Hypothesis testing for leakage assessment in side channel analysis

Making decisions is easy, making the right decision less so

Ileana Buhan, June 2023

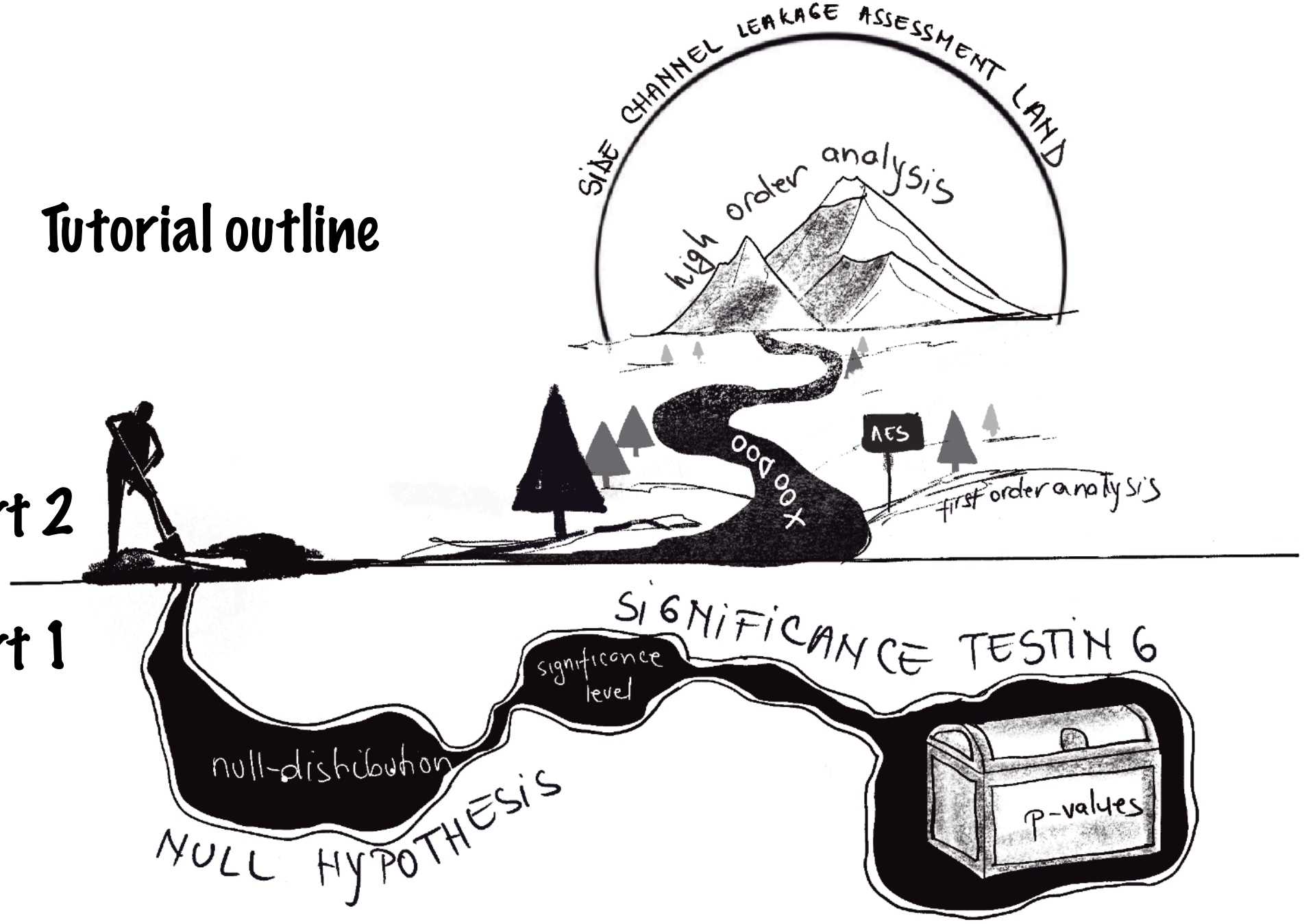
@ileanabuhan

Radboud University

Tutorial outline

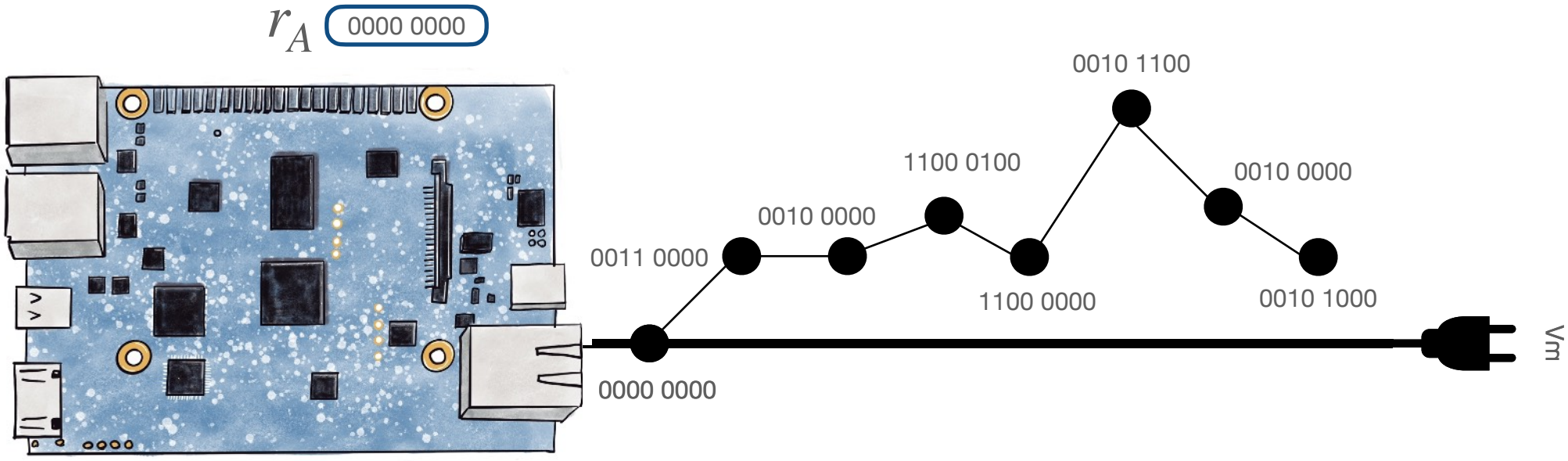
Part 2

Part 1



TVLA on real traces

What is leakage?



Leakage is the dependence between power consumption and the sensitive data.

Leakage detection in action

You are a developer who wants to ensure that your implementation does not leak, but not perform full attack. Ideas?

- Hint 1: we know that any dependency between the measured side-channel and the sensitive data is a potential side channel vulnerability;
- Hint 2: using the reverse logic, if there are no dependencies, there is no side-channel vulnerability;

Can we check vulnerability to side-channels without doing an attack?

- yes! measure the side channel for di erent input values and see if they are different;
- complicating fact: side channel measurements are influenced by many factors, not always straightforward;

Leakage detection in action

Test Vector Leakage Detection (TVLA) most popular leakage detection test.

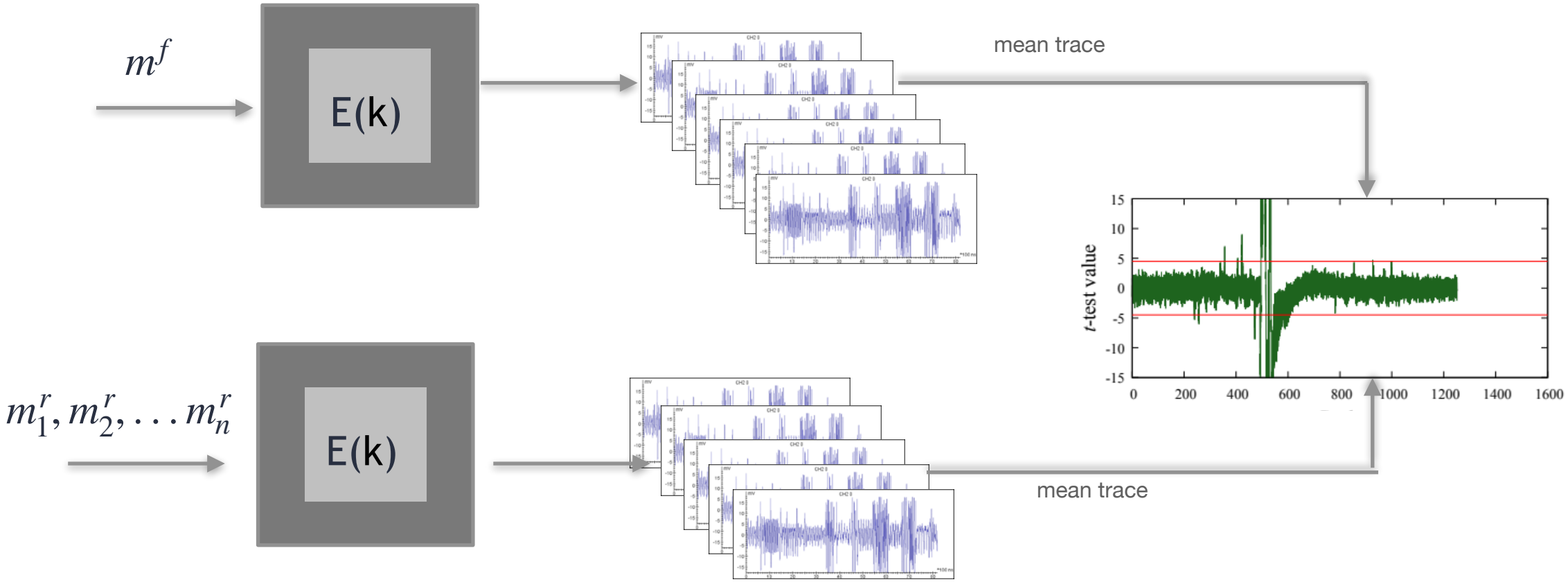
- Non-specific or general test: aims to detect any leakage that depends on input data (or key);

a.k.a fixed - vs - random;

- Specific-test: targets a specific intermediate value of the cryptographic algorithm that could be exploited to recover keys or other sensitive information.

a.k.a fixed - vs - fixed;

Collecting data for TVLA

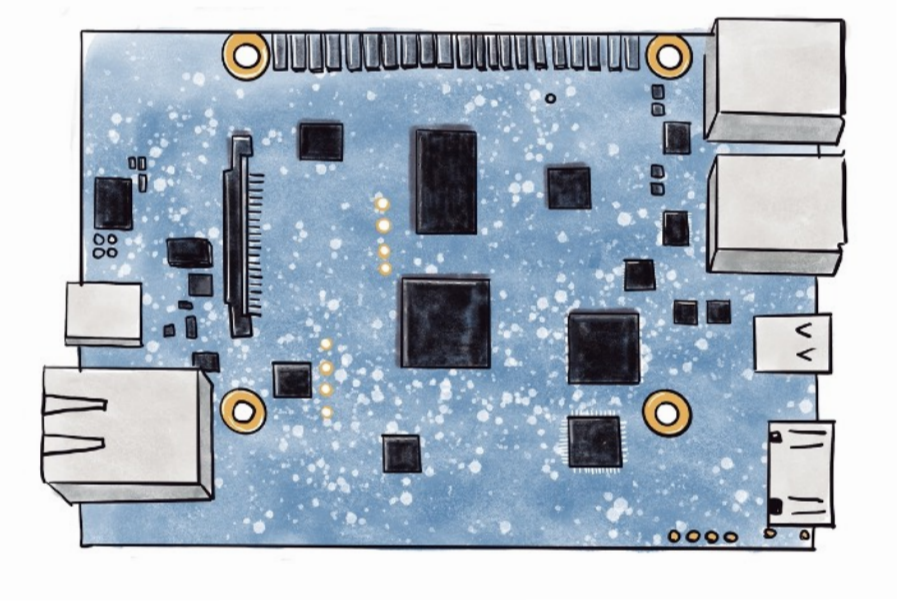


Suggested reading:

A testing methodology for sidechannel resistance validation Gilbert Goodwill, Benjamin Jun, Josh Jaffe, Pankaj Rohatgi: Cryptography Research Inc.

https://csrc.nist.gov/csrc/media/events/non-invasive-attack-testing-workshop/documents/08_goodwill.pdf

TVLA - two-sample t-test



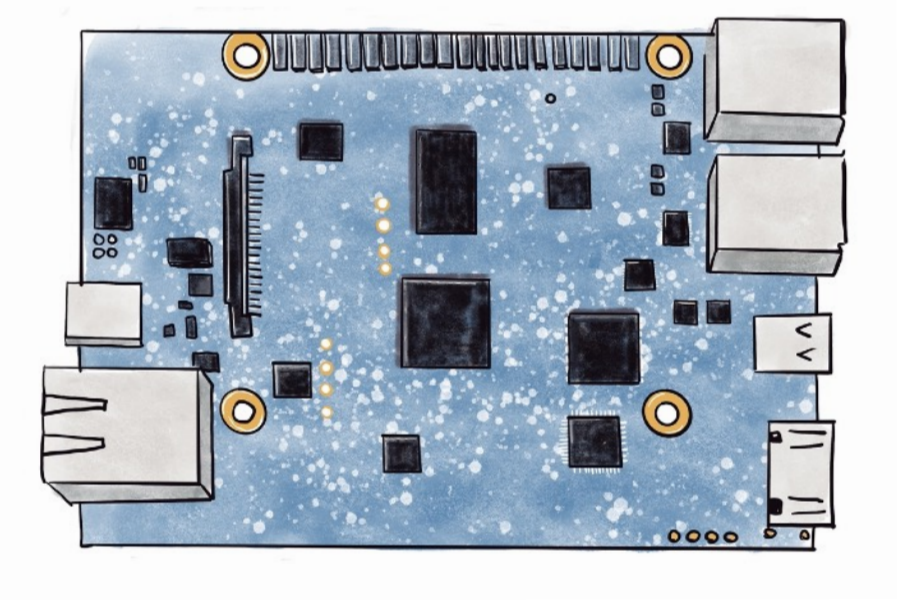
H_0: device is NOT guilty of leaking information

$$\mu_{fixed} = \mu_{random}$$

H_a: device is guilty of leaking information

$$\mu_{fixed} \neq \mu_{random}$$

Selecting the significance level

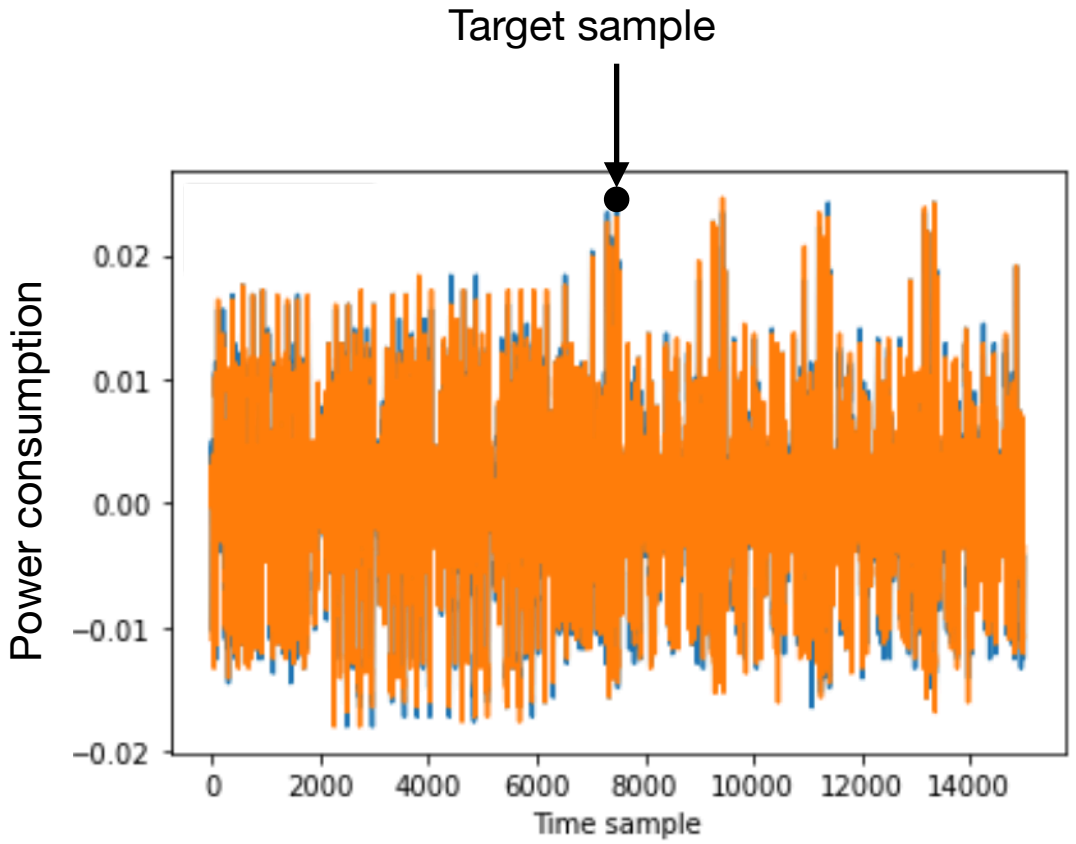


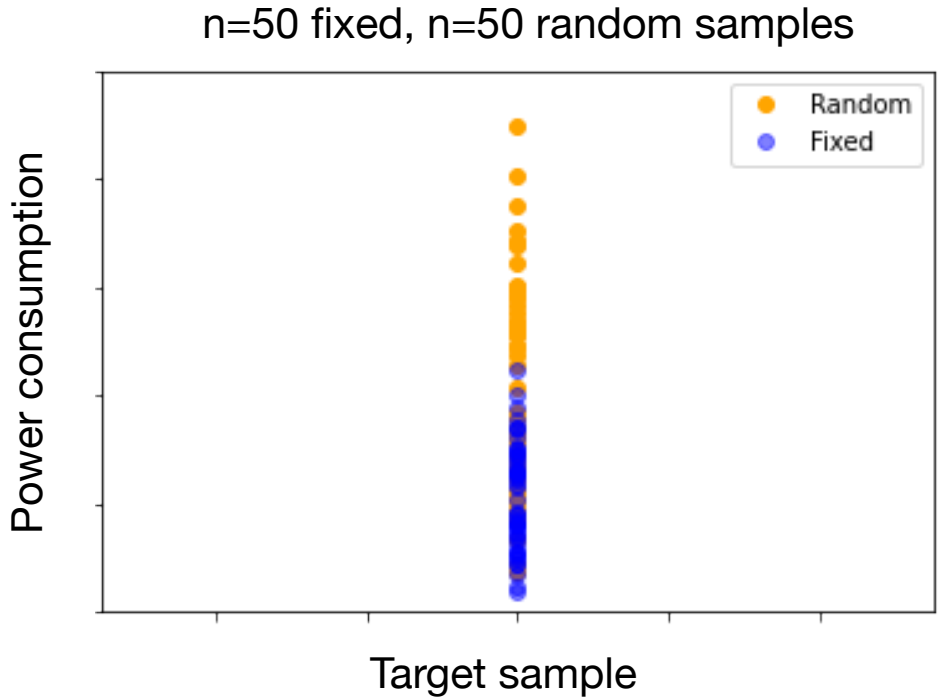
H_0: device is NOT guilty of leaking information

$$\mu_{fixed} = \mu_{random}$$

The significance level (α) is the probability at which we are prepared to reject the null hypothesis and conclude that the effect is statistically significant.

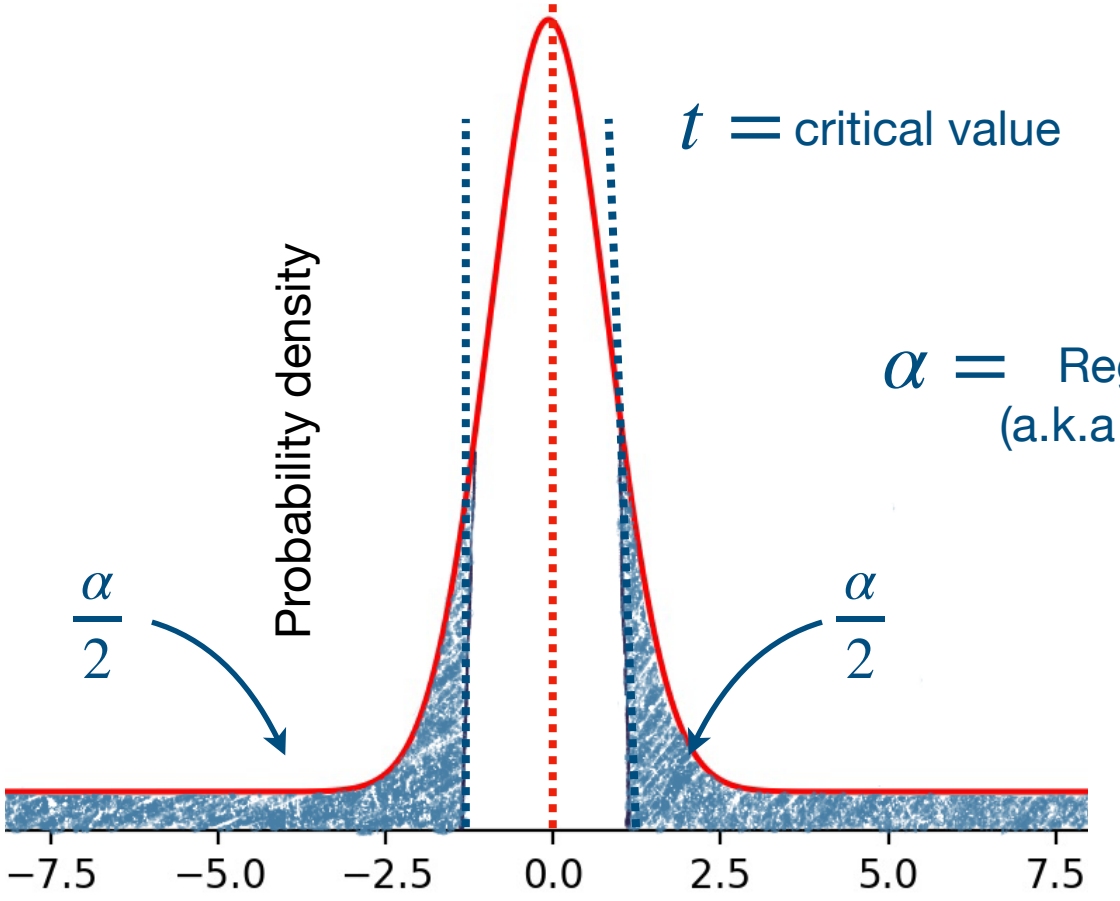
TVLA - two-sample t-test





First order t-test, we analyze each sample independently

Why the 4.5 value?



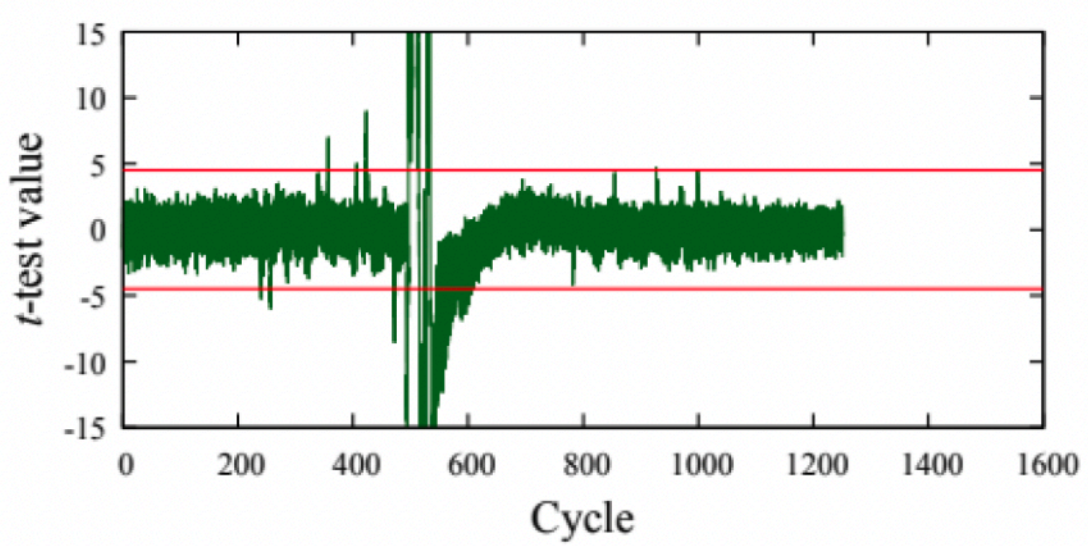
t = critical value

α = Region of rejection
(a.k.a significance level)

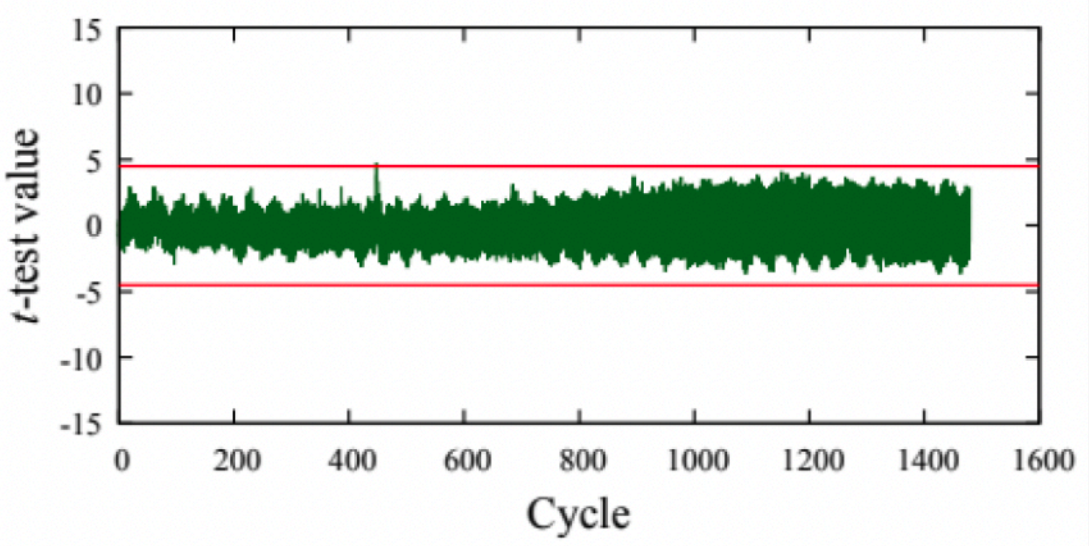
— T-distribution for $\mu_r = \mu_f$

$\mu_F - \mu_R$

What is the magic number 4.5?



(a) AES original implementation.

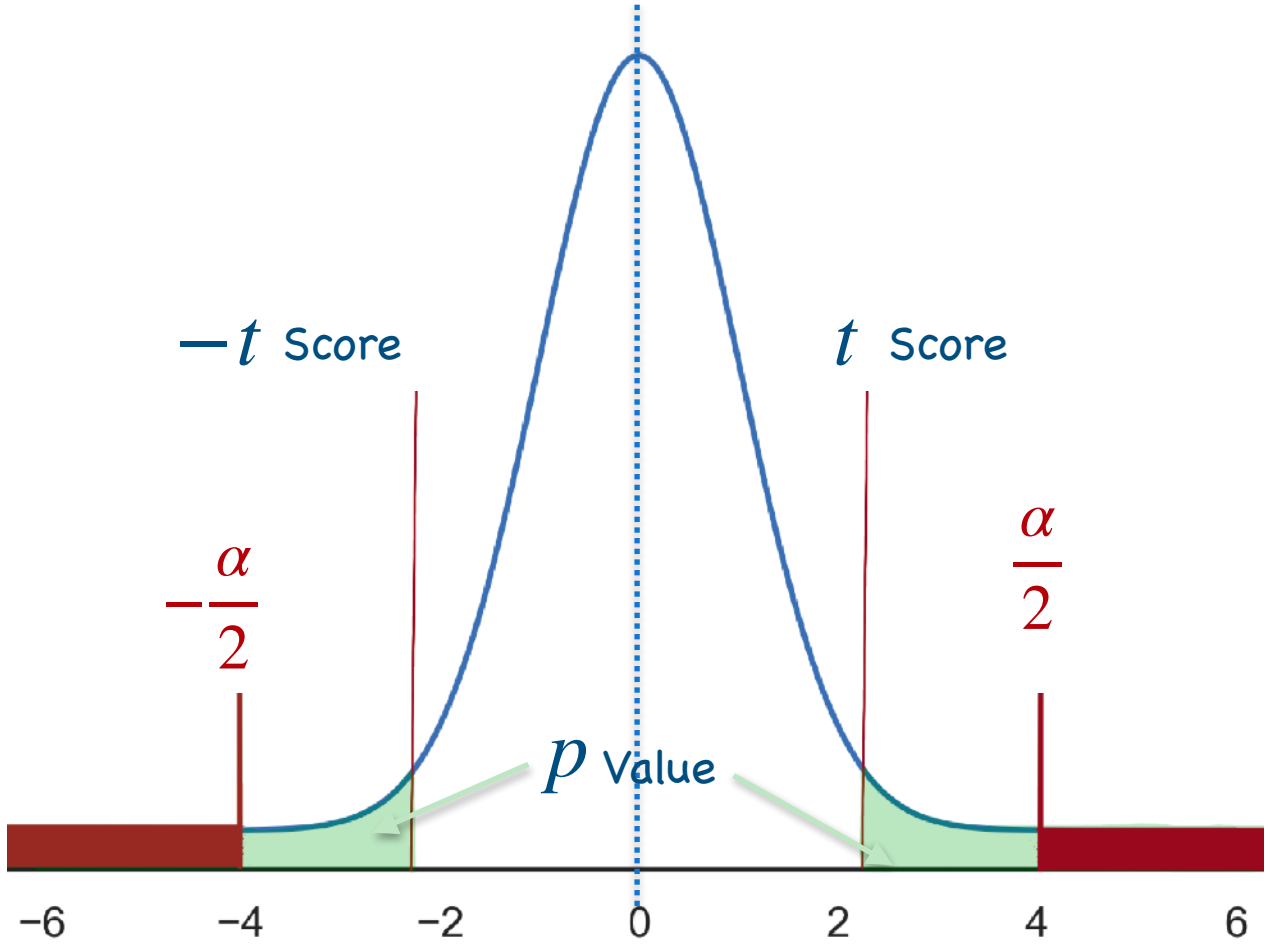


(d) AES fixed with ROSITA.

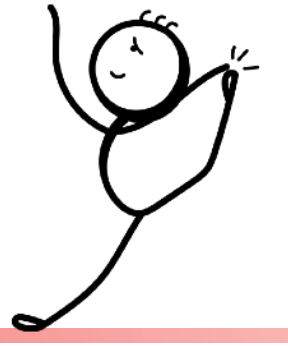
Source for the figure: Madura A Shelton and Niels Samwel and Lejla Batina and Francesco Regazzoni and Markus Wagner and Yuval Yarom *Rosita: Towards Automatic Elimination of Power-Analysis Leakage in Ciphers*, NDSS 2021

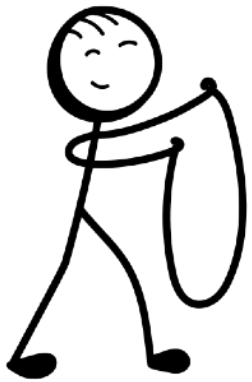
What is the magic number 4.5?

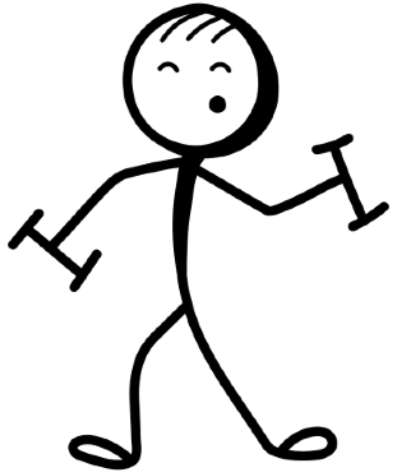
$\alpha = 0.001$

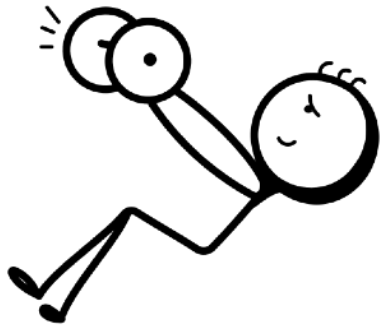


Exercise 4

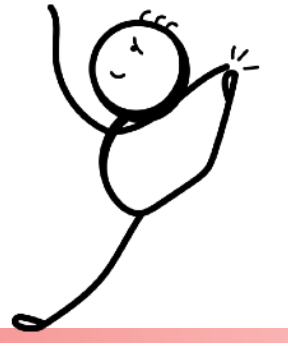


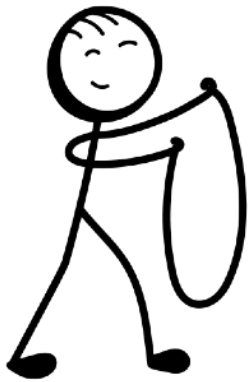


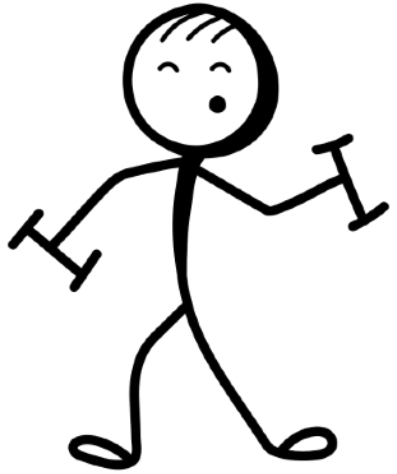


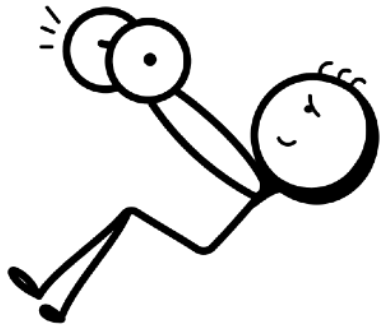


Exercise 5









Notes

IMPORTANT: TVLA test is qualitative measure of leakage, and NOT a quantitative measure.

If we are dealing with a high-order implementation, we always need to check if lower orders leak, there might be surprizes;

Final Notes

A lack of evidence to support the guilty verdict, does not mean the device is "innocent"; We say: "We fail to reject H0" and NOT "we accept H0"

Alternatively we say:

"The evidence supports the decision to reject H0 at significance level α".

Final Notes

A lack of evidence to support the guilty verdict, does not mean the device is "innocent"; We say: "We fail to reject H0" and NOT "we accept H0"

Alternatively we say:

"The evidence supports the decision to reject H0 at significance level α".

Why could TVLA to fail?

- Sample size - too small
- Effect size (the difference between the two means) is too small, because:
  - wrong fixed input;
  - too much noise (variance) in the sample data;
- Bad luck: statistical tests are probabilistic

Recommended reading

Carolyn Whitnall, Elisabeth Oswald:
A Critical Analysis of ISO 17825 ('Testing Methods for the Mitigation of Non-invasive Attack Classes Against Cryptographic Modules'). ASIACRYPT (3) 2019: 256-284

François-Xavier Standaert:
How (Not) to Use Welch's T-Test in Side-Channel Security Evaluations. CARDIS 2018: 65-79

Tobias Schneider, Amir Moradi:
Leakage Assessment Methodology - A Clear Roadmap for Side-Channel Evaluations. CHES 2015: 495-513

http://reassure.eu/leakage-detection-tutorial/

THANK YOU