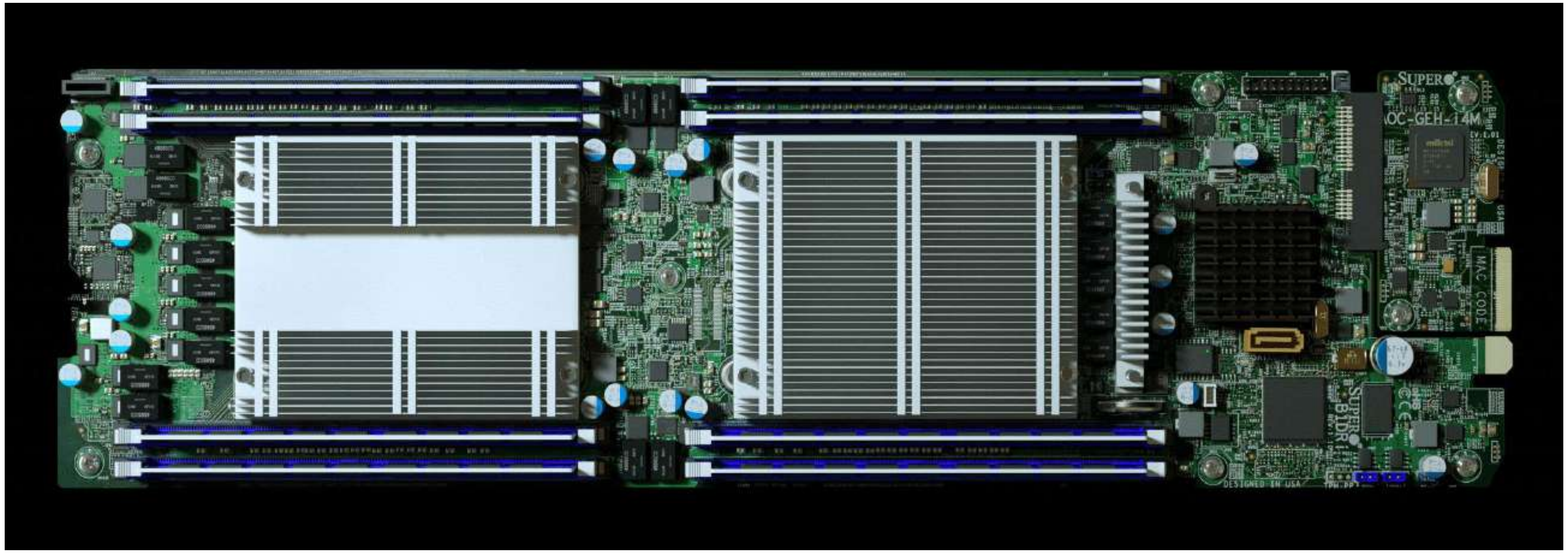
Hardware Trojan

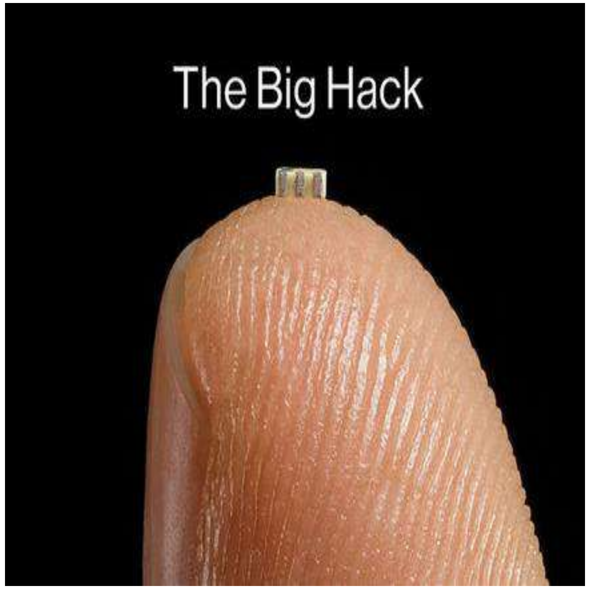
- What if smart lights at your home controlled by others over internet?
- Hardware Trojans (HT) are malicious modification inserted in an electronic device (processor, IC)
- HT can change the functionality of the circuit or leaks confidential information
- Trustworthiness of an electronic system is not guaranteed
- HTs are smaller compared to the design and are passive until it gets activated by a trigger (specific input/sequence)



The big hack

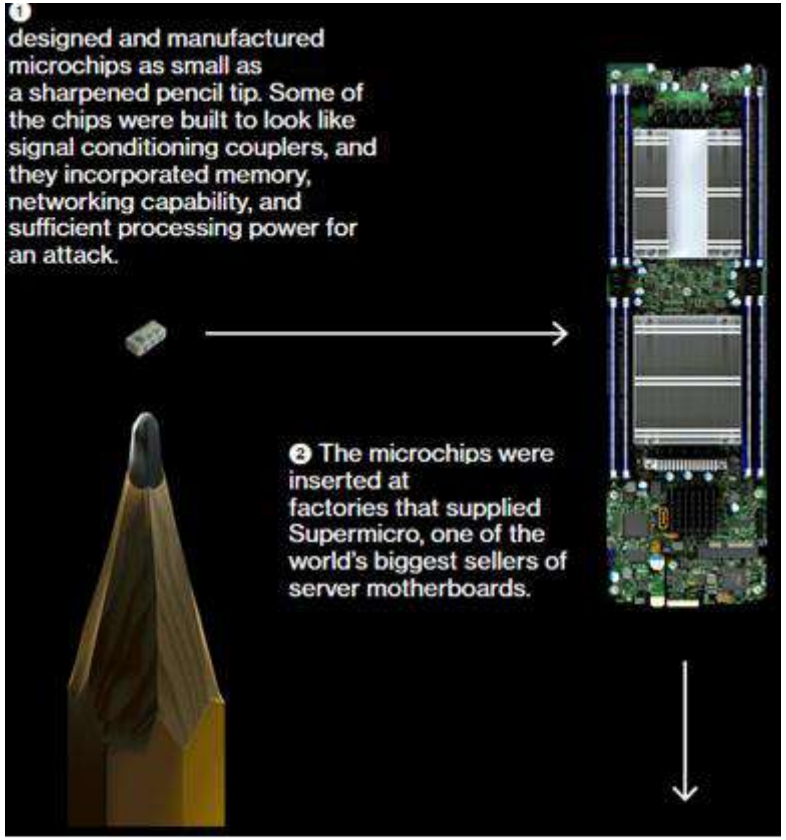
- Motherboard used in National Intelligence community and top companies
- Smaller than a grain of rice
- In jumper and between layers
- Hardware supply chain hack





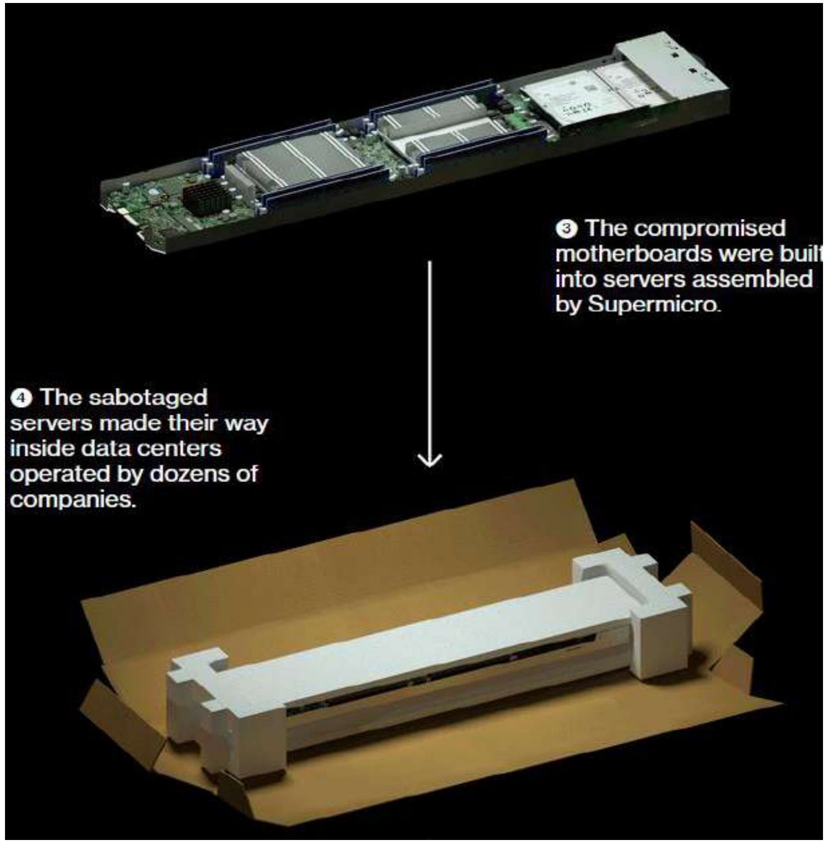
Source: www.bloomberg.com

How the Hack worked?



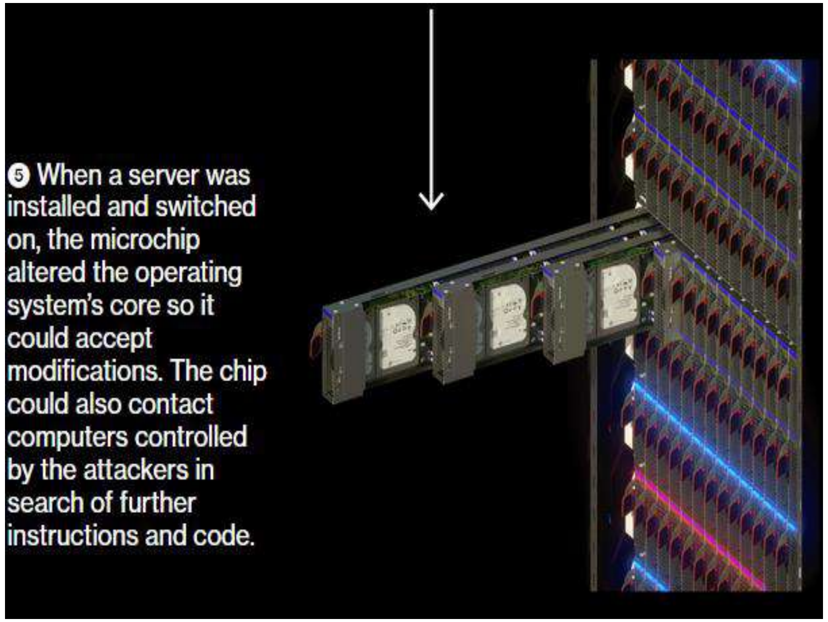
❶ designed and manufactured microchips as small as a sharpened pencil tip. Some of the chips were built to look like signal conditioning couplers, and they incorporated memory, networking capability, and sufficient processing power for an attack.

❷ The microchips were inserted at factories that supplied Supermicro, one of the world's biggest sellers of server motherboards.



❸ The compromised motherboards were built into servers assembled by Supermicro.

❹ The sabotaged servers made their way inside data centers operated by dozens of companies.

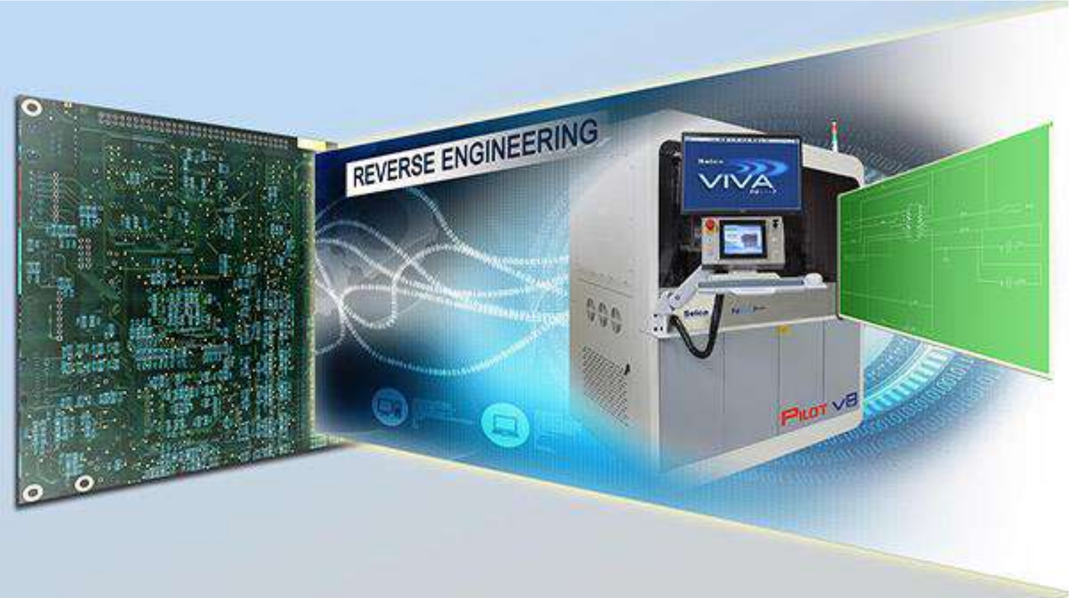


❺ When a server was installed and switched on, the microchip altered the operating system's core so it could accept modifications. The chip could also contact computers controlled by the attackers in search of further instructions and code.

Source: www.bloomberg.com

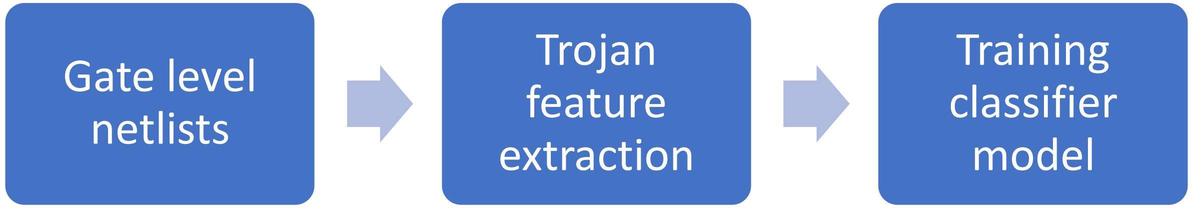
Hardware trust through Reverse Engineering

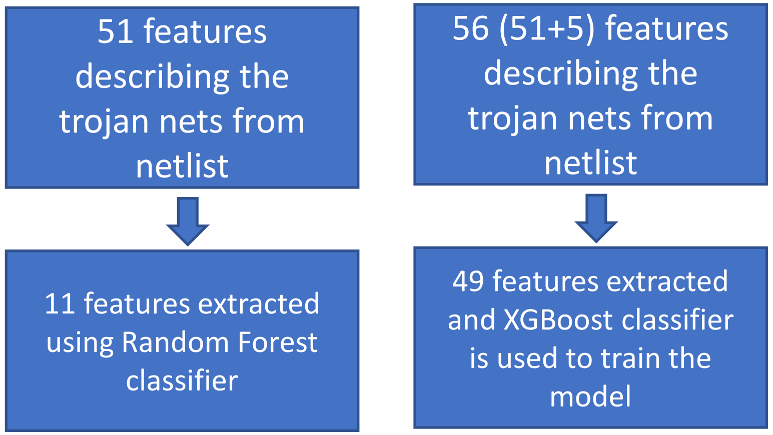
- Through electrical testing and/or physical inspection retrieving
  - An electronic design layout and/or netlist,
  - Stored information (memory contents, firmware, software, etc.)
  - Functionality/specification
- Golden data can be images from a known authentic chip or PCB, schematic, layout, or device, whose functionality, structural and electrical parametric signatures are available for comparison



Source: Botero, ULBERT J., et al. "Hardware Trust and Assurance through Reverse Engineering." *Association for Computing Machinery: New York, NY, USA* (2020).
www.seica.com

AI based Hardware Trojan detection




51 features describing the trojan nets from netlist
11 features extracted using Random Forest classifier
56 (51+5) features describing the trojan nets from netlist
49 features extracted and XGBoost classifier is used to train the model

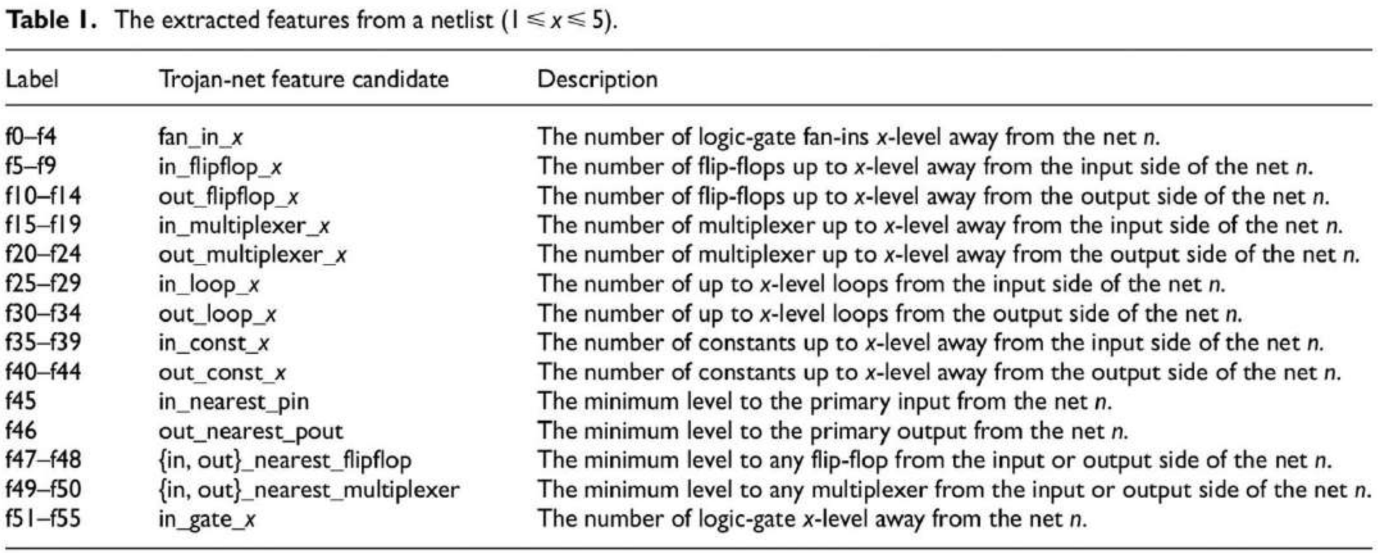


Table 1. The extracted features from a netlist (1 ≤ x ≤ 5).

Label	Trojan-net feature candidate	Description
f0–f4	fan_in_*x*	The number of logic-gate fan-ins *x*-level away from the net *n*.
f5–f9	in_flipflop_*x*	The number of flip-flops up to *x*-level away from the input side of the net *n*.
f10–f14	out_flipflop_*x*	The number of flip-flops up to *x*-level away from the output side of the net *n*.
f15–f19	in_multiplexer_*x*	The number of multiplexer up to *x*-level away from the input side of the net *n*.
f20–f24	out_multiplexer_*x*	The number of multiplexer up to *x*-level away from the output side of the net *n*.
f25–f29	in_loop_*x*	The number of up to *x*-level loops from the input side of the net *n*.
f30–f34	out_loop_*x*	The number of up to *x*-level loops from the output side of the net *n*.
f35–f39	in_const_*x*	The number of constants up to *x*-level away from the input side of the net *n*.
f40–f44	out_const_*x*	The number of constants up to *x*-level away from the output side of the net *n*.
f45	in_nearest_pin	The minimum level to the primary input from the net *n*.
f46	out_nearest_pout	The minimum level to the primary output from the net *n*.
f47–f48	{in, out}_nearest_flipflop	The minimum level to any flip-flop from the input or output side of the net *n*.
f49–f50	{in, out}_nearest_multiplexer	The minimum level to any multiplexer from the input or output side of the net *n*.
f51–f55	in_gate_*x*	The number of logic-gate *x*-level away from the net *n*.

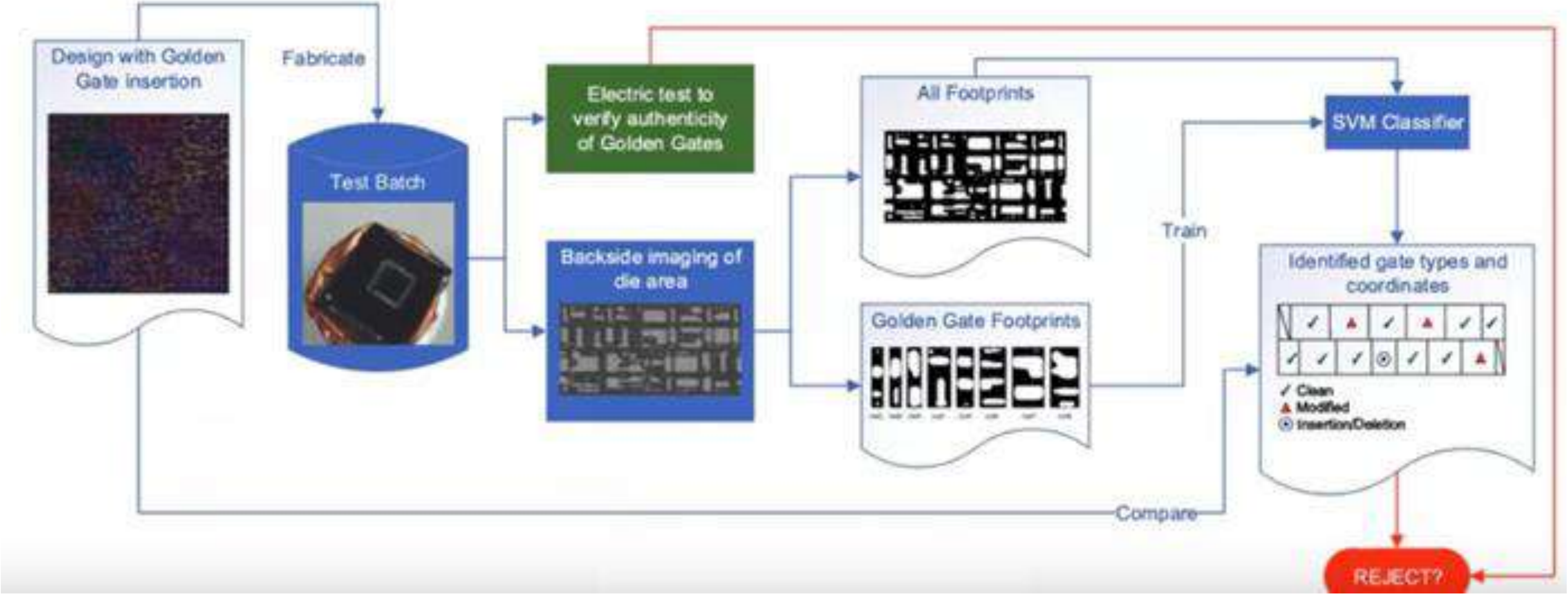
https://trust-hub.org/

Source:K. Hasegawa, M. Yanagisawa and N. Togawa, "Trojan-feature extraction at gate-level netlists and its application to hardware-Trojan detection using random forest classifier," *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1-4, doi: 10.1109/ISCAS.2017.8050827.

Source: Dong, C., Chen, J., Guo, W., & Zou, J. (2019). A machine-learning-based hardware-Trojan detection approach for chips in the Internet of Things. *International Journal of Distributed Sensor Networks*, *15*(12), 1550147719888098.

AI based Hardware Trojan detection

- When a chip is designed there is a ‘Golden netlist’ available
- SVM classifier – train with golden netlist
- Backside imaging of die area


Design with Golden Gate Insertion
Fabricate
Test Batch
Electric test to verify authenticity of Golden Gates
Backside imaging of die area
All Footprints
Golden Gate Footprints
Train
SVM Classifier
Identified gate types and coordinates
✓ Clean
▲ Modified
⦿ Insertion/Deletion
Compare
REJECT?

Source: Florida Institute for Cybersecurity Research

AI in Physical Inspection of Electronics (PIE)

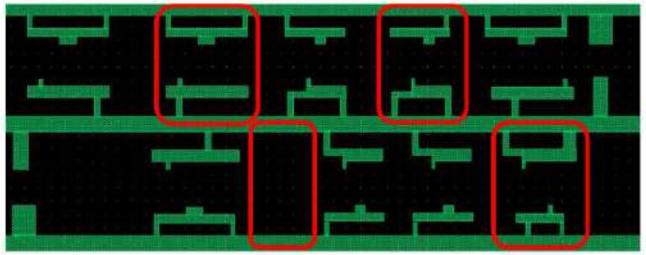
- The examination and verification of various hardware information (e.g. their physical patterns, connections, and functionalities) from their Scanning Electron Microscopy (SEM) images
- Interlayer and Surface analysis
- AI is used to examine and verify various hardware information from analyzing its SEM images
- Introducing computer vision models into the context of physical inspection of electronics
- Misalignment detection- CNN

Source:K. Hasegawa, M. Yanagisawa and N. Togawa, "Trojan-feature extraction at gate-level netlists and its application to hardware-Trojan detection using random forest classifier," *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1-4, doi: 10.1109/ISCAS.2017.8050827.

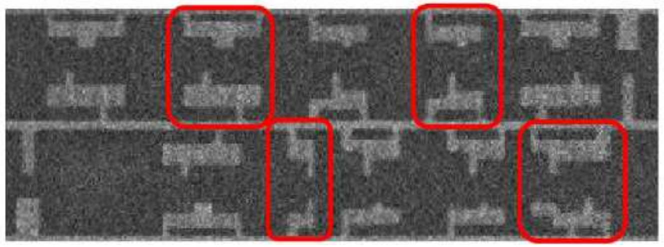
Source: Dong, C., Chen, J., Guo, W., & Zou, J. (2019). A machine-learning-based hardware-Trojan detection approach for chips in the Internet of Things. *International Journal of Distributed Sensor Networks*, *15*(12), 1550147719888098.

Scanning Electron Microscopy (SEM) images

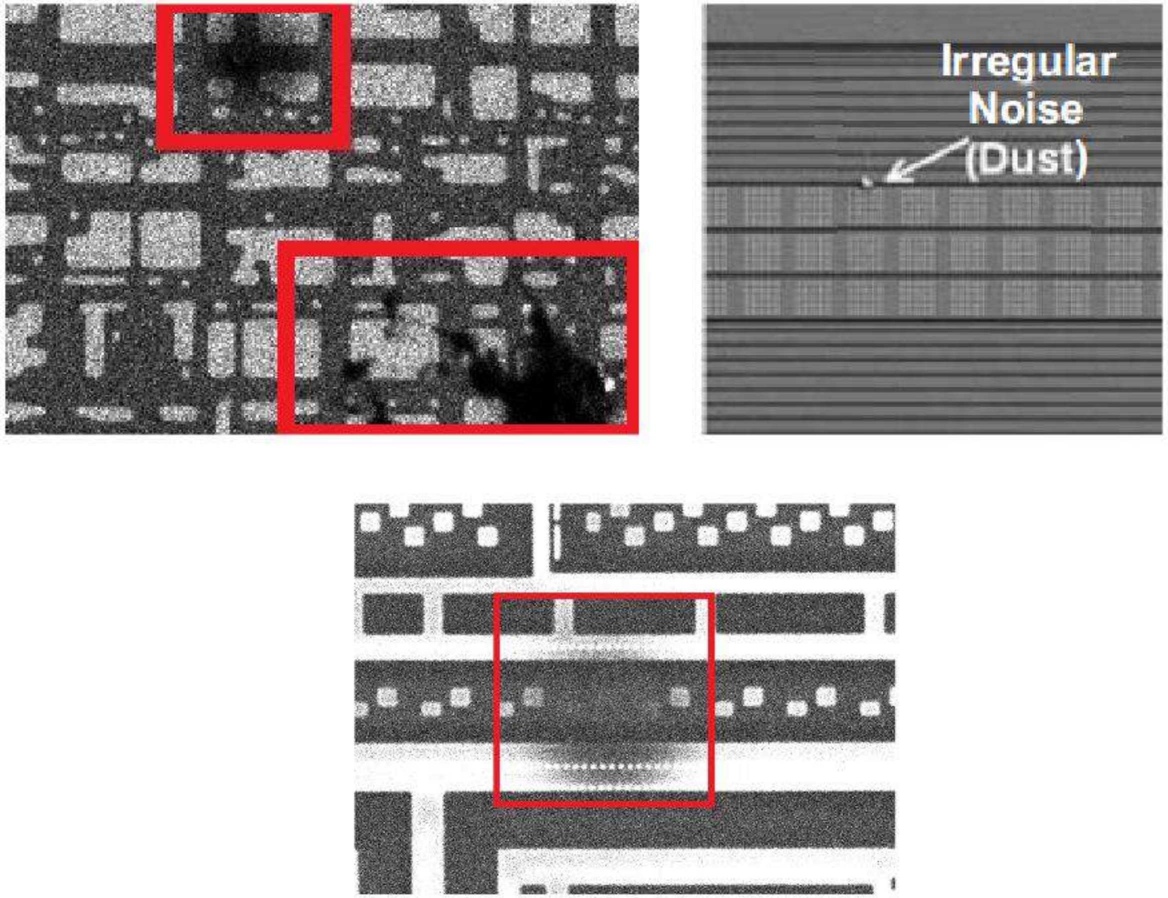
Layout of a smart card by authorized and trusted PCB manufacturer



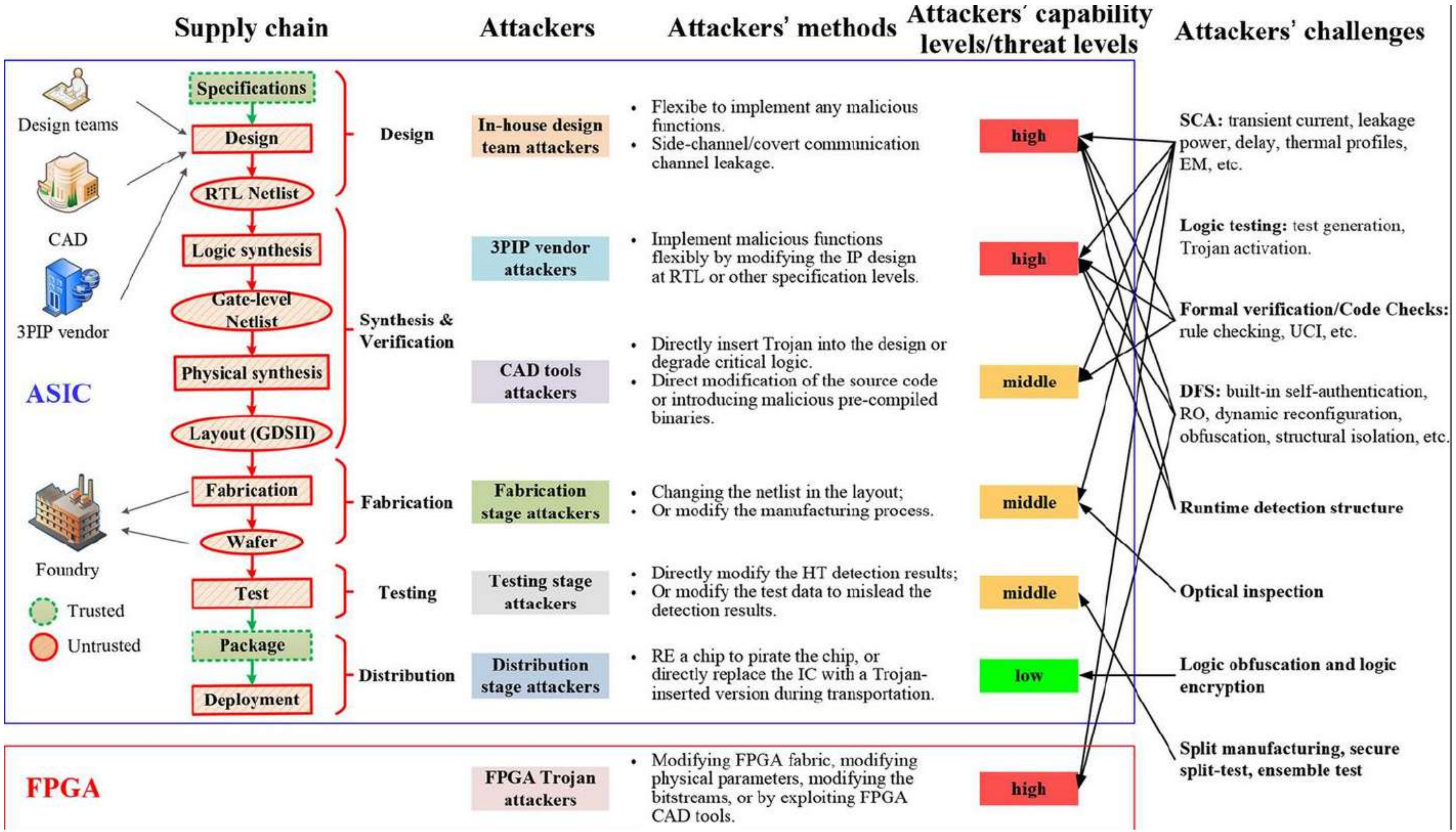
The designer’s layout to be compared



Challenges



Hardware Trojans



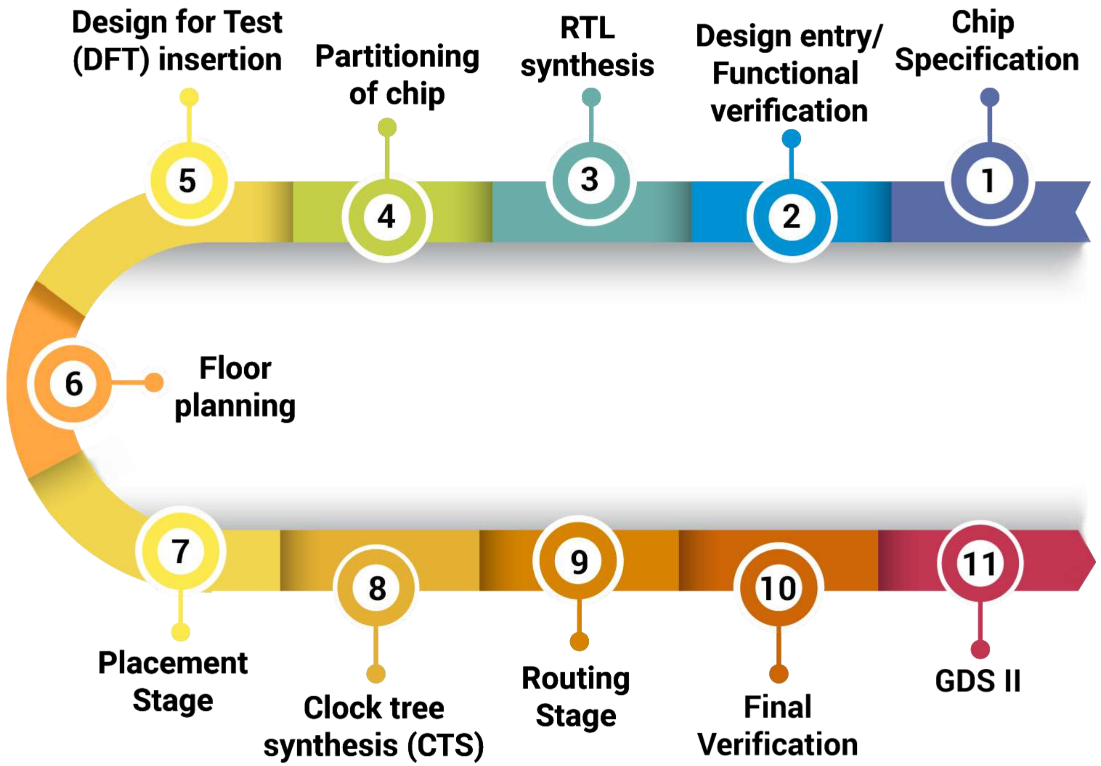
Supply chain	Attackers	Attackers' methods	Attackers´ capability levels/threat levels
Design	In-house design team attackers	• Flexibe to implement any malicious functions. • Side-channel/covert communication channel leakage.	high
Synthesis & Verification	3PIP vendor attackers	• Implement malicious functions flexibly by modifying the IP design at RTL or other specification levels.	high
	CAD tools attackers	• Directly insert Trojan into the design or degrade critical logic. • Direct modification of the source code or introducing malicious pre-compiled binaries.	middle
Fabrication	Fabrication stage attackers	• Changing the netlist in the layout; • Or modify the manufacturing process.	middle
Testing	Testing stage attackers	• Directly modify the HT detection results; • Or modify the test data to mislead the detection results.	middle
Distribution	Distribution stage attackers	• RE a chip to pirate the chip, or directly replace the IC with a Trojan-inserted version during transportation.	low
FPGA	FPGA Trojan attackers	• Modifying FPGA fabric, modifying physical parameters, modifying the bitstreams, or by exploiting FPGA CAD tools.	high

Attackers' challenges

- **SCA:** transient current, leakage power, delay, thermal profiles, EM, etc.
- **Logic testing:** test generation, Trojan activation.
- **Formal verification/Code Checks:** rule checking, UCI, etc.
- **DFS:** built-in self-authentication, RO, dynamic reconfiguration, obfuscation, structural isolation, etc.
- **Runtime detection structure**
- **Optical inspection**
- **Logic obfuscation and logic encryption**
- **Split manufacturing, secure split-test, ensemble test**

Source: Xue, Mingfu, et al. "Ten years of hardware Trojans: a survey from the attacker's perspective." *IET Computers & Digital Techniques* 14.6 (2020): 231-246.

AI based secure physical design flow



Hardware must be designed considering security

- Memory attacks
- Fault injection
- Side Channel Attacks

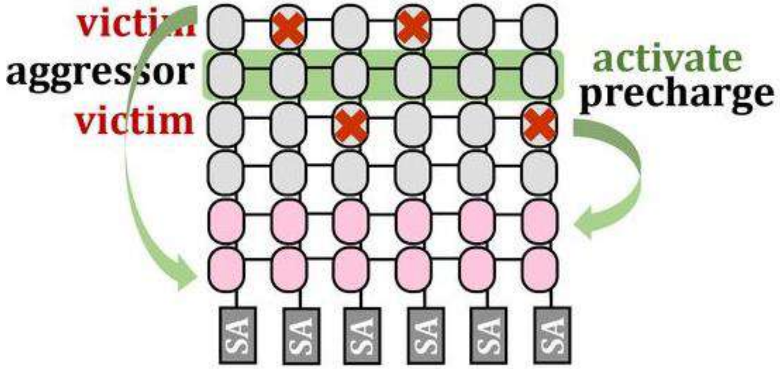
AI (Reinforcement Learning) can be adopted in ensuring secure physical design

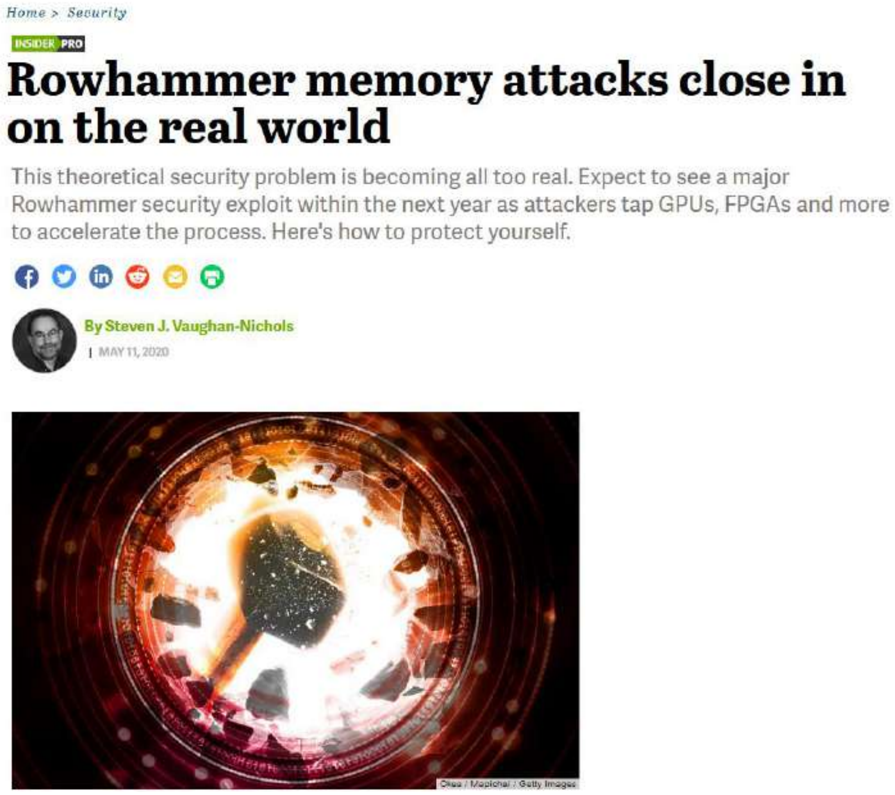
Source: Florida Institute for Cybersecurity Research, einfochips

Memory attacks

Rowhammer attack

- Flip bits stored in DRAM without accessing
- If a particular row is continuously accessed within refresh period; influences neighbouring rows (toggles the data)
- Privilege can be changed





Home > Security

INSIDER PRO

Rowhammer memory attacks close in on the real world

This theoretical security problem is becoming all too real. Expect to see a major Rowhammer security exploit within the next year as attackers tap GPUs, FPGAs and more to accelerate the process. Here's how to protect yourself.

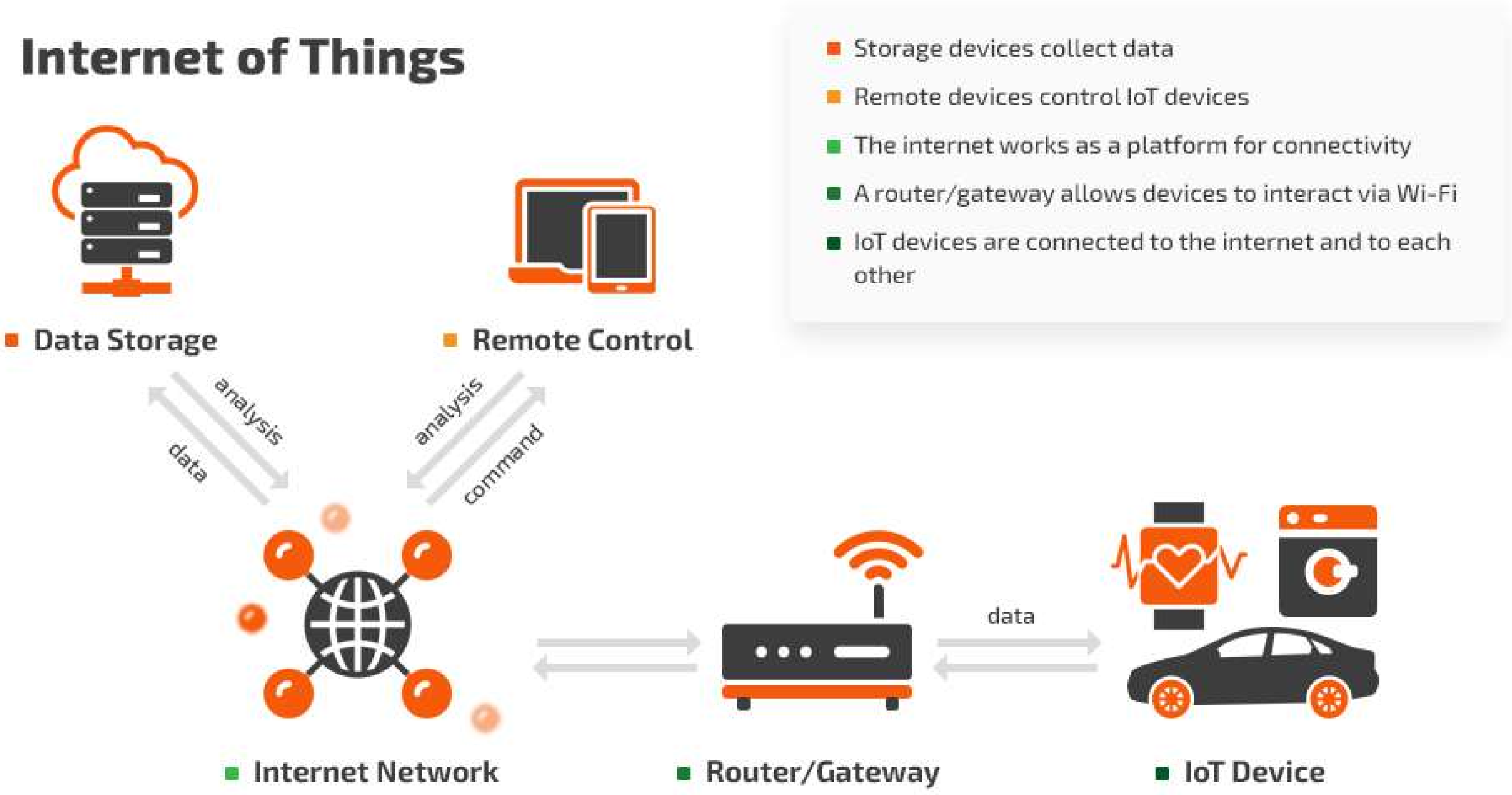
By Steven J. Vaughan-Nichols

| MAY 11, 2020

Olea / Mapichai / Getty Images

Source: In CROW: A Low-Cost Substrate for Improving DRAM Performance, Energy Efficiency, and Reliability

AI (TinyML) for IoT security



- Storage devices collect data
- Remote devices control IoT devices
- The internet works as a platform for connectivity
- A router/gateway allows devices to interact via Wi-Fi
- IoT devices are connected to the internet and to each other

3 Common IoT Attacks that Compromise Security

Unpatched vulnerabilities and design flaws in IoT devices have become a gateway for threat actors to penetrate user and corporate networks.

By **CISOMAG** - September 23, 2021

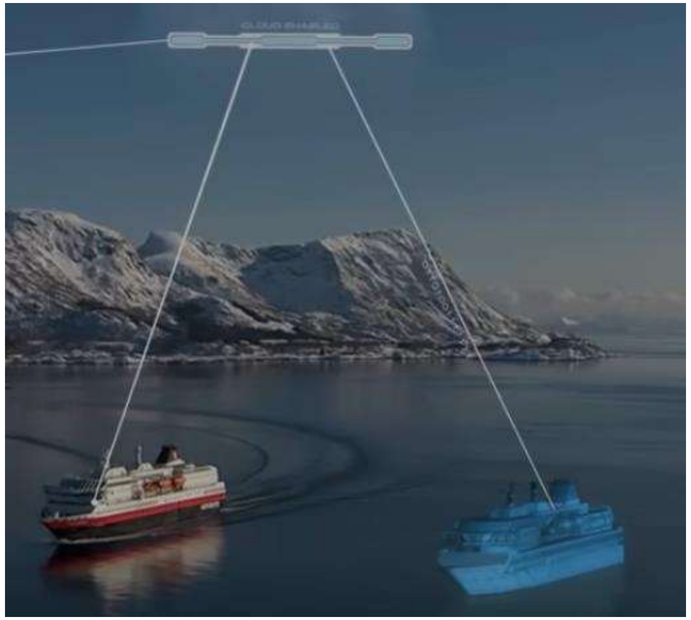
- Eavesdropping
- Privilege Escalation Attack
- Brute-Force attack

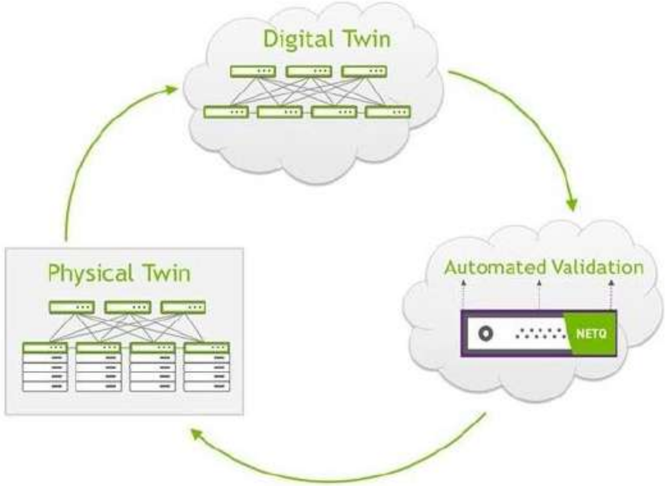
The high degree of device heterogeneity (different devices, some old and some new, but each with its own operating systems and particular vulnerabilities) makes IoT networks prime targets for hackers as they target weak links

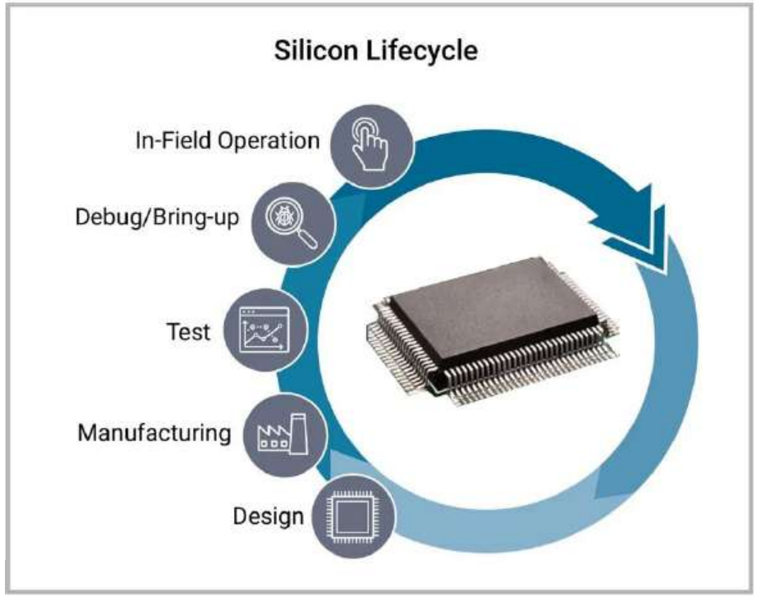
Source: www.apriorit.com

AI in Digital Twin for Hardware Security

- Virtual model of physical devices are modelled
- The model performs assessment using the information captured (data driven based on AI)
- Helps in finding the root cause of the fault and at which step of the Lifecycle



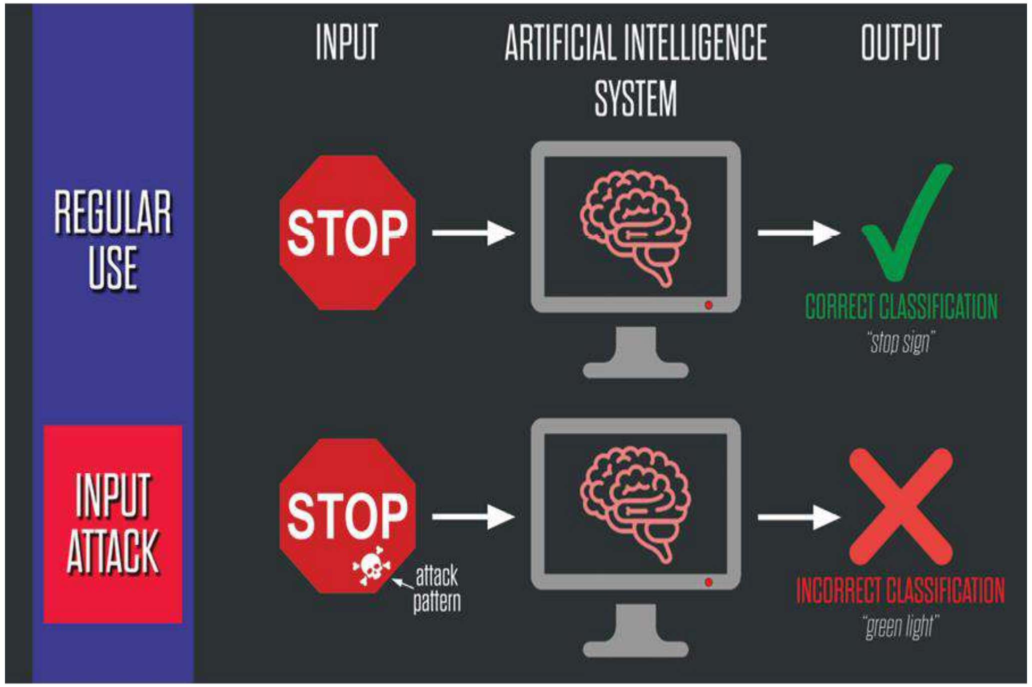




Source: Florida Institute for Cybersecurity Research
Source:https://semiengineering.com/microelectronics-and-the-ai-revolution/

AI against Hardware Security?

- Poisoning Attacks
  - Data poisoning attack – when attackers turn AI model against the developers
  - Data poisoning involves tampering with machine learning training data to produce undesirable outcomes
  - Algorithm Poisoning
  - Model Poisoning



Source: Jagielski, Matthew, et al. "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning." *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018.

Side Channel Attacks

- Side channels have the ability to render even mathematically robust cryptographic algorithms - vulnerable
- A side-channel adversary observes the physical properties of a cryptographic implementation, such as timing, power or electromagnetic emanations
- The attacker tries to infer the secret key by modelling a sensitive intermediate state of the design which then depends on the physical properties

Differential Power Analysis (DPA)

- The intermediate values depend on the plain text and key; exhaustive key search (2^n) technique can be followed for key retrieval
- For less secure algorithms

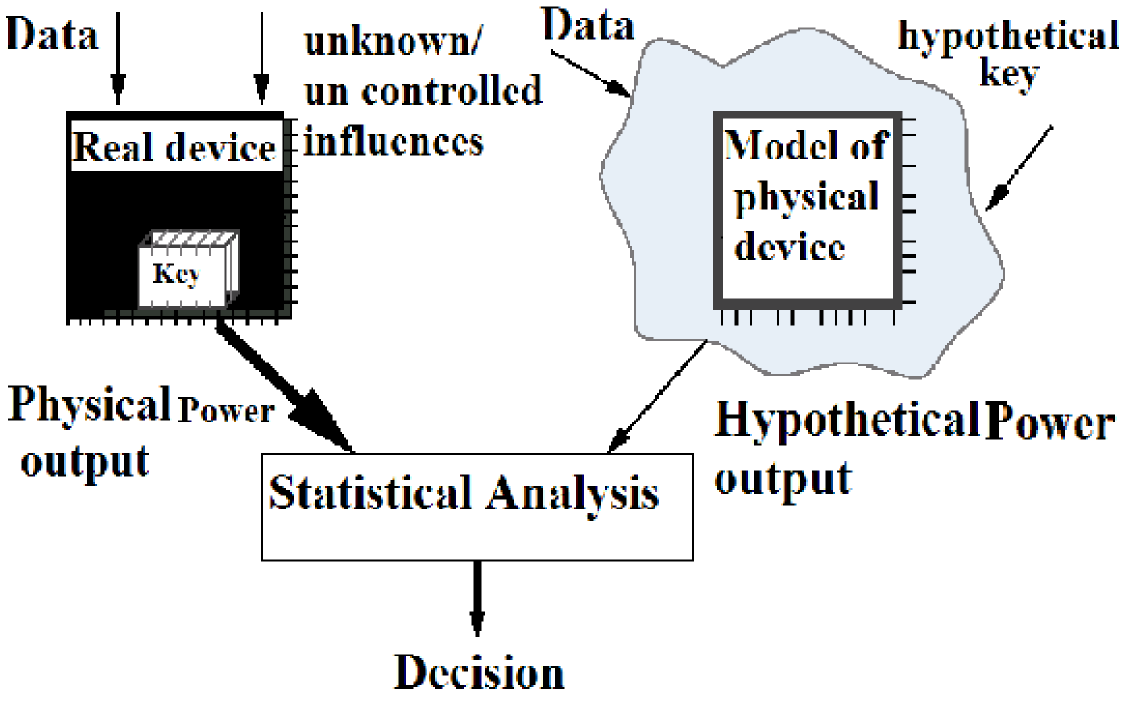
1 Choosing an intermediate result

2 Measuring the power consumption

3 Calculating hypothetical intermediate value

4 Mapping intermediate values to power consumption

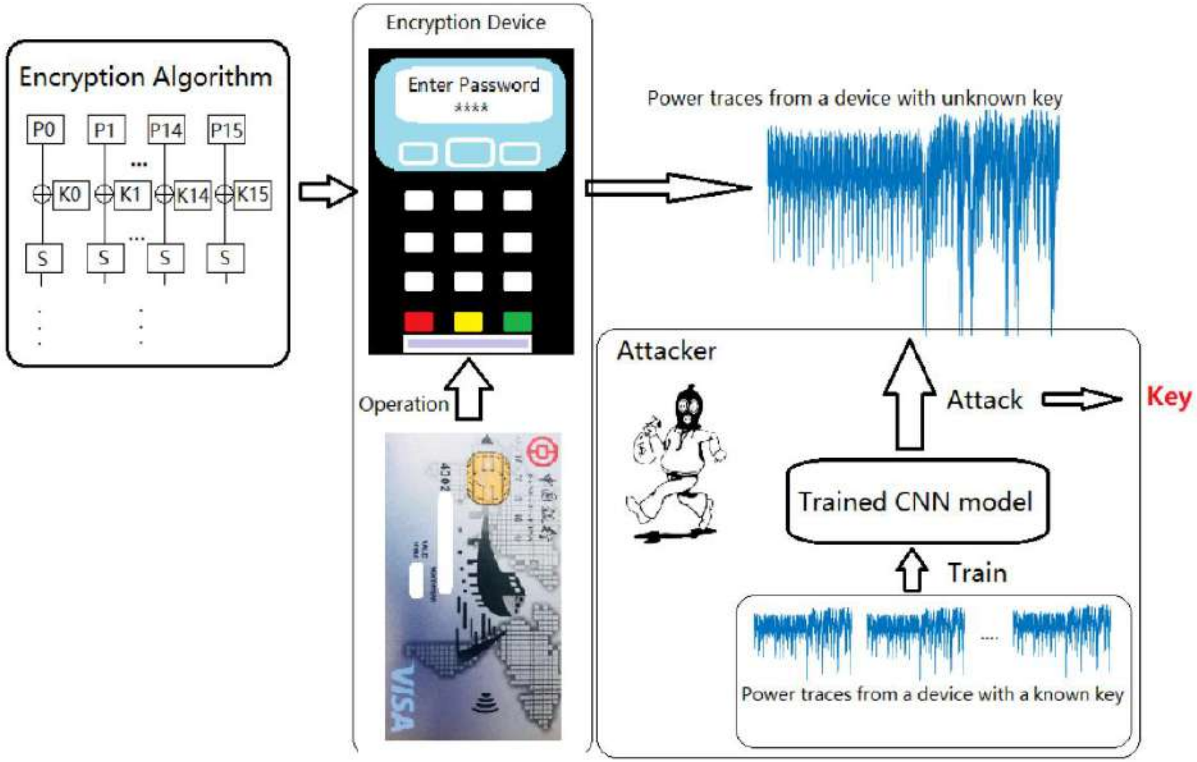
5 Finding correlation between hypothetical power consumption and power traces



$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}}$$

Kocher, P., Jaffe, J., & Jun, B. (1999). Differential power analysis. In *Annual international cryptology conference* (pp. 388-397). Springer, Berlin, Heidelberg.

AI in Key retrieval

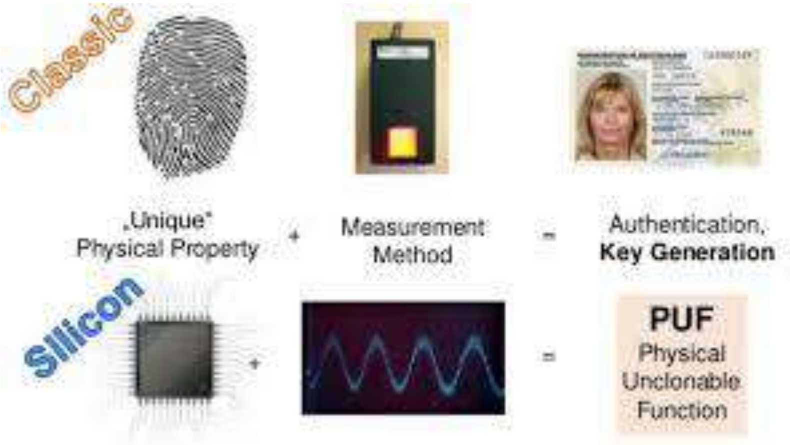


- The model is trained with the traces captured from the similar device (profiling device) of the target device (attack device) with known inputs
- Then the trained model is used to retrieve the key from the trace captured from the target device

Source: Huanyu Wang, Side-Channel Analysis of AES Based on Deep Learning, KTH Royal Institute of Technology

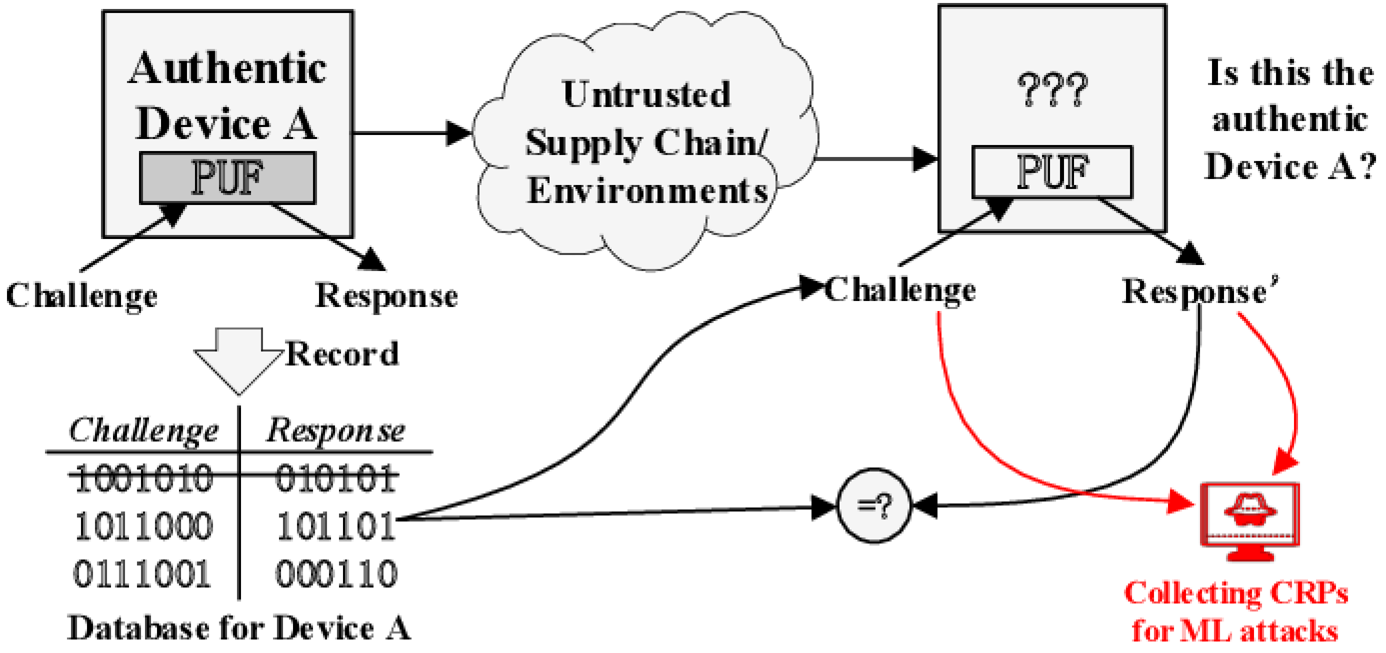
Physically Unclonable Function (PUF)

- PUF is a physical system used in hardware security that exploits inherent device variations, such that it's clone cannot be reproduced.
- Each PUF has it's own unique characteristics (fingerprint)
- PUFs depend on the uniqueness of their physical microstructure
- The microstructure depends on random physical factors introduced during manufacturing
- Used for key generation and storage
- Protocols with Challenge Response Pairs can be used for device authentication



Source: https://scl.engr.uconn.edu/courses/ece4451/Lec10_PUFs.pdf

AI-PUF


Authentic Device A
PUF
Challenge
Response
Untrusted Supply Chain/ Environments
???
PUF
Is this the authentic Device A?
Challenge
Response'
Record
=?
Collecting CRPs for ML attacks

Challenge	*Response*
~~1001010~~	~~010101~~
1011000	101101
0111001	000110

Database for Device A

Source: Xu, Xiaolin, and Jiliang Zhang. "Rethinking FPGA security in the new era of artificial intelligence." *2020 21st International Symposium on Quality Electronic Design (ISQED)*. IEEE, 2020.

Artificial Intelligence and Hardware Security

The ongoing cat-and-mouse game of security

- The relationship between machine learning and hardware security is:
  - Defenders can use AI with hardware-based observations to build models of an IC’s operation for attack detection against adversaries (in terms of software, hardware, and environmental conditions)
  - AI can help an adversary to attack a system to extract sensitive information from an IC, breaking trust assumptions in hardware security
- The AI circuitry itself can be the target of attacks

A cryptosystem should be secure even if everything about the system, except the key, is public knowledge

-Kerckhoffs's principle

References

- Benadjila R., Prouff E., Strullu R., Cagli E., Dumas C., (2018). Study of deep learning techniques for side-channel analysis and introduction to ASCAD database. ANSSI, France & CEA, LETI, MINATEC Campus, France.
- Botero, ULBERT J., et al. "Hardware Trust and Assurance through Reverse Engineering." *Association for Computing Machinery: New York, NY, USA* (2020).
- K. Hasegawa, M. Yanagisawa and N. Togawa, "Trojan-feature extraction at gate-level netlists and its application to hardware-Trojan detection using random forest classifier," 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 2017, pp. 1-4, doi: 10.1109/ISCAS.2017.8050827.
- Dong, C., Chen, J., Guo, W., & Zou, J. (2019). A machine-learning-based hardware-Trojan detection approach for chips in the Internet of Things. International Journal of Distributed Sensor Networks, 15(12), 1550147719888098.
- Florida Institute for Cybersecurity Research
- K. Hasegawa, M. Yanagisawa and N. Togawa, "Trojan-feature extraction at gate-level netlists and its application to hardware-Trojan detection using random forest classifier," 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 2017, pp. 1-4, doi: 10.1109/ISCAS.2017.8050827.
- Dong, C., Chen, J., Guo, W., & Zou, J. (2019). A machine-learning-based hardware-Trojan detection approach for chips in the Internet of Things. International Journal of Distributed Sensor Networks, 15(12), 1550147719888098.
- Xue, Mingfu, et al. "Ten years of hardware Trojans: a survey from the attacker's perspective." *IET Computers & Digital Techniques* 14.6 (2020): 231-246.
- A Low-Cost Substrate for Improving DRAM Performance, Energy Efficiency, and Reliability
- https://semiengineering.com/microelectronics-and-the-ai-revolution/
- Jagielski, Matthew, et al. "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning." *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018.
- Kocher, P., Jaffe, J., & Jun, B. (1999). Differential power analysis. In Annual international cryptology conference (pp. 388-397). Springer, Berlin, Heidelberg.
- Huanyu Wang, Side-Channel Analysis of AES Based on Deep Learning, KTH ROYAL INSTITUTE OF TECHNOLOGY
- Xu, Xiaolin, and Jiliang Zhang. "Rethinking FPGA security in the new era of artificial intelligence." *2020 21st International Symposium on Quality Electronic Design (ISQED)*. IEEE, 2020.
- https://scl.engr.uconn.edu/courses/ece4451/Lec10_PUFs.pdf
- https://www.toolbox.com/it-security/vulnerability-management/articles/what-is-hardware-security/#_001
- https://thehackernews.com/2022/01/detecting-evasive-malware-on-iot.html
- The big hack: www.bloomberg.com

Task Force Report on 'CYBSEC4AI'

https://www.psa.gov.in/psa-prod/publication/Taskforce-Report-CybSec4AI-SETS.pdf

Website: www.setsindia.in

Thank You

eswari@setsindia.net