Everything Everywhere AI: Securing next-generation embedded devices

Lejla Batina
Radboud University

KIVI JAARCONGRES
TU/e, March 11, 2026

Outline

- (Embedded) crypto is everywhere
- Side-channel analysis (SCA) attacks are a relevant threat today
- What AI has to do with it*?*
  - Example 1: Screen Gleaning
  - Example 2: Attacking AI using side-channel analysis
- Where do we go from here?

Our world is digital

and with crypto we
try to make it secure

Schoolbook cryptography

Alice

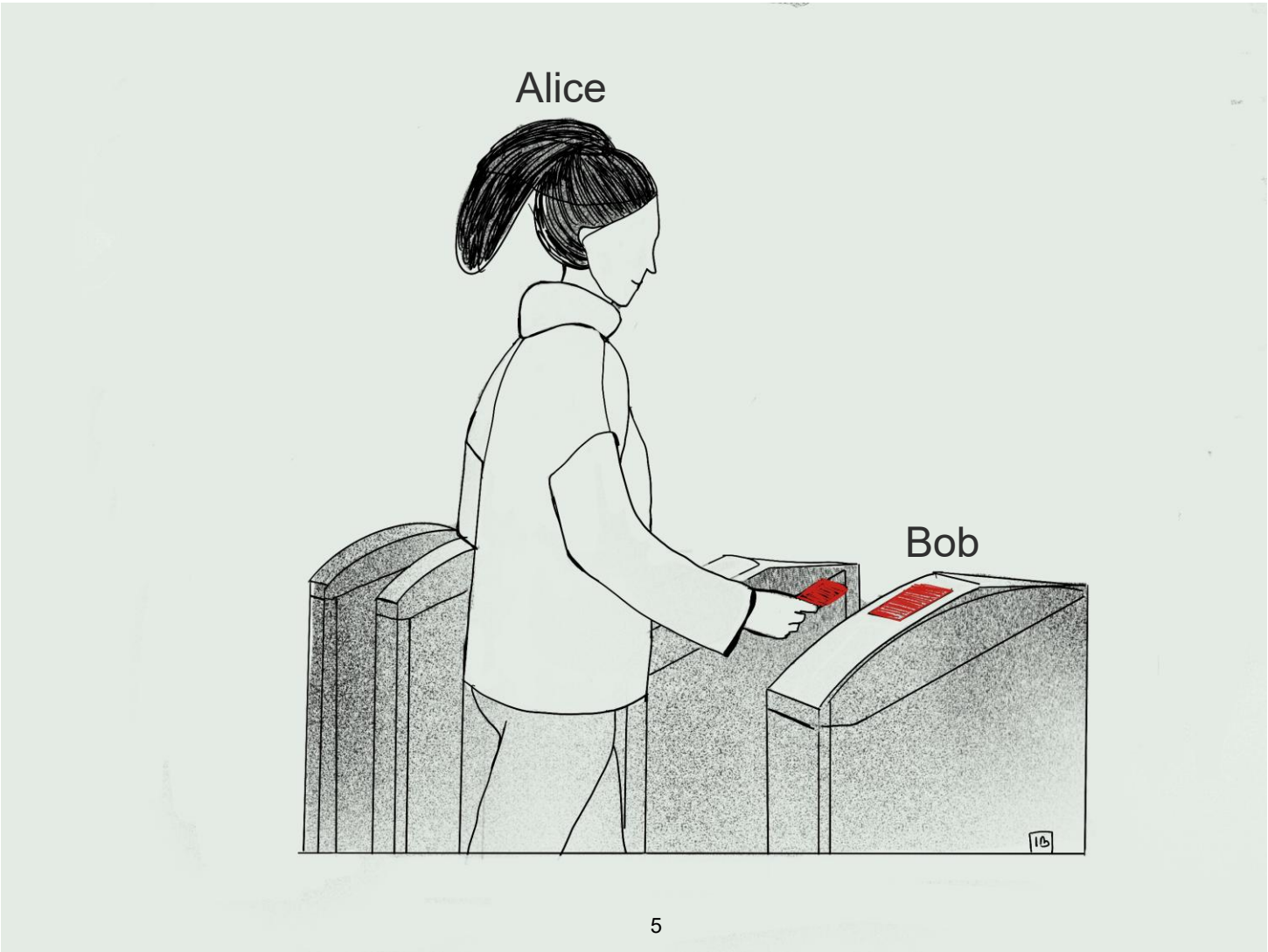
Hi, I'm Alice

Hi Alice, here is your secret data

Eve

Bob

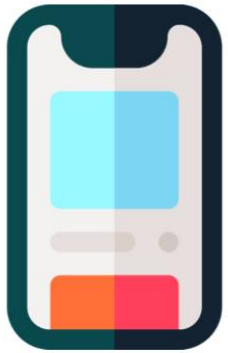
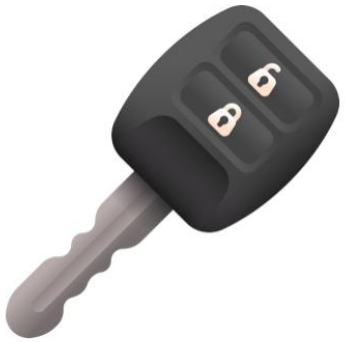
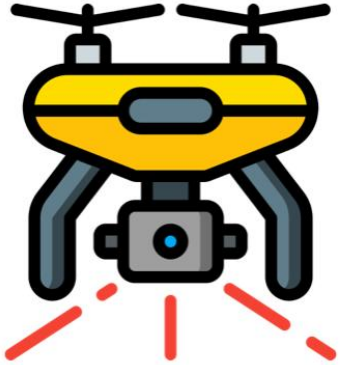
Crypto in the real world

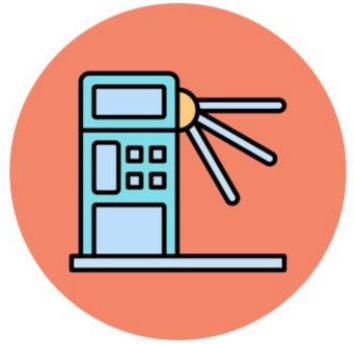


drawing by
Ileana Buhan

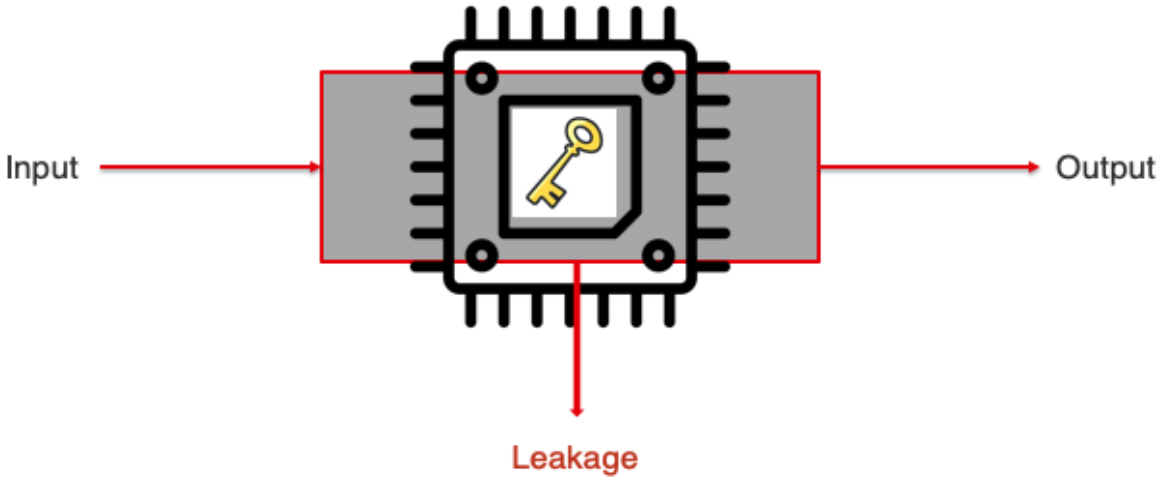
Devices with embedded crypto





Side-channel analysis (SCA)



Eve passive, can measure and process signals

Eve active, can insert her own signals

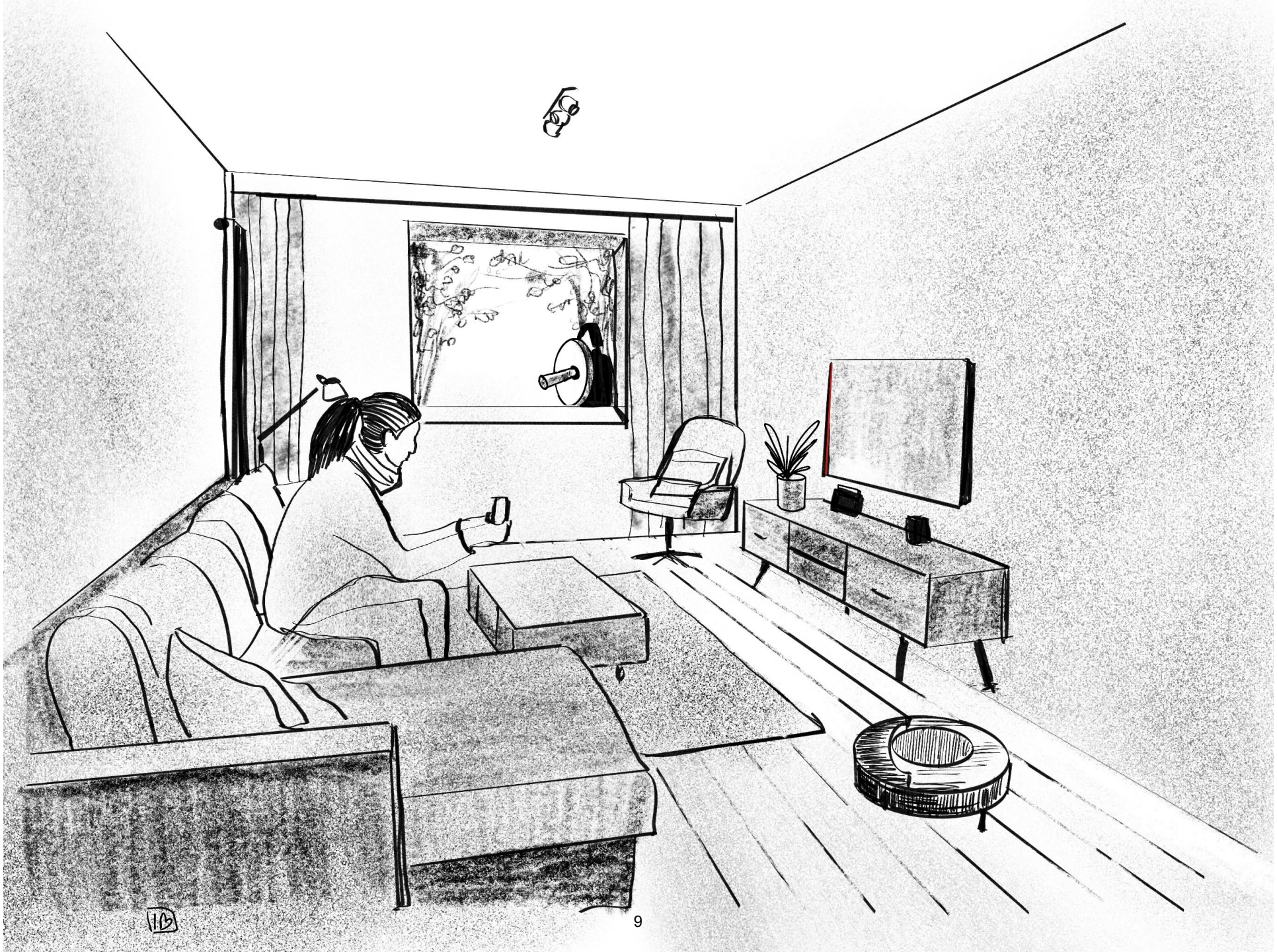
- Crypto is implemented on real devices such as microcontrollers and chips
- Eve's goal: secret key, message, IP, …

Relevance

https://arxiv.org/pdf/2308.01074.pdf

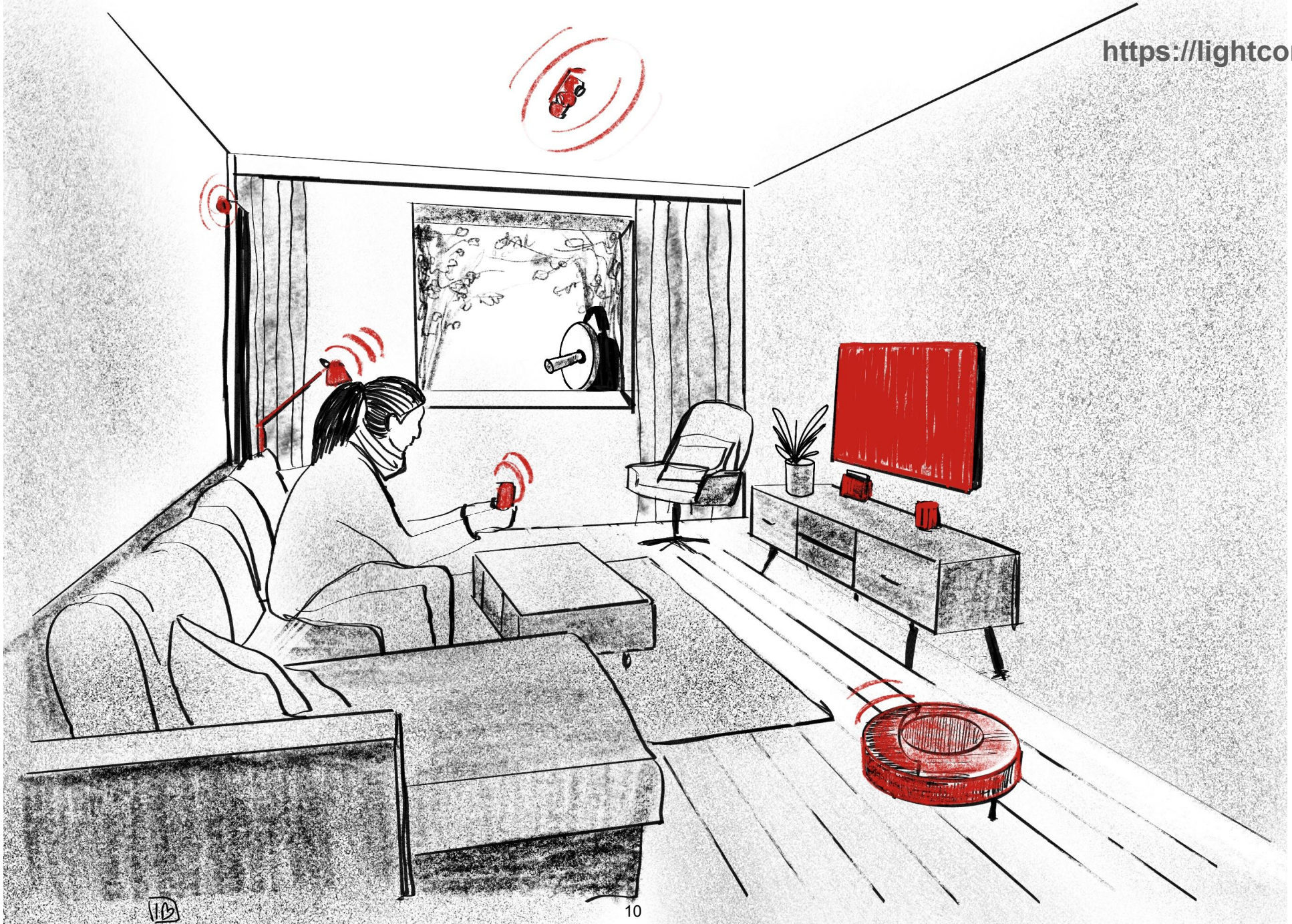
https://ninjalab.io/a-side-journey-to-titan/

Side-channel attacks exploit weaknesses in implementations.



drawing by
Ileana Buhan

https://lightcommands.com/



drawing by
Ileana Buhan

AI and Security

- Machine Learning (ML) applications: image recognition, natural languages, robotics, gaming etc.
- ML in IoT devices for image and speech recognition
- AI in small devices manipulating our data and affecting our privacy

But, can we trust it?

AI and SCA: How it all started

- Machine learning (ML) for SCA was introduced 15 years ago
  - In collaboration with Data Science@RU and Riscure/Keysight (a.o.)

- Deep learning (DL) for SCA became a very active research topic:
  - Neural nets (NNs) for SCA attacks
  - TEMPEST-like techniques e.g. screen gleaning

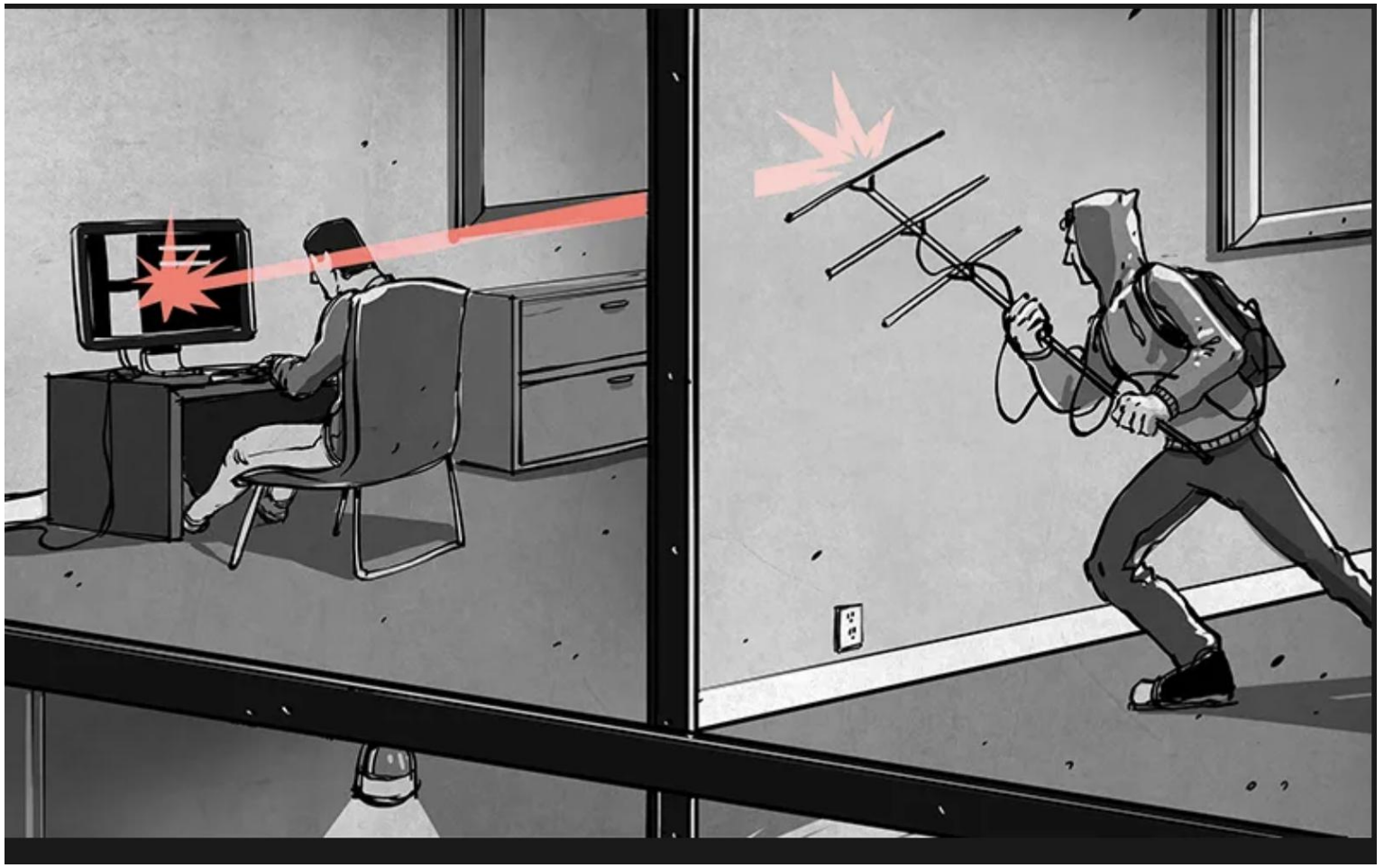
- SCA attacks on implementations of neural nets:
  - Reverse engineering and stealing IP
  - Input recovery from NN implementations

Ex. 1: Screen Gleaning

SCREEN GLEANING: A SCREEN READING TEMPEST ATTACK ON MOBILE DEVICES EXPLOITING AN ELECTROMAGNETIC SIDE CHANNEL

Z. LIU, N. SAMWEL, L. WEISSBART, Z. ZHAO, D. LAURET,
L. BATINA, M. LARSON

TEMPEST: eavesdropping on screens



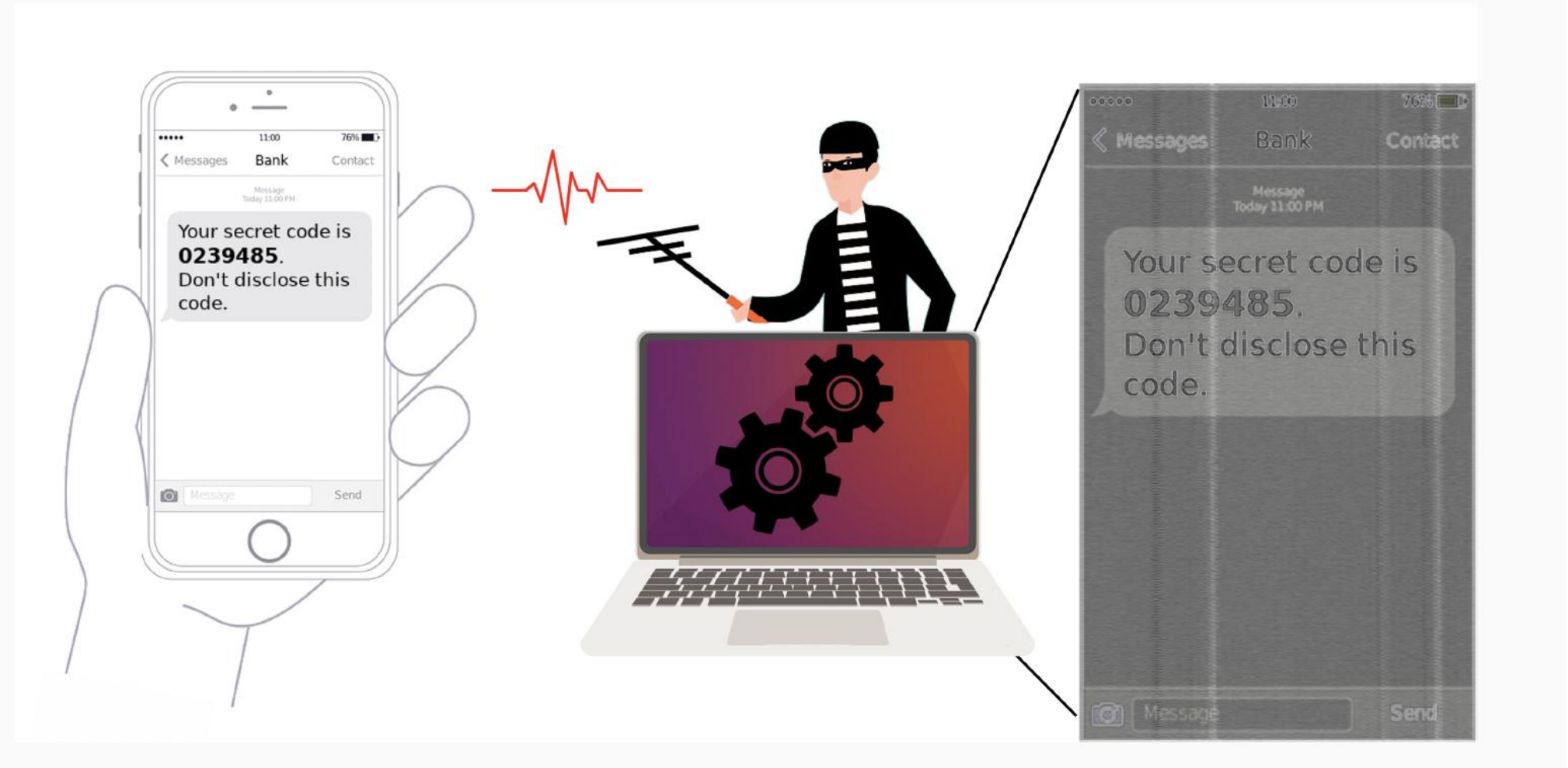
Source: Hackaday.com

TEMPEST history

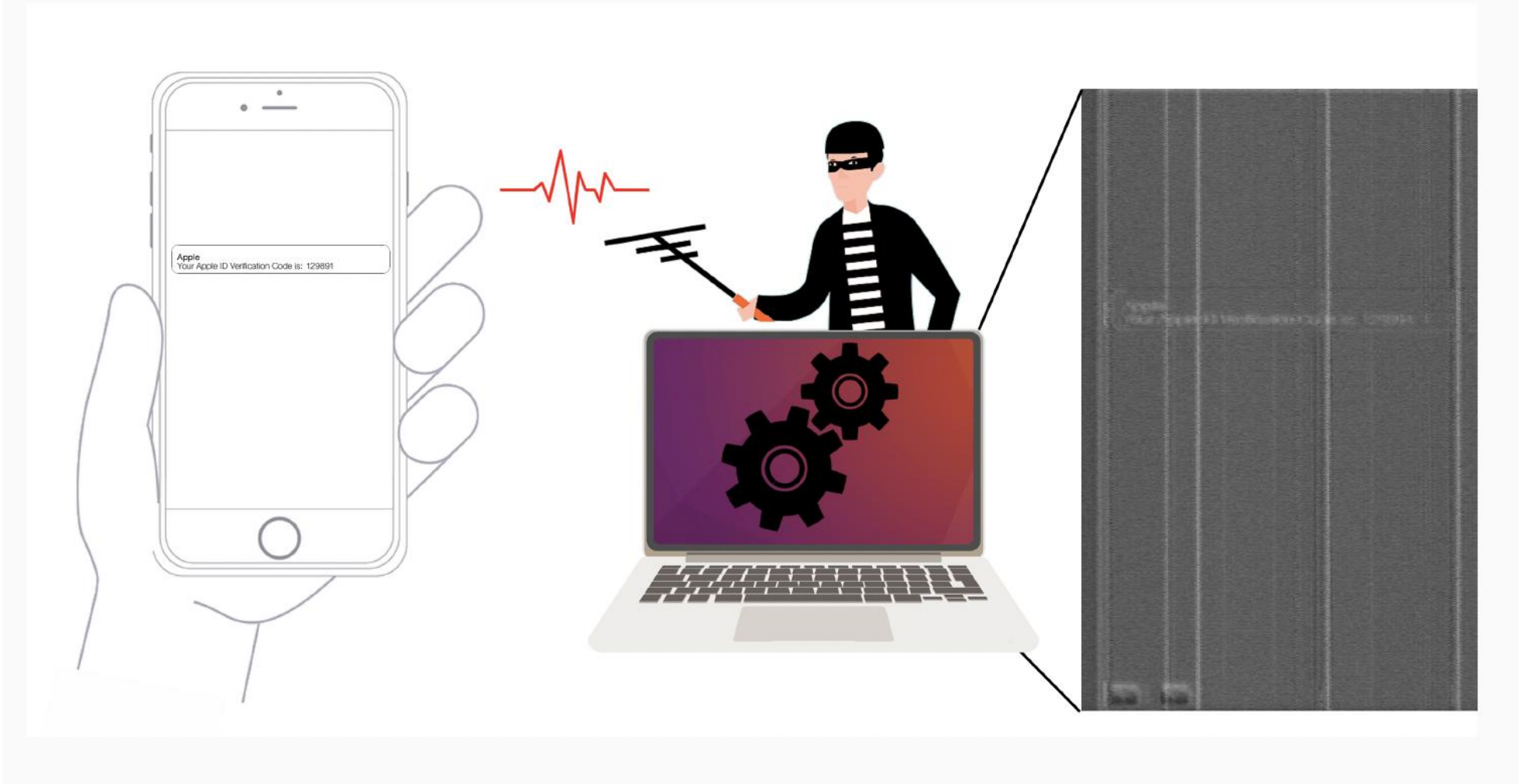
- WWII: Bell Labs noticed that messages from teleprinter communications can be recovered
- 1982: NSA published TEMPEST Fundamentals
- 1985: “Van Eck phreaking” the first unclassified analysis of the security risks of emanations from PC monitors

Source: Tomorrow’s world BBC

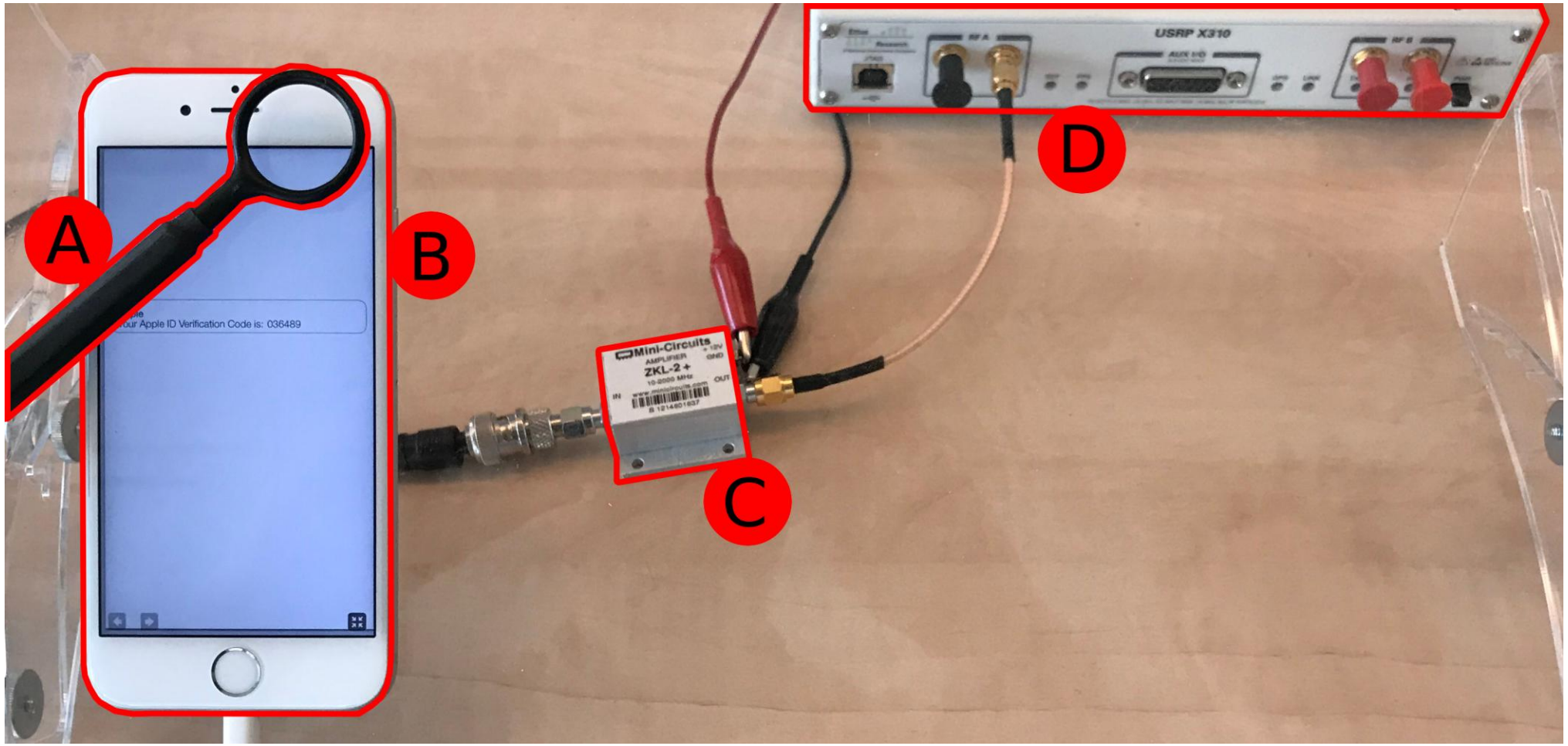
Modern TEMPEST: Screen gleaning on mobile phones



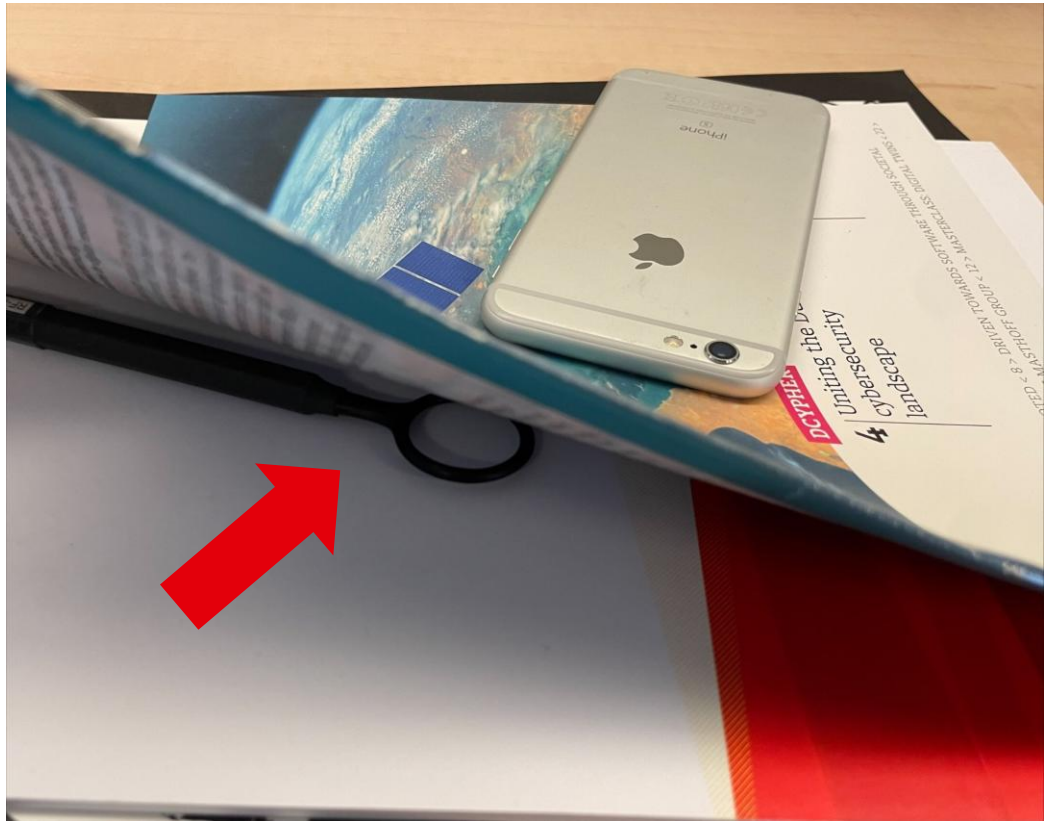
Screen gleaning in practice



Screen gleaning setup

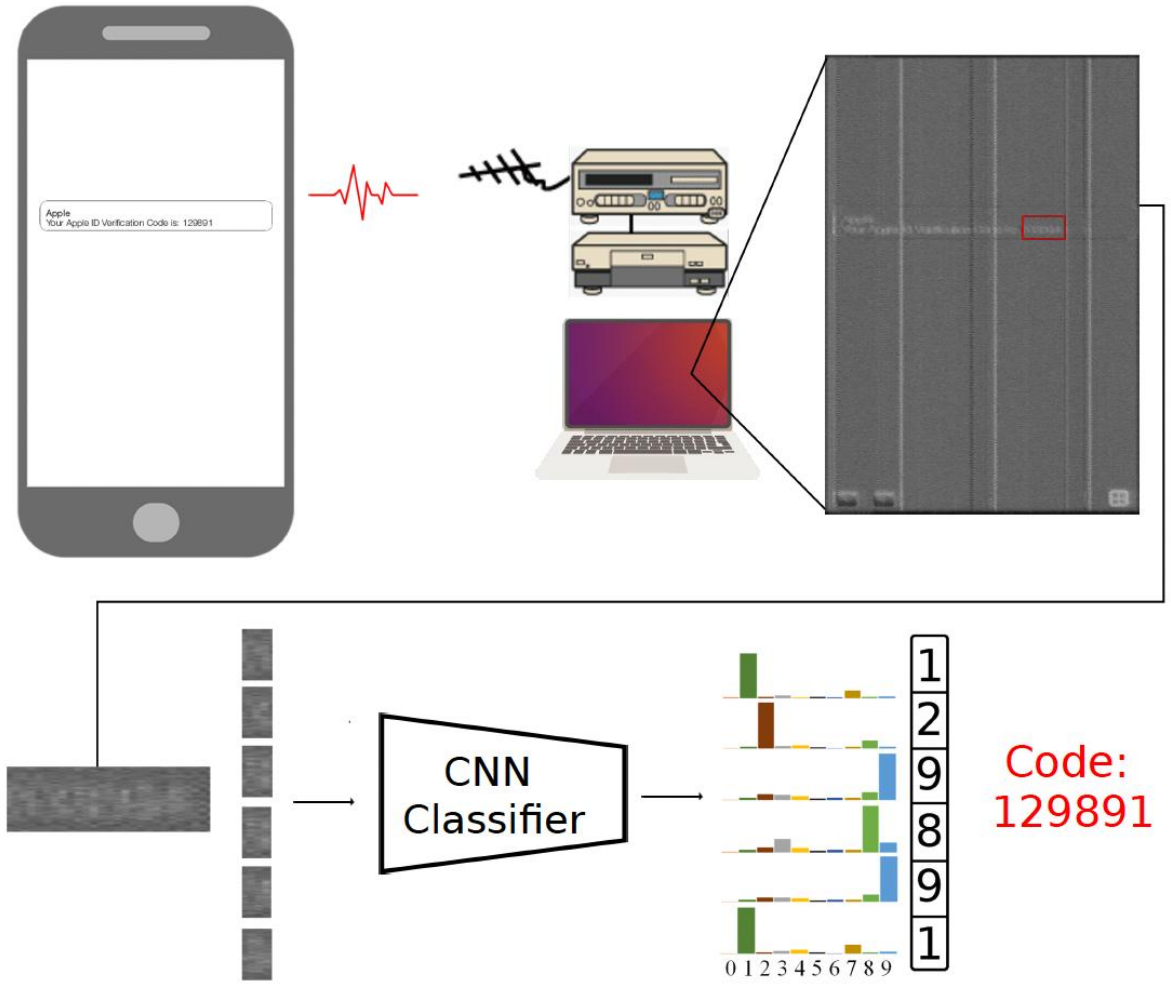


Motivating story and assumptions



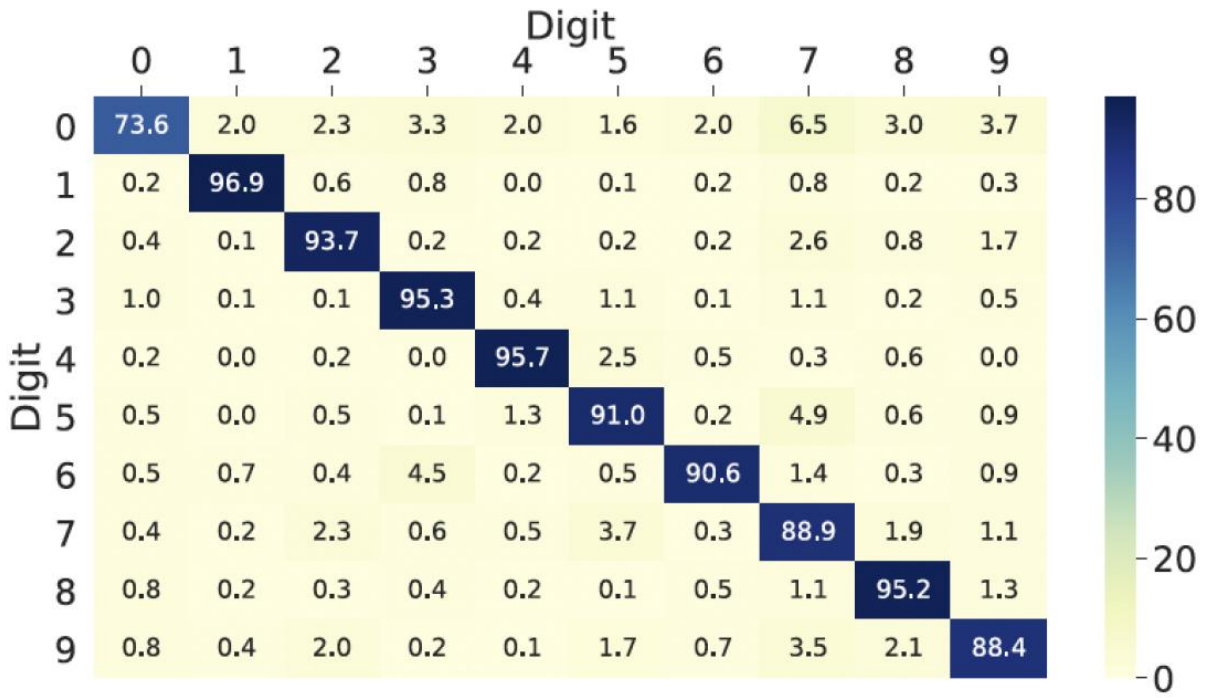
- The attacker has access to a profiling device that is “similar” to the target device
- The set of symbols displayed on the phone are digits only
- The attacker can collect EM traces from the target as representation of the image displayed on the screen

The attack model



- The target emits EM signal intercepted by antenna connected to SDR
- The leaked information is collected and reconstructed as a gray-scale image (emage)
- From emage, the 6-digit security code is cropped and fed into a CNN classifier

Security code results



	6 digits	**$\geq$ 5 digits**	**$\geq$ 4 digits**
Acc. (%)	50.5	89.5	99.0

More results

- Attack on different phones of the same model e.g. iPhone 6
  - ❖ Cross-device accuracy of 61.5% when the classifier is trained and tested on two different phones
- Attack on a different phone
  - ❖ Accuracy of 74% on Huawei Honor 6X
- Attack at a greater distance (through a magazine)
  - ❖ Accuracy of 65.8% on Huawei Honor 6X through 200 pages

Z. Liu, Niels Samwel, L. Weissbart, Z. Zhao, D. Lauret, L. Batina, M. Larson, Screen Gleaning: A Screen Reading TEMPEST Attack on Mobile Devices Exploiting an Electromagnetic Side Channel, NDSS 2021.

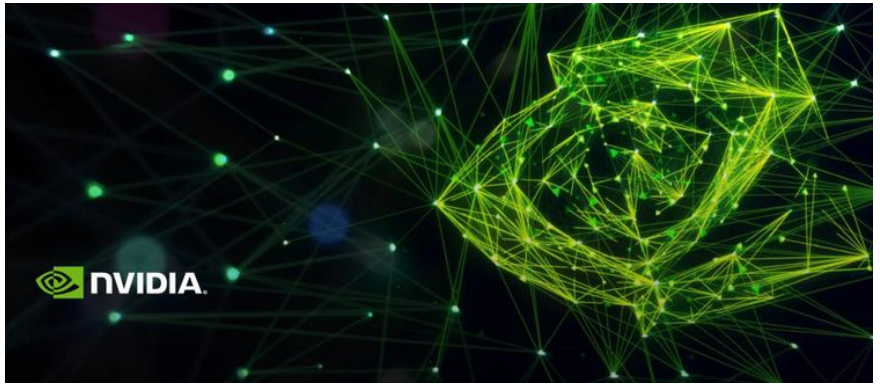
CSI NN: REVERSE ENGINEERING OF NEURAL NETWORK ARCHITECTURES THROUGH ELECTROMAGNETIC SIDE CHANNEL

L. BATINA, S. BHASIN, D. JAP, S. PICEK

Ex. 2: Rev engineering of NNs

BARRACUDA: GPUS DO LEAK DNN WEIGHTS

P. HORVATH, L. CHMIELEWSKI, L. WEISSBART, L. BATINA, Y. YAROM



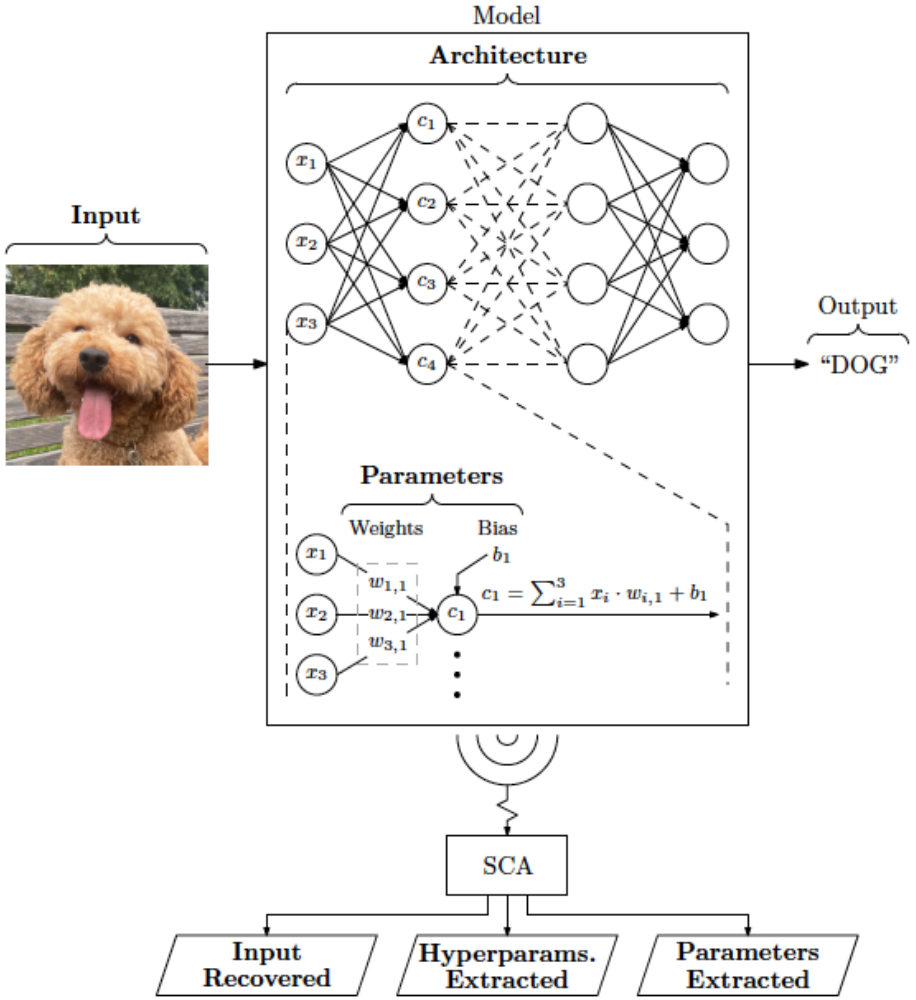
Everything Everywhere embedded AI

Motivation

- Well-trained models are valuable for some industries
- Neural nets are being deployed on various platforms, some of those are constrained
- This makes the NN architectures and their parameters target for adversaries

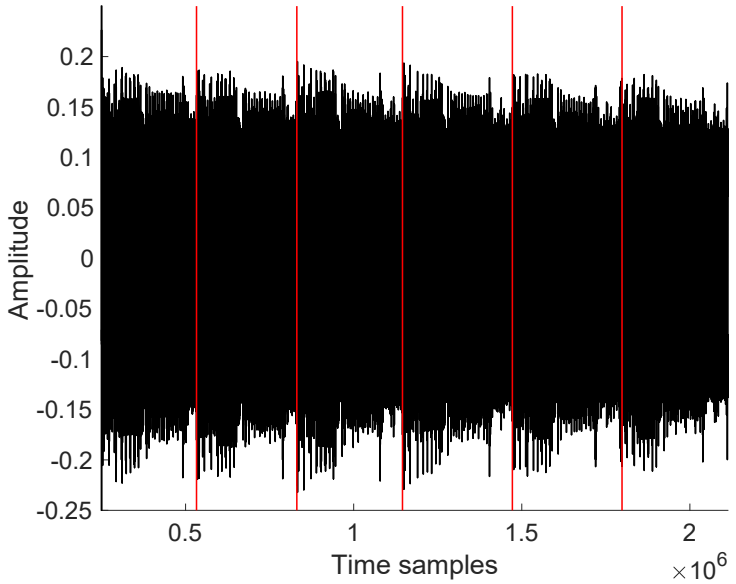
S. Picek and L. Batina: Cybersecurity and Artificial Intelligence: A Perfect Couple? DATA, CYBERSECURITY & PRIVACY 03-2024 #15

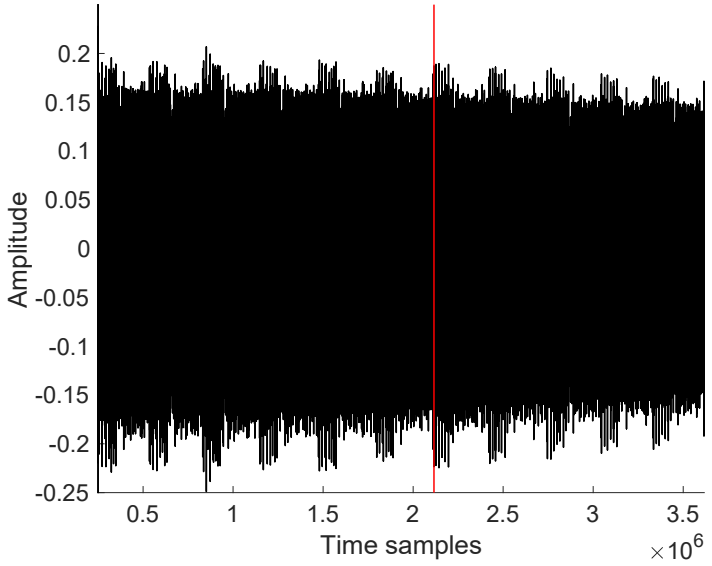
New field: Physical SCA on neural nets



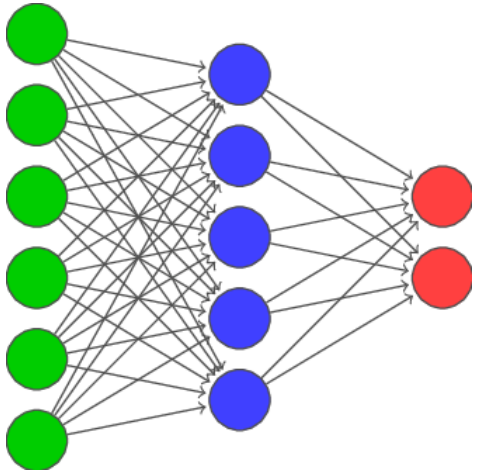
- Main components of interest for adversaries:
  - architecture and hyperparameters, including activation function
  - trained parameters
  - input (image or text)

Simple EMA on neurons

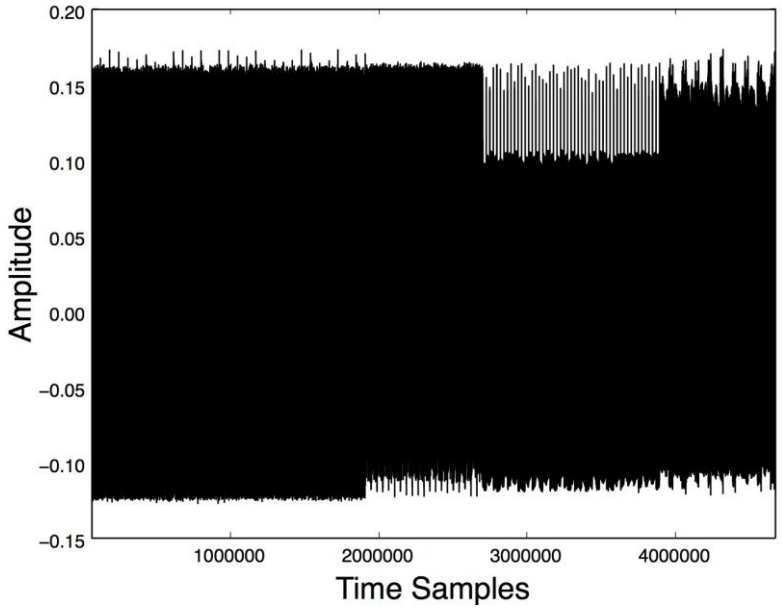




- 6 neurons = 6 repeating patterns
- Each neuron executes a number of multiplications, followed by activation f.

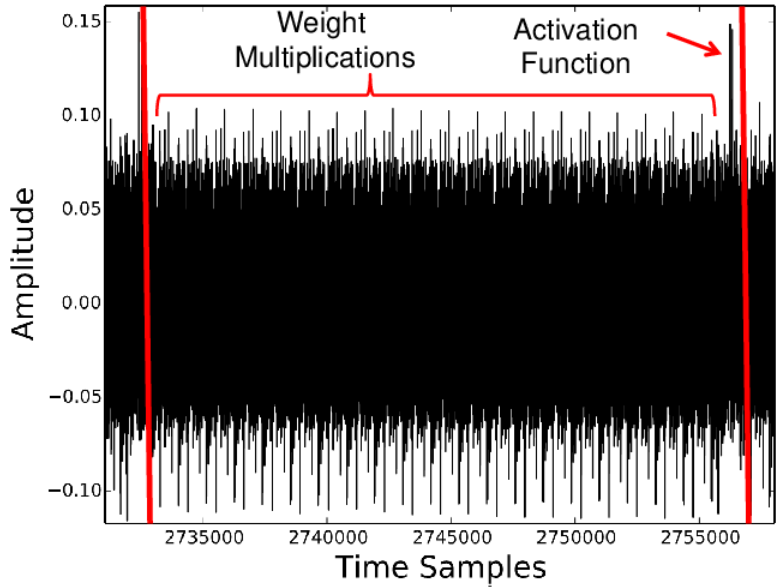


Results for MLP on ARM Cortex-M3



Four hidden layers (50, 30, 20, 50)

Zoom in one neuron in 3rd layer



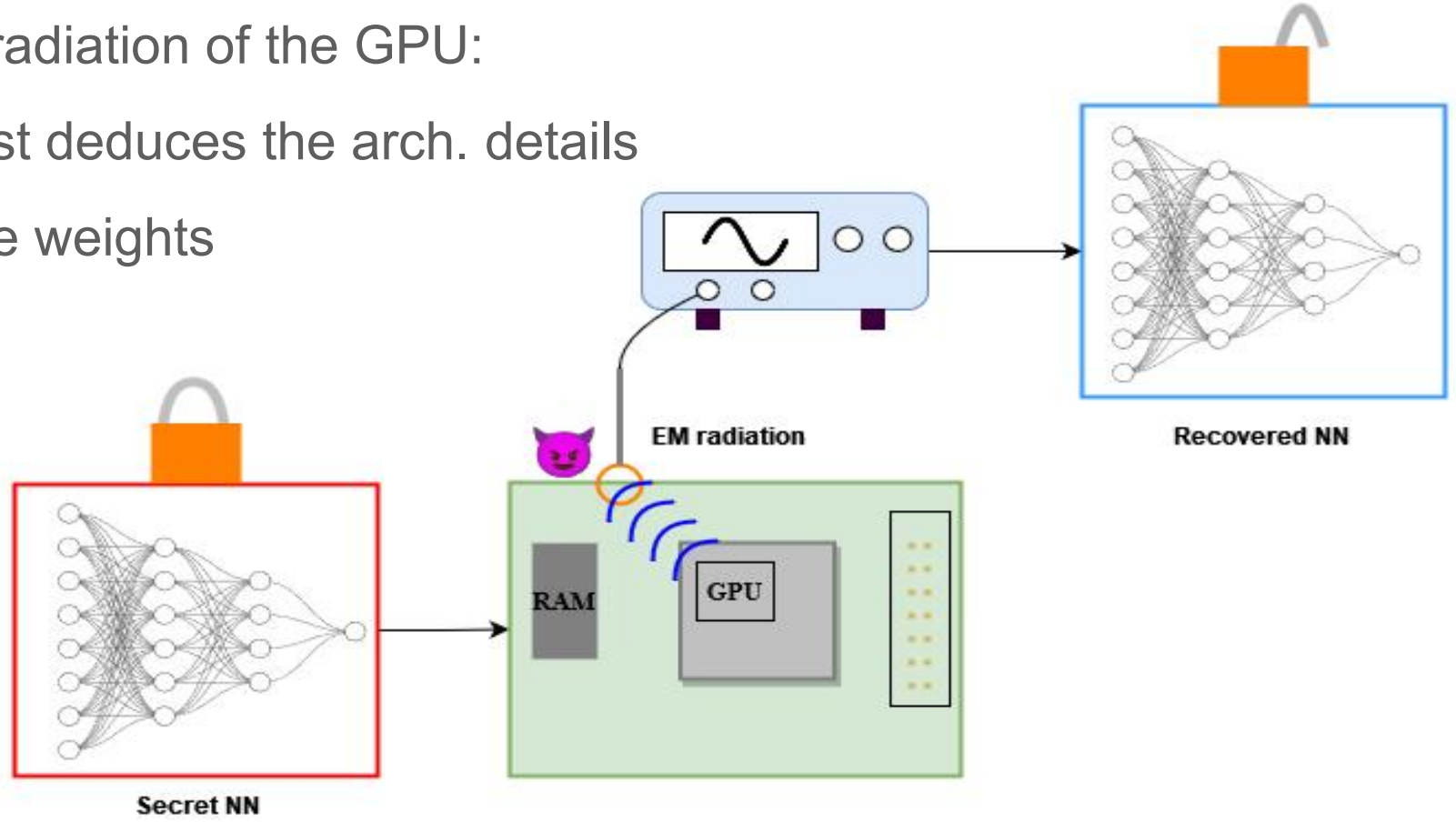
One neuron in 3rd hidden layer: 20 multiplications and 1 ReLU

With MNIST: Accuracy 98.16% (original) vs 98.15% (reverse engineered)
Average weight error: 0.0025.

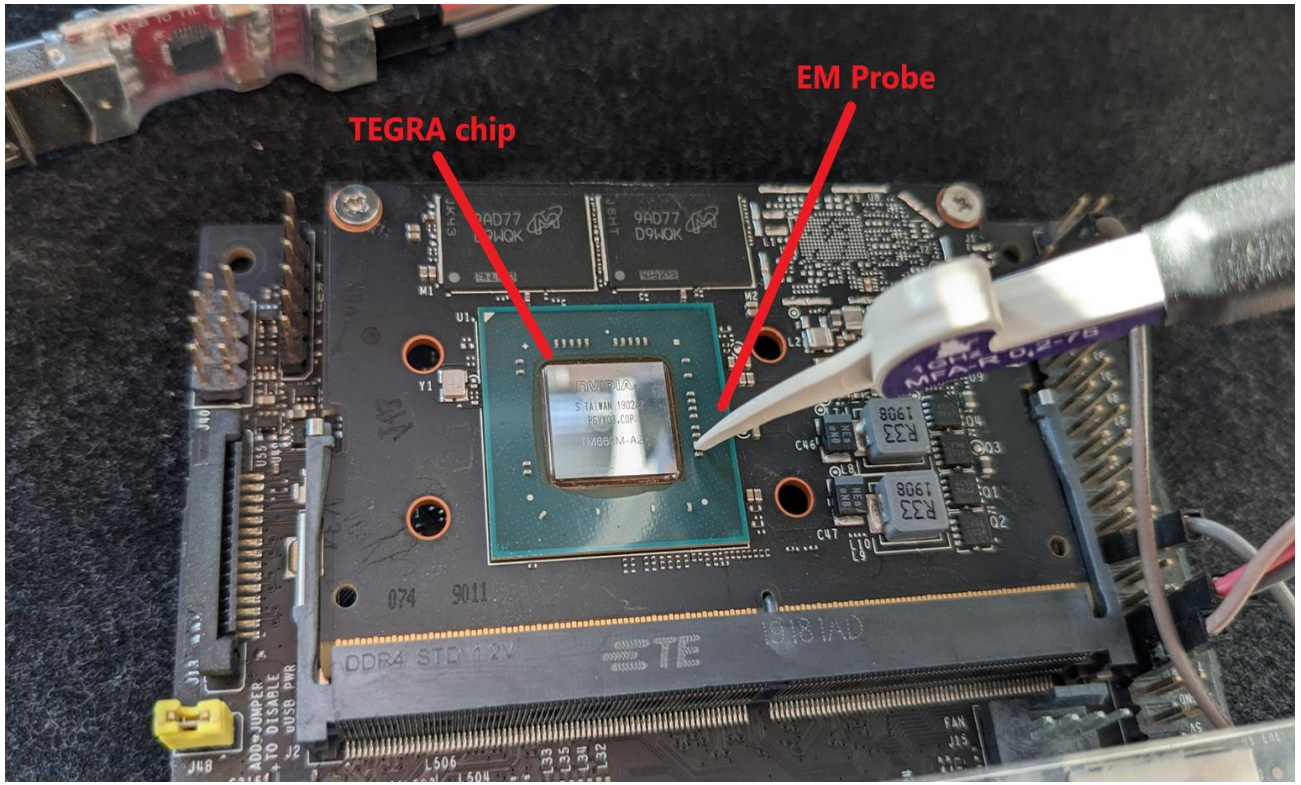
Lejla Batina, Shivam Bhasin, Dirmanto Jap, Stjepan Picek: CSI NN: Reverse Engineering of Neural Network Architectures Through Electromagnetic Side Channel. USENIX Security Symposium 2019: 515-532.

New attack BarraCUDA: Attacking Nvidia GPUs

- Goal: recover the trade secrets encoded in NNs parameters from an ML model
- By monitoring the EM radiation of the GPU:
  - the adversary first deduces the arch. details
  - then recovers the weights

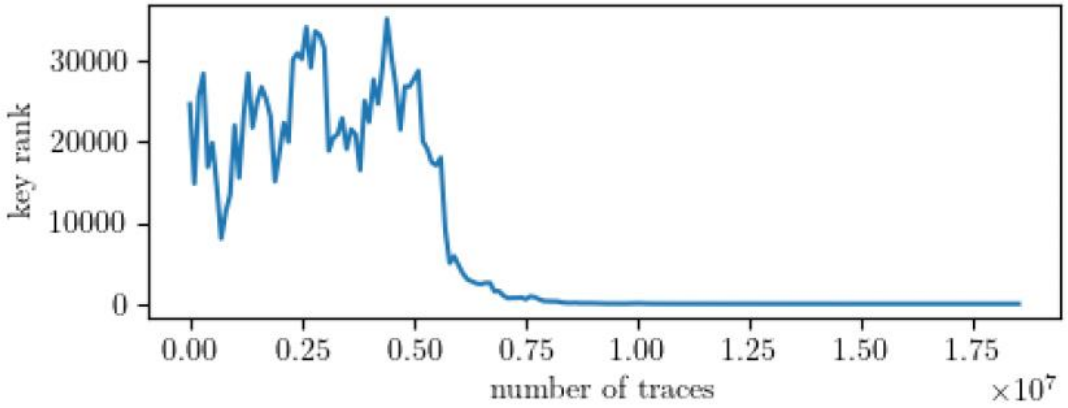


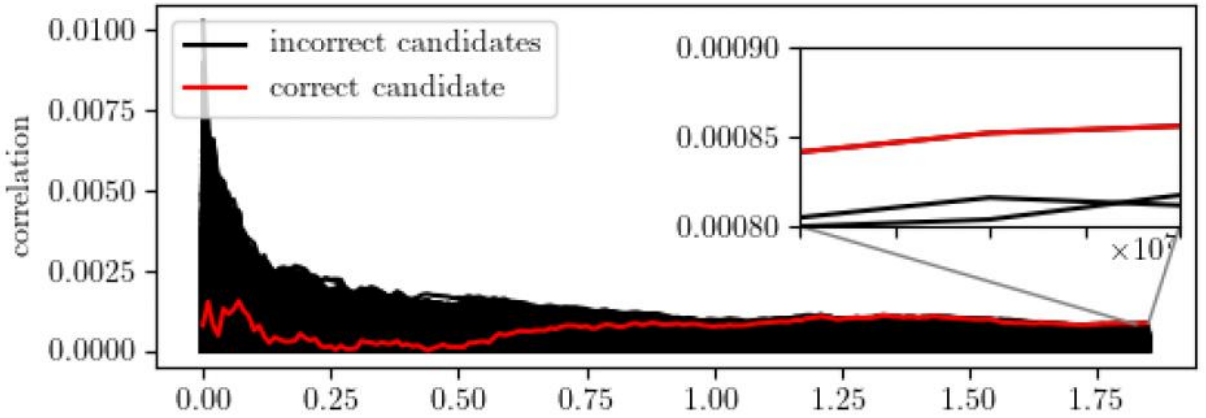
Setup and the threat model



- Attacker learned the architecture details by using SCA techniques
- Attacker has a physical access to the device and can monitor EM during the inference for known inputs

Results for weights recovery





Key rank and correlations of the 9th weight in the first layer in the real-world CNN architecture.

P. Horvath, L. Chmielewski, L. Weissbart, L. Batina, Y. Yarom. BarraCUDA: GPUs do Leak DNN Weights, USENIX Security Symposium, 2025.

BarraCUDA: What is new

- Performed parameter extraction from the EfficientNet model running on an industry-strength Jetson Nano device
- Reverse engineering the closed source TensorRT library
- Attack has a large complexity, which required developing a special CUDA-based attack implementation
- The attack on Jetson Nano requires 11-12 days (traces collection and alignment), and the parameters are recovered in 6 min per weight, but the results improved a lot on Jetson Orin

AI and Security: Where do we go from here?

- AI can improve side-channel and fault attacks on embedded devices
- Security and privacy vulnerabilities when implementing AI should be considered
- Other topics:
  - Security and privacy of AI
  - Malware recognition
  - Fuzzing
  - Network intrusion detection

CESCA group @Radboud University

https://cescalab.cs.ru.nl/

Thank You!