Cryptography and Embedded System Security
CRAESS_I

Xiaolu Hou

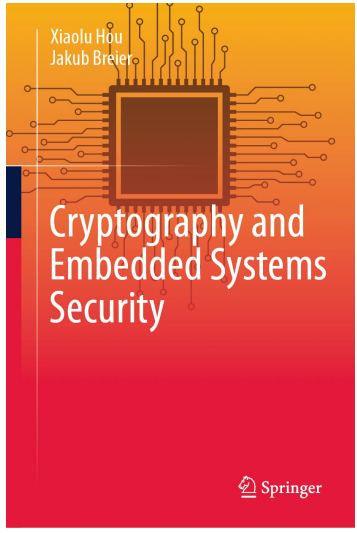
FIIT, STU
xiaolu.hou @ stuba.sk

Course Outline

- Abstract algebra and number theory
- Introduction to cryptography
- Symmetric block ciphers and their implementations
- RSA, RSA signatures, and their implementations
- Probability theory and introduction to SCA
- SPA and non-profiled DPA
- Profiled DPA
- SCA countermeasures
- FA on RSA and countermeasures
- FA on symmetric block ciphers
- FA countermeasures for symmetric block cipher
- Practical aspects of physical attacks
  - Invited speaker: Dr. Jakub Breier, Senior security manager, TTControl GmbH

Recommended reading

- Textbook
  - Sections
    - 1.7;
    - 4.1;
    - 4.2.1, 4.2.2, 4.2.3.



Lecture Outline

- Probability Measure
- Random Variable
- Side-channel Analysis
- Side-channel Leakages
- Leakage Assessment

Probability theory and introduction to SCA

- Probability Measure
- Random Variable
- Side-channel Analysis
- Side-channel Leakages
- Leakage Assessment

Random Experiments

- Probability theory studies the mathematical theory behind random experiments.
- A random experiment is an experiment whose output cannot be predicted with certainty in advance.
- However, if the experiment is repeated many times, we can see "regularity" in the average output.
- For example, if we roll a die, we cannot predict the output of one roll.
- But if we roll it many times, we would expect to see the number 1 in $1/6$ of the outcomes assuming the die is fair.

Sample Space and Events

- For a given random experiment, we define *sample space*, denoted by Ω, to be the set of all possible outcomes.
- A subset A of Ω is called an *event*.
- If the outcome of the experiment is contained in A, then we say that A has *occurred*.
- The empty set $\emptyset$ denotes the event that consists of no outcomes.
- $\emptyset$ is also called the *impossible event*.

Example

- When the random experiment is rolling a die, the sample space $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$. $A = \{ 1, 2, 3 \} \subseteq \Omega$ is an event.
- When the random experiment is rolling two dice, $\Omega = \{ (i, j) \mid 1 \leq i, j \leq 6 \}$. One possible event is $A = \{ (1, 2), (1, 1) \}$.

Events

- Recall that we have defined complement, unions, and intersections between sets in the first week.
- Fix a sample space Ω. Take two events, A and B.
- We say that $A \cup B$ occurs if either A or B occurs.
- Similarly, $\bigcup_{i=1}^{n} A_i$ occurs when at least one A_i occurs.
- And we say $A \bigcap B$ occurs if both A and B occur, $\bigcap_{i=1}^{m} A_i$ occurs if all of the events A_i occur.
- If $A \cap B = \emptyset$, then A and B cannot both occur, they are called *mutually exclusive*.
- The complement of A, A^c, contains events in Ω, but not in A.

Sample space and its power set

- Ω: sample space
- $\mathcal{A}$: power set of Ω, 2^{Ω}

Example

Let us consider the random experiment of tossing a coin, the sample space $\Omega = \{ H, T \}$. $\mathcal{A} = 2^{\Omega} = \{ \emptyset, \Omega, \{ H \}, \{ T \} \}$.

Probability

Definition

A *probability measure* defined on $(\Omega, \mathcal{A})$ is a function $P : \mathcal{A} \to [0, 1]$ such that

- $P(\Omega) = 1$, $P(\emptyset) = 0$.
- For any $A_1, A_2, \ldots \in \mathcal{A}$ that are pairwise disjoint, i.e. $A_{i_1} \cap A_{i_2} = \emptyset$ for $i_1 \neq i_2$, *countable additivity*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

$P(A)$ is called the *probability* of A.

Example

Tossing a coin, $\Omega = \{ H, T \}$. $\mathcal{A} = 2^{\Omega} = \{ \emptyset, \Omega, \{ H \}, \{ T \} \}$. P defined as follows is a probability measure on $(\Omega, \mathcal{A})$:

$$P(\emptyset) = 0, \quad P(\Omega) = 1, \quad P(\{ H \}) = \frac{1}{2}, \quad P(\{ T \}) = \frac{1}{2}.$$

Uniform probability

Definition

Let Ω be a set with finite cardinality. $\mathcal{A} = 2^{\Omega}$. A probability measure P on $(\Omega, \mathcal{A})$ is called *uniform* if

$$P(\{\,\omega\,\}) = \frac{1}{|\Omega|},\ \forall \omega \in \Omega.$$

We note that if P is a uniform probability measure on $(\Omega, \mathcal{A})$, then for any $A \in \mathcal{A}$, $P(A) = \frac{|A|}{|\Omega|}$.

Example

Let $\Omega = \{\,1, 2, 3, 4, 5, 6\,\}$ and $\mathcal{A} = 2^{\Omega}$. The uniform probability measure on $(\Omega, \mathcal{A})$ is given by P such that

$$P(\{\,i\,\}) = \frac{1}{6}$$

for $i \in \Omega$. Let $A = \{\,1, 2, 3\,\},\ B = \{\,2, 4\,\}$, then

$$P(A) = 1/2, \quad P(B) = 1/3.$$

Probability measure on a countable set

Example

Let Ω be a countable set (finite or countably infinite). Let $\mathcal{A} = 2^{\Omega}$. Then, any probability measure on $(\Omega, \mathcal{A})$ is a function such that for any $A \in \mathcal{A}$,

$$P(A) = \sum_{\omega \in A} P(\{\, \omega \,\}), \text{ where } P(\{\, \omega \,\}) \geq 0 \text{ and } \sum_{\omega \in \Omega} P(\{\, \omega \,\}) = 1.$$

Conditional probability

- Take any $A, B \in \mathcal{A}$ such that $P(B) > 0$.
- We would like to compute the probability of A occurring given the knowledge that B has occurred.
- We do not need to consider $A \cap B^c$ since B has already occurred.
- Instead, we look at $A \cap B$, which occurs when both A and B occur.
- This leads to the definition of the *conditional probability of A given B*:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Example

Let $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$ and $\mathcal{A} = 2^{\Omega}$. The uniform probability measure on $(\Omega, \mathcal{A})$ is given by P such that

$$P(\{ i \}) = 1/6$$

for $i \in \Omega$. $A = \{ 1, 2, 3 \}$, $B = \{ 2, 4 \}$. Then

$$A \cap B = ? \quad P(A \cap B) = ? \quad P(A|B) = ?$$

Conditional probability

- Take any $A, B \in \mathcal{A}$ such that $P(B) > 0$.
- We would like to compute the probability of A occurring given the knowledge that B has occurred.
- We do not need to consider $A \cap B^c$ since B has already occurred.
- Instead, we look at $A \cap B$, which occurs when both A and B occur.
- This leads to the definition of the *conditional probability of A given B*:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Example

$A = \{ 1, 2, 3 \}, \ B = \{ 2, 4 \}$

$$A \cap B = \{ 2 \}, \quad P(A \cap B) = \frac{1}{6}, \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/3} = \frac{1}{2}$$

Independent events

Definition

Two events A, B are said to be *independent* if $P(A \cap B) = P(A)P(B)$. Otherwise, we say that they are *dependent*.

When $P(B) > 0$, the condition $P(A \cap B) = P(A)P(B)$ is equivalent to

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

That is, the probability of A occurring given the knowledge that B has occurred is the same as the probability of A occurring without the knowledge that B has occurred.

Independent events

Example

Let $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$ and $\mathcal{A} = 2^{\Omega}$. The uniform probability measure on $(\Omega, \mathcal{A})$ is given by P such that

$$P(\{ i \}) = 1/6,$$

for $i \in \Omega$.

$$A = \{ 1, 2, 3 \}, \quad B = \{ 2, 4 \}, \quad P(A) = 1/2, \quad P(B) = 1/3,$$

$$A \cap B = \{ 2 \}, \quad P(A \cap B) = \frac{1}{6}, \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/3} = \frac{1}{2}$$

$$P(A)P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}.$$

By definition, A and B are independent. We also note that

$$P(A|B) = P(A) = 1/2.$$

Bayes' Theorem

Theorem (Bayes' Theorem)

If $P(A) > 0$ and $P(B) > 0$, then

$$P(B)P(A|B) = P(A)P(B|A).$$

Proof.

Since

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

we have

$$P(B)P(A|B) = P(A \cap B), \quad P(A)P(B|A) = P(A \cap B).$$

□

Partition of Ω

Definition

A set of events $\{ E_1, E_2, \dots \mid E_i \in \mathcal{A} \}$, is called a *partition of* Ω if they are pairwise disjoint, $P(E_i) > 0$ for all i, and $\cup_i E_i = \Omega$. If the set of events is finite, it is called a *finite partition of* Ω, otherwise, it is called a *countable partition of* Ω.

Example

Let $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$, $\mathcal{A} = 2^{\Omega}$, and P be the uniform probability measure on $(\Omega, \mathcal{A})$. Let $E_1 = \{ 1, 2, 3 \}$, $E_2 = \{ 4, 5 \}$, $E_3 = \{ 6 \}$. Then, $\{ E_1, E_2, E_3 \}$ is a finite partition of Ω. We can also calculate that

$$P(E_1) = \frac{1}{2}, \quad P(E_2) = \frac{1}{3}, \quad P(E_3) = \frac{1}{6}.$$

Lemma

Lemma

Let $\{ E_1, E_2, \dots \mid E_i \in \mathcal{A} \}$ *be a finite or countable partition of* Ω*. Then, for any* $A \in \mathcal{A}$*, we have*

$$P(A) = \sum_i P(A|E_i)P(E_i).$$

Example

Continue from the previous example, $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$, $\mathcal{A} = 2^{\Omega}$, P is the uniform probability measure on $(\Omega, \mathcal{A})$. $E_1 = \{ 1, 2, 3 \}$, $E_2 = \{ 4, 5 \}$, $E_3 = \{ 6 \}$. Let $A = \{ 2, 4 \}$, then

$$P(A) =? \quad A \cap E_1 =? \quad A \cap E_2 =? \quad A \cap E_3 =?$$

$$P(A|E_1) =? \quad P(A|E_2) =? \quad P(A|E_3) =? \quad \sum_{i=1}^{3} P(A|E_i)P(E_i) =?$$

Lemma

Lemma

Let $\{ E_1, E_2, \dots \mid E_i \in \mathcal{A} \}$ *be a finite or countable partition of* Ω*. Then, for any* $A \in \mathcal{A}$*, we have*

$$P(A) = \sum_i P(A|E_i)P(E_i).$$

Example

$\Omega = \{ 1,2,3,4,5,6 \}$, $E_1 = \{ 1,2,3 \}$, $E_2 = \{ 4,5 \}$, $E_3 = \{ 6 \}$, $A = \{ 2,4 \}$.

$$P(A) = 1/3, \quad A \cap E_1 = \{ 2 \}, \quad A \cap E_2 = \{ 4 \}, \quad A \cap E_3 = \emptyset.$$

$$P(A|E_1) = \frac{1/6}{1/2} = \frac{1}{3}, \quad P(A|E_2) = \frac{1/6}{1/3} = \frac{1}{2}, \quad P(A|E_3) = 0.$$

$$\sum_{i=1}^{3} P(A|E_i)P(E_i) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{3} = P(A).$$

A generalized version of Bayes' Theorem

Theorem

Let $\{ E_1, E_2, \ldots \mid E_i \in \mathcal{A} \}$ *be a finite or countable partition of* Ω. *For any* $A \in \mathcal{A}$ *with* $P(A) > 0$ *and any* m, *we have*

$$P(E_m|A) = \frac{P(A|E_m)P(E_m)}{\sum_i P(A|E_i)P(E_i)}.$$

Proof.

By Bayes' Theorem

$$P(E_m|A) = \frac{P(A|E_m)P(E_m)}{P(A)}.$$

The result then follows from the previous Lemma. □

Probability theory and introduction to SCA

- Probability Measure
- Random Variable
- Side-channel Analysis
- Side-channel Leakages
- Leakage Assessment

Definition

Definition

A *random variable* X is a function $X : \Omega \to \mathbb{R}$, such that certain conditions are satisfied.

Example

Let us consider the random experiment of tossing a coin, the sample space $\Omega = \{ H, T \}$. Define $X : \Omega \to \mathbb{R}$ such that $X(H) = 0,\ X(T) = 1$. Then X is a random variable.

Example of a random variable – indicator function

Example

Fix $A \in \mathcal{A}$, the *indicator function*, denoted 1_A, for A is defined as follows:

$$1_A : \mathcal{A} \to \mathbb{R}, \quad 1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}.$$

1_A is a random variable.

Distribution

Let X be a random variable, define P^X as follows:

$$P^X : \mathcal{R} \to [0,1], \quad B \mapsto P(X^{-1}(B)).$$

We say that P^X is *induced* by X and it is called the *distribution* of X.
Here $\mathcal{R}$ denotes the *Borel set*, a set of subsets of $\mathbb{R}$, which contains open sets, closed set, etc.

Remark

For simplicity, we will write $P(X \in B)$ instead of $P(X^{-1}(B))$. For example, we write $P(X \leq x)$ instead of $P(X^{-1}((-\infty, x]))$.

CDF

The *cumulative distribution function (CDF)* of X, denoted F, is defined as

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0,1] \\ x &\mapsto P^X((-\infty, x]) = P(X^{-1}((-\infty, x])) = P(X \leq x) \end{aligned}$$

Continuous random variable

When the distribution function $F(x) = P(X \leq x)$ has the form

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

we say that X has *probability density function (PDF)* f and X is called a *continuous random variable*.

Remark

When Ω is a countable set (finite or countably infinite), X is a discrete random variable. But we will focus on continuous random variables.

Example – continuous random variable

Example

Define $f(x) = 1$ for $x \in (0, 1)$ and 0 otherwise. $F(x) = \int_{-\infty}^{x} f(y)dy$, is given by

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}.$$

If X is a random variable that has F as its CDF, then we say that X induces a *uniform distribution* on $(0, 1)$.

Standard normal distribution

Example

A random variable Z that induces a *standard normal distribution* has probability density function

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

and cumulative distribution function

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{y^2}{2}\right) dy.$$

- Standard normal distribution will be very useful in later parts of the course and we use $\Phi(z)$ instead of $F(z)$ to denote its CDF.
- We say that Z is a *standard normal random variable*.

Standard normal distribution

Example

The following figure shows that $f(z)$ is a bell-shaped curve that is symmetric about 0. The symmetry is also apparent from the formula for $f(z)$.

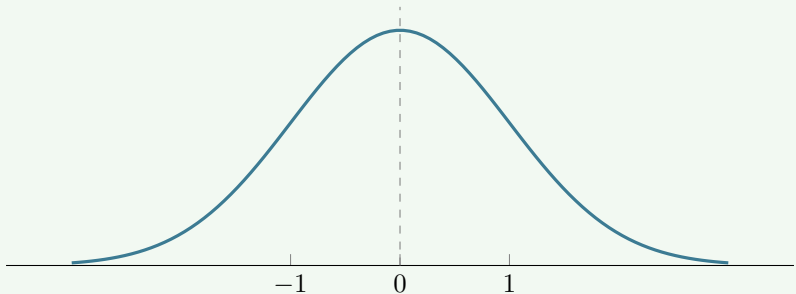


Figure: Probability density function of the standard normal random variable

Expectations and Variances

- The *expectation/mean* of a random variable X is the expected average value of X.
- And the *variance* of X is the average squared distance from the mean.
  - By squaring the distances, the small deviations from the mean are reduced and the big ones are enlarged.
  - Thus the variance measures how the values of X vary from the mean or how "spread out" the values of X are.

Expectation of a continuous random variable

When X is a continuous random variable with PDF f, its *expectation/mean* is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

provided the integral exists.

Example

X induces a *uniform distribution* on $(0, 1)$. X has PDF $f(x) = 1$ for $x \in (0, 1)$ and 0 otherwise.

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}.$$

Expectation of the standard normal random variable

When X is a continuous random variable with PDF f, its *expectation/mean* is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

provided the integral exists.

Example

Let Z be a random variable that induces the standard normal distribution. Then

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

and

$$\mathrm{E}[Z] = \int_{-\infty}^{\infty} z f(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp\left(-\frac{z^2}{2}\right) dz = 0.$$

Integral of an odd function with limits $-\infty$ and ∞ is zero.

Expectation of the standard normal random variable

Example

As shown in the following figure, $f(x)$ is symmetric about 0, so it is not surprising that the expected average value of Z is 0.

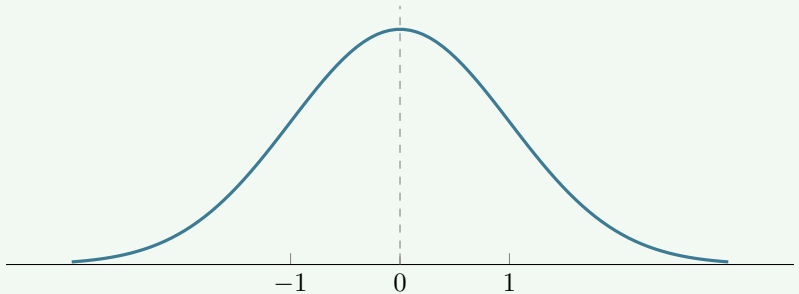


Figure: Standard normal distribution.

Some properties of expectations

- Given two random variables X, Y such that $\mathrm{E}[|X|] < \infty, \mathrm{E}[|Y|] < \infty$
- For any $a, b \in \mathbb{R}$

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y], \quad \mathrm{E}[aX+b] = a\mathrm{E}[X] + b, \quad \mathrm{E}[b] = b.$$

Variance

The *variance* of a random variable X is given by

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right] = \mathrm{E}\left[X^2\right] - \mu^2,$$

where μ denotes the expectation of X, $\mathrm{E}[X]$.
When X is a continuous random variable with PDF $f(x)$,

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f(x) dx$$

Example

X induces a *uniform distribution* on $(0,1)$. X has PDF $f(x) = 1$ for $x \in (0,1)$ and 0 otherwise. We have computed that $\mathrm{E}[X] = 0.5$

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - \mathrm{E}[X]^2 = \int_0^1 x^2 dx - \frac{1}{2^2} = \frac{x^3}{3}\Big|_0^1 - \frac{1}{4} = \frac{1}{12}.$$

Properties of variances

Given two random variables X, Y such that $\mathrm{E}\left[|X|\right] < \infty, \mathrm{E}\left[|Y|\right] < \infty$. Take any $a, b \in \mathbb{R}$. We have seen that

$$\mathrm{E}\left[X+Y\right] = \mathrm{E}\left[X\right] + \mathrm{E}\left[Y\right], \quad \mathrm{E}\left[aX+b\right] = a\mathrm{E}\left[X\right] + b, \quad \mathrm{E}\left[b\right] = b.$$

Then

$$\begin{aligned} \mathrm{Var}(aX+b) &= \mathrm{E}\left[(aX+b-\mathrm{E}\left[aX+b\right])^2\right] = \mathrm{E}\left[(aX+b-a\mathrm{E}\left[X\right]-b)^2\right] \\ &= a^2\mathrm{E}\left[(X-\mathrm{E}\left[X\right])^2\right] = a^2\mathrm{Var}(X). \end{aligned}$$

In particular, we have

$$\mathrm{Var}(b) =? \quad \mathrm{Var}(X+b) =? \quad \mathrm{Var}(aX) =?$$

Properties of variances

Given two random variables X, Y such that $\mathrm{E}\left[|X|\right] < \infty, \mathrm{E}\left[|Y|\right] < \infty$. Take any $a, b \in \mathbb{R}$. We have seen that

$$\mathrm{E}\left[X+Y\right] = \mathrm{E}\left[X\right] + \mathrm{E}\left[Y\right], \quad \mathrm{E}\left[aX+b\right] = a\mathrm{E}\left[X\right] + b, \quad \mathrm{E}\left[b\right] = b.$$

Then

$$\begin{aligned} \mathrm{Var}(aX+b) &= \mathrm{E}\left[(aX+b-\mathrm{E}\left[aX+b\right])^2\right] = \mathrm{E}\left[(aX+b-a\mathrm{E}\left[X\right]-b)^2\right] \\ &= a^2\mathrm{E}\left[(X-\mathrm{E}\left[X\right])^2\right] = a^2\mathrm{Var}(X). \end{aligned}$$

In particular, we have

$$\mathrm{Var}(b) = 0, \quad \mathrm{Var}(X+b) = \mathrm{Var}(X), \quad \mathrm{Var}(aX) = a^2\mathrm{Var}(X).$$

Variance of the standard normal random variable

Example

Let Z be a random variable that induces the standard normal distribution. We know that

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

and $\mathrm{E}[Z] = 0$

$$\mathrm{Var}(Z) = \mathrm{E}\left[Z^2\right] - 0 = \int_{-\infty}^{\infty} z^2 f(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) dz = 1.$$

We write $Z \sim \mathcal{N}(0, 1)$ to indicate that Z induces a standard normal distribution with mean 0 and variance 1.

z_α

Given any $\alpha \in (0,1)$, we define z_α such that

$$P(Z > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha, \quad \text{i.e.} \quad \Phi(z_\alpha) = 1 - \alpha.$$

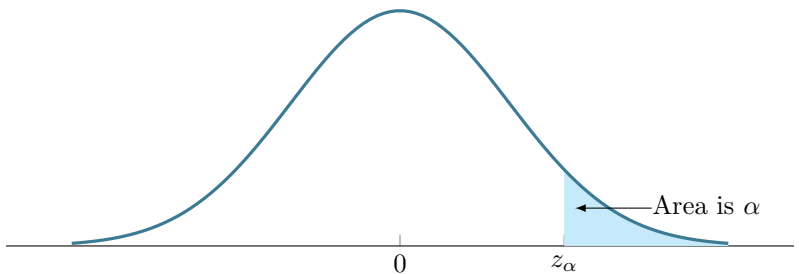


Figure: Probability density function $f(z)$ for $Z \sim \mathcal{N}(0,1)$. $P(Z > z_\alpha) = \alpha$, α corresponds to the area under $f(z)$ for $z > z_\alpha$.

z_α

α	0.1	0.05	0.01	0.005	0.001
$1-\alpha$	0.900	0.950	0.990	0.995	0.999
z_α	1.282	1.645	2.326	2.576	3.090

Table: A few values of z_α with corresponding α.

Normal random variable

- Let $Z \sim \mathcal{N}(0, 1)$ be a standard normal random variable
- Take any $\sigma, \mu \in \mathbb{R}$ with $\sigma^2 > 0$
- Define $Y = \sigma Z + \mu$
- It can be shown that Y has PDF

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

And

$$\mathrm{E}[Y] = \mu, \quad \mathrm{Var}(Y) = \sigma^2$$

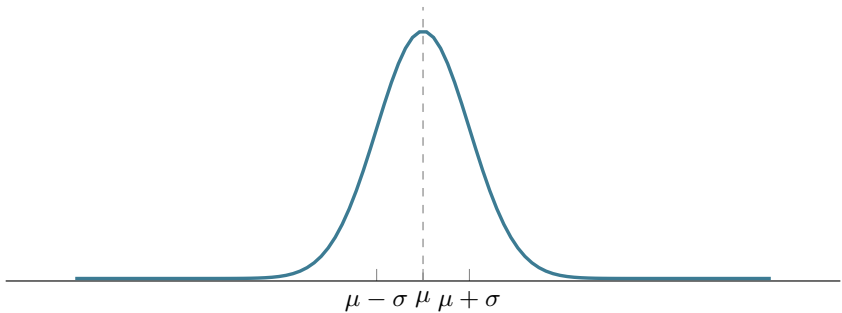
- We say that Y induces a *normal distribution with mean μ and variance σ^2*, written $Y \sim \mathcal{N}(\mu, \sigma^2)$. Y is also called *normal/a normal random variable*. We note that the mean and variance fully define a normal distribution.

Normal random variable

$f(y)$ is a bell-shaped curve symmetric about μ and obtains its maximum value of

$$\frac{1}{\sigma\sqrt{2\pi}} \approx \frac{0.399}{\sigma}$$

at $y = \mu$



Independent random variable

Definition

Given two random variables $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$, they are said to be *independent* if for any $A, B \in \mathcal{R}$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Here $\mathcal{R}$ denotes the *Borel set*, a set of subsets of $\mathbb{R}$, which contains open sets, closed set, etc.

If two random variables $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ are independent, it can be proven that

$$\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y] \quad \text{if} \quad \mathrm{E}[|X|] < \infty \text{ and } \mathrm{E}[|Y|] < \infty.$$

Covariance

- To analyze the relation between two random variables X and Y, we define the *covariance* of X and Y to be

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\right].$$

- It can be shown that

$$\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].$$

- It is easy to see that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ and $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

Definition

Let X and Y be two random variables. If $\mathrm{Cov}(X, Y) = 0$, we say that X and Y are *uncorrelated*. Otherwise, we say that X and Y are correlated.

Correlation coefficient

Definition

Let X and Y be two random variables with finite variances. The *correlation coefficient* of X and Y is given by

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

- $-1 \leq \rho \leq 1$
- Answer the question: if large values of X tend to be paired with large Y values or small Y values.
- If when X is large (or small), Y is also large (or small), then the signs of $X - \mathrm{E}[X]$ and $Y - \mathrm{E}[Y]$ will tend to be the same. And the absolute value of ρ will be bigger.
- If when X is large (or small), Y is small (or large), then the signs of $X - \overline{X}$ and $Y - \overline{Y}$ will tend to be different. And the absolute value of ρ will be bigger.
- In the special case when X and Y are uncorrelated, $\rho = 0$.
- In particular, if X and Y are independent, then $\rho = 0$

Probability theory and introduction to SCA

- Probability Measure
- Random Variable
- Side-channel Analysis
- Side-channel Leakages
- Leakage Assessment

Why are we interested in physical attacks?

- Cryptography provides algorithms that enable secure communication in theory
- In the real world, these algorithms have to be implemented on real devices:
  - software implementations: general-purpose devices
  - hardware implementations: dedicated secure hardware devices
- To evaluate the security level of cryptographic implementations, it is necessary to include a physical security assessment

Targets and Attack Goals

Targets

- Credit cards
- Passports
- Key Fob
- ...

Goals:

- Recovery of the secret key
- Privilege escalation
- IP theft
- ...

picture source: https://goodtimes.ca/how-many-credit-cards-is-too-many/

Different physical attack methods

- Side-channel analysis attacks
  - EM/Power analysis
  - Timing analysis
  - Cache attacks
- Fault attacks
  - Optical fault injection
  - Electromagnetic fault injection
  - Clock/voltage glitch
- Hardware Trojans
- ...

picture source: https://nl.dreamstime.com/stock-foto-s-hamer-die-computer-boos-tonen-raken-image34210923

Side-channel analysis attacks

- Side-channel analysis attacks target cryptographic implementations passively.
- The attacks exploit the possibility of the attacker observing the physical characteristics of a device that is running the cryptographic algorithm.
- The attacker obtains the side-channel information, e.g. power consumption, and execution time, then utilizes such information to recover the secret key.
- In this course, we will focus on power analysis attacks that exploit power consumption information.
- The attack methodologies can be used in a similar manner when electromagnetic (EM) emanation is analyzed.

Remark

We use the terminology *side-channel analysis* attacks only in the narrower meaning which refers to power analysis attacks. In short, we also write side-channel analysis as *SCA*.

Device under test (DUT)

- The device that we take measurement of is called the *device under test (DUT)*
  - A microcontroller running a software implementation
  - An FPGA or ASIC realizing a hardware implementation.
- We assume the attacker has certain knowledge of the implementation.
  - How to interface with the encryption routine
  - Whether the computation is executed serially or in parallel
  - Whether some types of countermeasures are present
  - Generally, this type of information can be also obtained by reverse engineering, visual inspection of the side-channel measurements, or sometimes just with a simple trial-and-error technique.

Attacker goal

- The ultimate goal of the attacker is to recover the master key of a symmetric block cipher or the private key of a public-key cipher.

Non-profiled SCA

- If the attacker does not have access to a similar device, just the target device or just the measurements coming from the target device, we talk about a *non-profiled SCA*.
- In a general scenario, this attack utilizes a set of measurements where a fixed secret key is used to encrypt multiple (random) plaintexts.

Profiled SCA

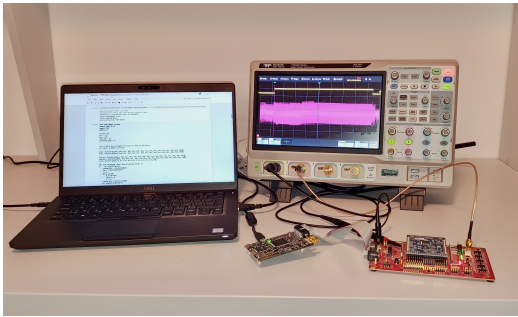
- If we assume the attacker has access to a clone device of the target device, then the attacker can carry out a *profiled SCA*.
- The attack operates in two phases.
- In the profiling phase, the attacker acquires side-channel measurements for known plaintext/ciphertext and known key pairs.
  - This set of data is used to characterize or model the device.
- Then the attacker acquires a few measurements from the target device, usually identical to the clone device, with known plaintext/ciphertext but the key is secret.
  - These measurements from the target device are then tested against the characterized model from the clone device.

Measurement device

- Power analysis measures the *power consumption* of the device under attack.
- The power consumption is in the form of a voltage change.
- The measurement is normally done with a digital sampling oscilloscope – a device that takes samples of the measured voltage signal over time.
- We refer to each sample point as a *time sample*.

Setup

- Ready-to-use measurement platform NewAE ChipWhisperer-Lite (black in the middle) – handling the communication with the DUT and the acquisition.
- CW 308 UFO board (red) – a breakout board with the DUT – ARM Cortex-M4 (blue) mounted on top.
- The controlling and data processing were done from a laptop, from the Jupyter environment available for the ChipWhisperer platform.
- Teledyne T3DSO3504 benchtop oscilloscope, used mainly for convenience purposes – to precisely locate the time intervals in the initial analysis stage.



Trigger

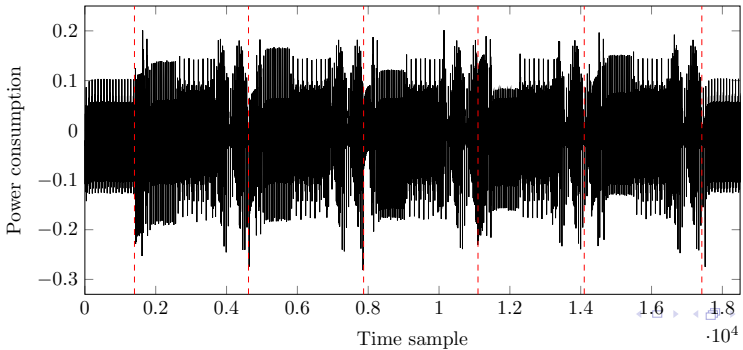
- An important task during the acquisition is capturing the correct time window corresponding to the operations we want to measure.
- In laboratory conditions, it is common to use an artificial trigger signal that indicates the start/end of the encryption.
- In real-world settings, it is necessary to identify the correct position by examining the captured signal – this is usually done based on the evaluator's expertise.

Power trace

- In our experiments, a trigger signal is raised to high during the computation that we want to capture and lowered afterward.
- One measurement consists of the voltage values for each *time sample* in this duration.
- It can be stored in an array of length equal to the total number of time samples in the measured time interval.
- It can also be drawn in a graph where the $x-$axis corresponds to time samples and the $y-$axis records the voltage values.
- Thus, we refer to the result of one measurement as a *(power) trace*.
- Note that, in the case of ChipWhisperer, which will be used for our experiments and analysis, the $y-$axis does not show the actual voltage value but a 10-bit value proportional to the current going through the shunt resistor

One trace

- One power trace for the first five rounds of PRESENT encryption.
- A sequence `nop` operations before and after the cipher computation.
- Certain patterns can be seen from the trace and we can deduce the corresponding operations in each time interval.
- From time sample $0 - 312$ and from time sample $2778 - 3100$ we have `nop` instructions.
- Five repeated patterns, indicated by red dotted lines $\rightarrow$ duration of each round



Datasets

There are four datasets that will be analyzed in more detail in the course.
All the datasets:

- Capture one round of software implementation of PRESENT.
- `nop` operations before and after PRESENT computation
- Each trace has 3600 time samples

More details on individual datasets:

- *Fixed dataset A*: This dataset contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a fixed plaintext `0xABCDEF1234567890`.
- *Fixed dataset B*: This dataset contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a fixed plaintext `0x84216BA484216BA4`.
- *Random plaintext dataset*: This dataset contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a random plaintext for each trace.
- *Random dataset*: This dataset contains 10000 traces with a random round key and a random plaintext for each trace.

PRESENT

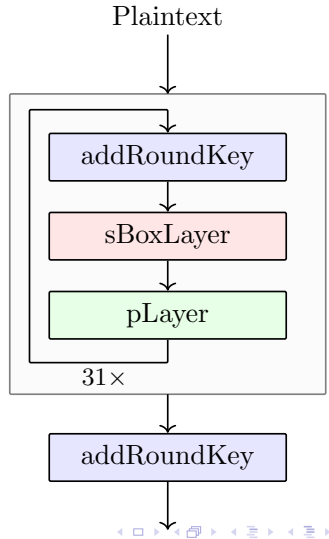
- Proposed in 2007 as a symmetric block cipher optimized for hardware implementation.
- Block length: $n = 64$
- Number of rounds: $\texttt{Nr} = 31$
- Key length: either 80 or 128.

PRESENT – encryption

- Round function: addRoundKey, sBoxLayer, and pLayer.
- After 31 rounds, addRoundKey is applied again before the ciphertext output

Remark

For PRESENT specification, we consider the 0th bit of a value as the rightmost bit in its binary representation. For example, the 0th bit of $3 = 011_2$ is 1, the 1st bit is 1 and the 2nd bit is 0.



Attack methods

- Classical power analysis attack methods
  - *Simple power analysis (SPA)*
  - *Differential power analysis (DPA)*
- SPA assumes the attacker has access to only one or a few measurements corresponding to some fixed inputs.
- DPA assumes the attacker can take measurements for a potentially unlimited number of different inputs.

Probability theory and introduction to SCA

- Probability Measure
- Random Variable
- Side-channel Analysis
- Side-channel Leakages
- Leakage Assessment

Leakages

- In the later parts of the course, we will see that by analyzing the power consumption, we can deduce the secret key. Consequently, we also refer to the power consumption as the *leakage* of the device.
- We consider the leakage consists of two parts: *signal* and *noise*.
- Signal refers to the part of the leakage that contains useful information for our attack and the rest is noise.
  - For example, if we would like to recover the hamming weight of an intermediate value, then the part of the leakage correlated to the hamming weight of that intermediate value is our signal.

Leakage is dependent on the operations being executed

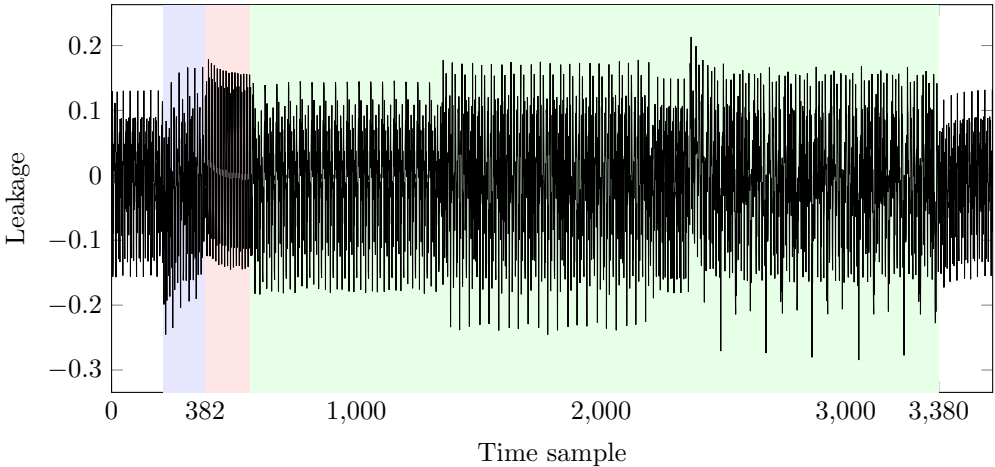


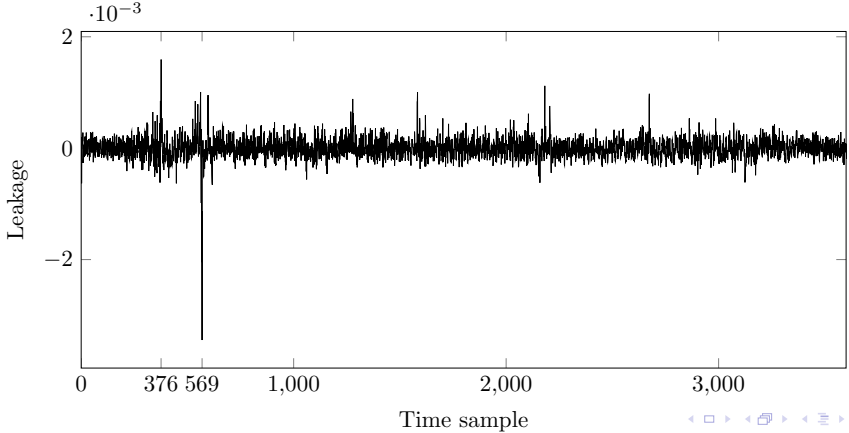
Figure: The averaged trace for 5000 traces from the *Fixed dataset A*. The blue, pink, and green parts of the trace correspond to addRoundKey, sBoxLayer, and pLayer respectively.

Leakage is dependent on the data being processed

- 1000 traces: each for a random plaintext with the 0th bit equal to 0; Take the average
- 1000 traces: each for a random plaintext with the 0th bit equal to 1; Take the average
- Take the difference trace of those two averages

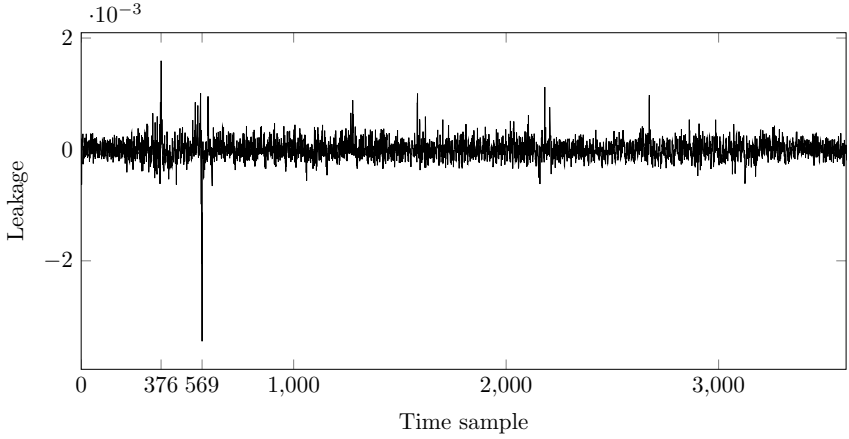
Leakage is dependent on the data being processed

- A few peaks in the difference trace and apart from those peaks, most of the points are close to zero.
- Those peaks indicate that at the corresponding time samples, the 0th bit of the plaintext is involved in the computations.



Leakage is dependent on the data being processed

- Compared with the previous figure, we can guess that the first and second peaks most likely correspond to addRoundKey and sBoxLayer.
- The later peaks are probably related to the pLayer operation.



Remark

- SPA exploits typically the relationship between the executed operations and the leakage
- DPA focuses on the relationship between the processed data and the leakage

Distribution of leakages

- For a fixed time sample t, let L_t, X_t, and N_t denote the random variables corresponding to the leakage, signal, and noise respectively

$$L_t = X_t + N_t.$$

- We consider X_t and N_t to be independent – "independent noise assumption"
- We fix the operation and the data, and we get a constant signal, i.e. X_t is a constant
  - The variants in the leakage will be caused by the noise
  - Let us take the *Fixed dataset A*: This dataset contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a fixed plaintext `0xABCDEF1234567890`.

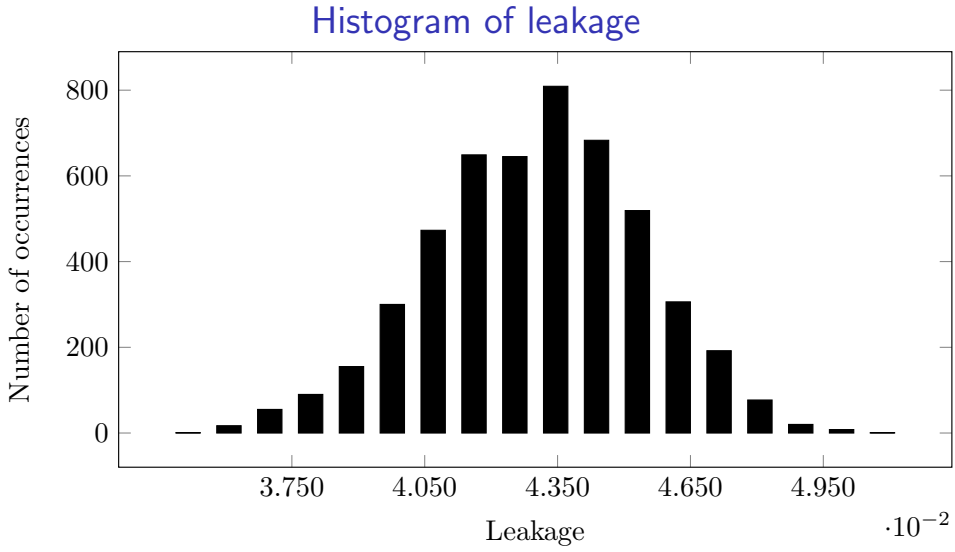


Figure: Histogram of leakages at $t = 3520$ across 5000 traces from the *Fixed dataset A*. $t = 3520$ corresponds to `nop` operations

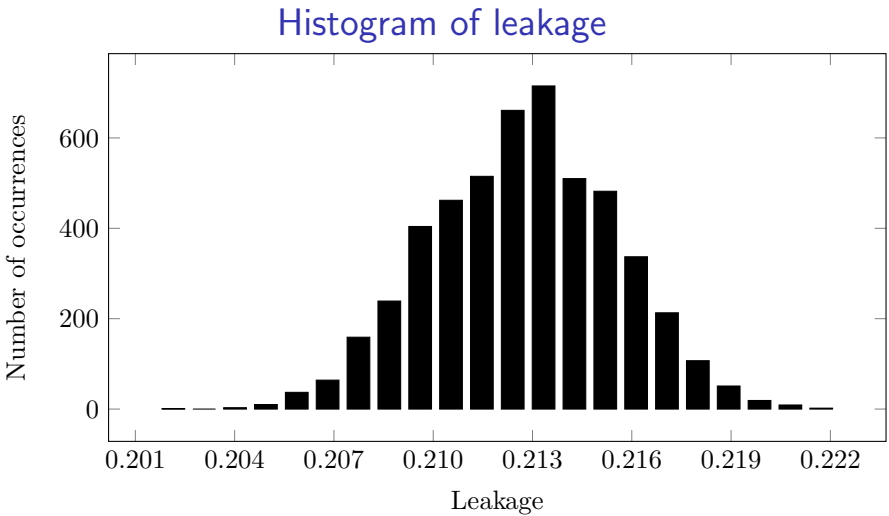


Figure: Histogram of leakages at $t = 2368$ across 5000 traces from the *Fixed dataset A*. $t = 2368$ corresponds to the highest peak in the averaged trace.

Distribution of leakages

- For a fixed time sample t, let L_t, X_t, and N_t denote the random variables corresponding to the leakage, signal, and noise respectively

$$L_t = X_t + N_t.$$

- We consider X_t and N_t to be independent
- We fix the operation and the data, and we get a constant signal, i.e. X_t is a constant
  - The variants in the leakage will be caused by the noise
  - We approximate the distribution induced by L_t with a normal distribution

$$L_t \sim \mathcal{N}(\mu_t, \sigma_t^2).$$

$$\mathrm{E}\left[L_t\right] = X_t + \mathrm{E}\left[N_t\right], \quad \mathrm{Var}(L_t) = \mathrm{Var}(N_t).$$

Leakage model

- Assume a value $\boldsymbol{v}$ is being processed in the DUT
- Let noise $\sim \mathcal{N}(0, \sigma^2)$ be a normal random variable with mean 0 and variance σ^2.
- *Identity leakage model*
  - The leakage is correlated to $\boldsymbol{v}$

$$\mathcal{L}(\boldsymbol{v}) = \boldsymbol{v} + \text{noise}.$$

- *Hamming weight model*
  - The leakage will then be correlated to $\mathrm{wt}(\boldsymbol{v})$, the Hamming weight of $\boldsymbol{v}$[1]

$$\mathcal{L}(\boldsymbol{v}) = \mathrm{wt}(\boldsymbol{v}) + \text{noise}.$$

Example

$\boldsymbol{v} = \texttt{A}$

- Identity leakage model: $\mathcal{L}(\boldsymbol{v}) = 10 + \text{noise}$
- Hamming weight leakage model: $\mathcal{L}(\boldsymbol{v}) = 2 + \text{noise}$

[1] The Hamming weight of vector $\boldsymbol{v} \in \mathbb{F}_2^m$ is defined to be the number of 1s in $\boldsymbol{v}$

Leakage model

$$\mathcal{L}(\boldsymbol{v}) = \boldsymbol{v} + \text{noise}, \quad \mathcal{L}(\boldsymbol{v}) = \mathrm{wt}\,(\boldsymbol{v}) + \text{noise}.$$

- Even though the actual leakage may not be exactly equal to $\mathcal{L}(\boldsymbol{v})$, those leakage models can be used to approximate the behavior of the actual leakages or for statistical analysis.
- For example, our previous experiments have demonstrated that the identity leakage model is realistic since when the data is fixed, the distribution of leakages is close to a normal distribution.

Probability theory and introduction to SCA

- Probability Measure
- Random Variable
- Side-channel Analysis
- Side-channel Leakages
- Leakage Assessment

Motivation

- In the next few weeks, we will see various SCA attacks on cryptographic implementations.
- As a developer, one might want to evaluate the DUT and conclude if it is vulnerable against SCA or not.
- Different new attacks are being developed and it is impractical to verify the security of our device against all of them.
- Leakage assessment aims to solve this problem by analyzing the power trace and answering the question of whether any input-dependent information can be detected in the traces of the DUT.
- We will see a method based on the student's $t-$test.
- The methodology is also referred to as *test vector leakage assessment (TVLA)*.

Remark

- Leakage assessment methods do not provide any conclusions in cases where data-dependent leakage is not detected.
- The absence of data-dependent leakage indicated by a particular method does not prove that the implementation is leakage-free.

Properties of expectations and variances

Given two random variables X, Y such that $\mathrm{E}[|X|] < \infty, \mathrm{E}[|Y|] < \infty$. Take any $a, b \in \mathbb{R}$. Then

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y], \quad \mathrm{E}[aX+b] = a\mathrm{E}[X] + b, \quad \mathrm{E}[b] = b.$$

And

$$\mathrm{Var}(b) = 0, \quad \mathrm{Var}(X+b) = \mathrm{Var}(X), \quad \mathrm{Var}(aX) = a^2\mathrm{Var}(X).$$

Modelling leakages

- Let us fix a time sample t.
- L_t: the random variable corresponding to the leakage at t.
- When the data being processed in the DUT is fixed, $L_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$
- Since we focus on one time sample, the operation is also fixed. We know that the signal X_t, which is dependent on only operation and/or data, is fixed.
- Since

$$L_t = X_t + N_t,$$

- According to the properties of expectations and variances,

$$\mu_t = X_t + \mathrm{E}\left[N_t\right], \quad \sigma_t^2 = \mathrm{Var}(N_t).$$

when the data being processed is fixed in the DUT.

Leakages for two plaintexts

- Let us consider a DUT running PRESENT encryption and let L_t, L'_t denote the leakages corresponding to encryptions of two different plaintexts at time sample t.
- We can write

$$L_t = X_t + N_t, \quad L'_t = X'_t + N'_t,$$

with

$$L_t \sim \mathcal{N}(\mu_t, \sigma_t^2), \quad L'_t \sim \mathcal{N}(\mu'_t, \sigma_t'^2).$$

- We take the signal to be part of the leakage related to the plaintext value.
- Since the noise is independent of the signal, we have $N_t = N'_t$.
- And

$$\sigma_t^2 = \sigma_t'^2, \quad \mu_t - X_t = \mu'_t - X'_t.$$

Equal signal

- We have seen before that the leakage L_t is dependent on the data being processed in the device.
- In fact, certain SCA attacks (e.g. DPA) exploit the dependency of the leakage on data.
- If the leakage is *not* exploitable, we would expect, at least, that the signals at time sample t should be the same when the only difference is the values of the data being processed.
- With our notations, this means that we would like to test if

$$X_t = X'_t, \quad \text{and equivalently} \quad \mu_t = \mu'_t.$$

Sample

- We have discussed that a random experiment is an experiment whose output cannot be predicted with certainty in advance.
- However, if the experiment is repeated many times, we can see "regularity" in the average output.
- For a given random experiment, the *sample space*, denoted by Ω, is the set of all possible outcomes.
- A random variable $X : \Omega \to \mathbb{R}$.
- $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$: normal random variable
- We repeat the random experiment n times and record the outcomes.
- Then the possible outcomes $\{ X_1, X_2, \ldots, X_n \}$ are n independent identically distributed random variables.
- We refer to this set as a *sample*.

Sample mean and sample variance

The *sample mean (empirical mean)*, denoted $\overline{X}$, is given by

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The *sample variance (empirical variance)*, denoted S_x^2, is given by

$$S_x^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Remark

The sample mean and sample variance are random variables. A realization of $\overline{X}$ and S_x^2 are represented as $\overline{x}$ and s_x^2.

Statistical hypothesis

- *Statistical hypothesis*: a statement about the parameters of an unknown distribution
- We call such a statement hypothesis because it is not known whether or not it is true.
- We will use **samples** from an unknown distribution to draw certain conclusions regarding a given hypothesis about this distribution.
- A procedure for determining whether or not the values of a sample are consistent with the hypothesis.
- The decision will then be either *accept* the hypothesis, or *reject* it.
- By accepting a hypothesis, we conclude that *the resulting data from the sample appear to be consistent with it*.

Example (Example of a hypothesis)

If we are interested in whether $\mu_t = \mu_t'$, we can set a hypothesis that $\mu_t = \mu_t'$.

Samples for TVLA

- Two datasets, each with M traces
- Encryption of one plaintext
- Encryption of another plaintext
- We will use
  - *Fixed dataset A*: This dataset contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a fixed plaintext

$$p_1 = \texttt{ABCDEF1234567890}.$$

  - *Fixed dataset B*: This dataset contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a fixed plaintext

$$p_2 = \texttt{0x84216BA484216BA4}.$$

- L_t: leakage for encryption of p_1 at time sample t. We will take *Fixed dataset A* as a sample for L_t
- L'_t: leakage for encryption of p_2 at time sample t. We will take *Fixed dataset B* as a sample for L_t

Statistical hypothesis

- The hypothesis that we want to test is called the *null hypothesis*, denoted by H_0.
- For example
$$H_0 : \mu_t = \mu'_t, \quad H_0 : \mu_t \geq \mu'_t.$$
- We will test the null hypothesis against an *alternative hypothesis*, denoted by H_1. For example
$$H_1 : \mu_t \neq \mu'_t, \quad H_1 : \mu_t > \mu'_t.$$
- For TVLA, the null and alternative hypotheses are:
$$H_0 : \mu_t = \mu'_t, \quad H_1 : \mu_t \neq \mu'_t.$$

Example

Let $t = 392$, the null and alternative hypotheses are

$$H_0 : \mu_{392} = \mu'_{392}, \quad H_1 : \mu_{392} \neq \mu'_{392}.$$

$t-$value

$$t - \text{value}_t := \frac{\overline{L_t} - \overline{L'_t}}{\sqrt{\frac{S_t^2 + S_t'^2}{M}}},$$

- $\overline{L_t}$: sample mean of L_t
- S_t^2: sample variance of L_t
- $\overline{L'_t}$: sample mean of L'_t
- $S_t'^2$: sample variance of L'_t
- M: the number of traces.

$t-$value – Example

Example

Let $t = 392$. We can compute the sample mean and sample variance of L_{392} with *Fixed dataset A*:

$$\overline{l_{392}} \approx -0.0525, \quad s^2_{392} \approx 1.5141 \times 10^{-6}.$$

With *Fixed dataset B*, we compute the sample mean and sample variance of L'_{392}:

$$\overline{l'_{392}} \approx -0.0501, \quad s'^2_{392} \approx 1.4801 \times 10^{-6}.$$

We set the following hypotheses

$$H_0 : \mu'_{392} = \mu_{392}, \quad H_1 : \mu'_{392} \neq \mu_{392}.$$

According to student's $t-$test, we calculate

$$t - \text{value}_t := \frac{\overline{l_t} - \overline{l'_t}}{\sqrt{\frac{s^2_t + s'^2_t}{M}}} = \frac{\overline{l'_{392}} - \overline{l_{392}}}{\sqrt{\frac{s^2_{392} + s'^2_{392}}{5000}}} = \frac{0.0024}{\sqrt{\frac{1.5141 \times 10^{-6} + 1.4801 \times 10^{-6}}{5000}}} \approx -98.1.$$

z_α

Given any $\alpha \in (0,1))$, we define z_α such that

$$P(Z > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha, \quad \text{i.e.} \quad \Phi(z_\alpha) = 1 - \alpha.$$

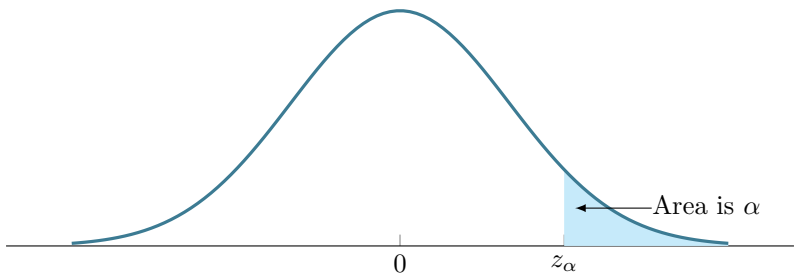


Figure: Probability density function $f(z)$ for $Z \sim \mathcal{N}(0,1)$. $P(Z > z_\alpha) = \alpha$, α corresponds to the area under $f(z)$ for $z > z_\alpha$.

Student's $t-$test

- *level of significance of the test*, α: when H_0 is true, the probability of rejecting it is not bigger than α.

$$P\left(\text{reject } H_0 | H_0 \text{ is true}\right) \leq \alpha.$$

- Given α, the procedure of hypothesis testing is to find the correct condition for rejecting/accepting H_0
- Student's $t-$test: reject H_0 if $|t - \text{value}_t| > z_{\alpha/2}$

$$P\left(|t - \text{value}_t| > z_{\alpha/2} \;\middle|\; H_0 \text{ is true}\right) = \alpha.$$

α	0.1	0.05	0.01	0.005	0.001
z_α	1.282	1.645	2.326	2.576	3.090

Example

For $t = 392$, we have computed $|t|-\text{value}_{392} \approx 98.1$. What is the conclusion of the student's $t-$test with a level of significance 0.01?

Student's $t-$test

$$H_0 : \mu_{392} = \mu'_{392}, \quad H_1 : \mu_{392} \neq \mu'_{392}.$$

Student's $t-$test: reject H_0 if $t - \text{value}_t > z_{\alpha/2}$

$$P\left(t - \text{value}_t > z_{\alpha/2} | H_0 \text{ is true}\right) = \alpha.$$

α	0.1	0.05	0.01	0.005	0.001
z_α	1.282	1.645	2.326	2.576	3.090

Example

- For $t = 392$, we have computed $|t|-\text{value}_{392} \approx 98.1$. Take $\alpha = 0.01$, then $z_{\alpha/2} = 2.576$.
- $98.1 > 2.576$. By the student's $t-$test with a level of significance, 0.01 is to reject H_0.
- We conclude that $\mu_{392} \neq \mu'_{392}$
- The probability that our conclusion is wrong is $\alpha = 0.01$.

What does the conclusion mean to us

- Certain SCA attacks (e.g. DPA) exploit the dependency of the leakage on data.
- If the leakage L_t is *not* exploitable, we would expect, at least, that the signals at time sample t should be the same when different values of the data are being processed at this specific time.
- This means:

$$X_t = X'_t, \quad \text{and equivalently} \quad \mu_t = \mu'_t.$$

- When we take the signal to be part of the leakage correlated to the plaintext value, our example concludes that the signals at time sample 392 for encryption of plaintexts `ABCDEF1234567890` and `84216BA484216BA4` are very likely to be different, according to our measurements *Fixed dataset A* and *Fixed dataset B*.
- The probability of the conclusions being wrong is 0.01.

TVLA threshold

Following the convention for TVLA, we set $z_{\alpha/2} = 4.5$. By definition, this threshold corresponds to

$$\frac{\alpha}{2} = 1 - \Phi(z_\alpha) = 1 - \Phi(4.5) = 1 - 0.9999966023268753 \approx 3.4 \times 10^{-6}.$$

The significance level is given by

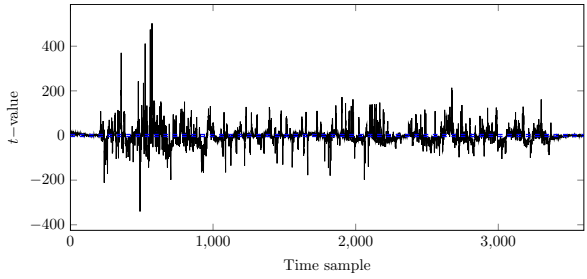
$$\alpha \approx 6.8 \times 10^{-6}.$$

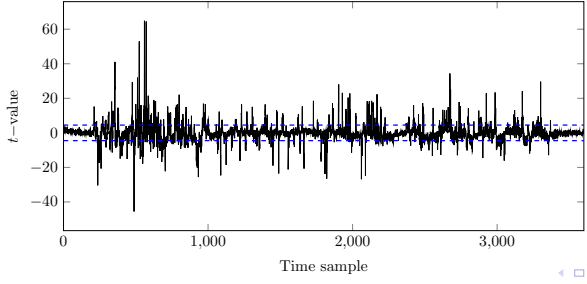
This means that there is a 6.8×10^{-4} percent chance that we would reject the null hypothesis (i.e. conclude that the means are different) in case it is true (i.e. the means are in fact the same).

TVLA procedure

- Fix the master key
- Measure M traces for one fixed plaintext
- Measure M traces for another fixed plaintext
- For each time sample t
  - Compute the $t-$values with the two sets of traces
  - Compare the $t-$value with the threshold 4.5 and -4.5

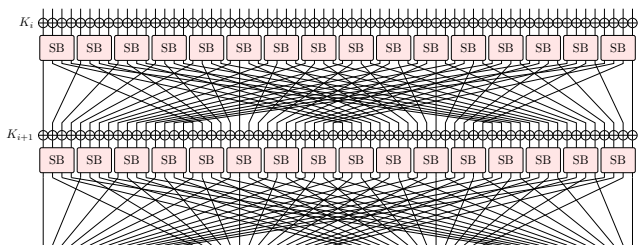
TVLA results – 5000 and 50 traces



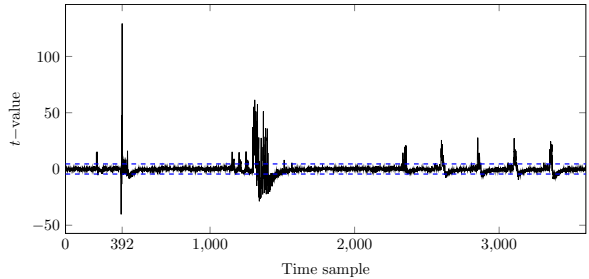


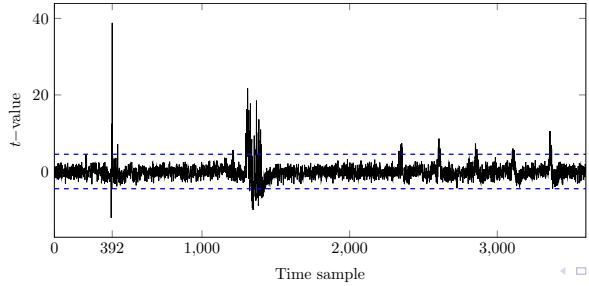
Other settings

- We can also have different numbers of traces for those two measurements - another formula for $t-$value
- We have seen fixed vs. fixed (two different fixed plaintexts). Can also have fixed vs. random (a fixed plaintext, random plaintexts) – Welch's $t-$test
- Instead of fixing the plaintext, we can also fix one intermediate value, e.g. the 0th Sbox output of PRESENT



TVLA results – Sbox output – fixed vs. fixed – *Random dataset*





Code

- A simplified code for TVLA with Sbox output as the fixed intermediate value can be found here

https://github.com/XIAOLUHOU/SCA-measurements-and-analysis----Experimental-results-for-textbook/blob/main/Assignment_materials/TVLA.ipynb

- All datasets and analysis code related to SCA can be found here

https://github.com/XIAOLUHOU/SCA-measurements-and-analysis----Experimental-results-for-textbook/tree/main

Observations

- When more traces are used (i.e. when the sample size is bigger), it is more likely for us to capture information about the inputs from the leakages. We will see next week that more traces indeed indicate higher chances for the attacks to be successful.
- When we take signal to be the 0th Sbox output, the highest $|t|-$value is obtained at 392. We will see that this is our *point of interest* (POI) for our attack – sample points that give the best attack results
- Compared to the signal being the plaintext, the $|t|-$values are in general smaller with much fewer times samples crossing the threshold when the signal is given by an Sbox output. This is unsurprising as we would expect more computations to be correlated with the plaintext rather than a single Sbox output.

More about leakage assessment

- TVLA was first proposed in 2011[1]
- More discussions on how to set the threshold[2]
- Another prominent leakage assessment method - Person's χ^2-test[3]

[1]Gilbert Goodwill, B. J., Jaffe, J., & Rohatgi, P. (2011, September). A testing methodology for side-channel resistance validation. In NIST non-invasive attack testing workshop (Vol. 7, pp. 115-136).

[2]Ding, A. A., Zhang, L., Durvaux, F., Standaert, F. X., & Fei, Y. (2017, November). Towards sound and optimal leakage detection procedure. In International Conference on Smart Card Research and Advanced Applications (pp. 105-122). Springer, Cham.

[3]Schneider, T., & Moradi, A. (2015). Leakage assessment methodology: A clear roadmap for side-channel evaluations. In Cryptographic Hardware and Embedded Systems–CHES 2015: 17th International Workshop, Saint-Malo, France, September 13-16, 2015, Proceedings 17 (pp. 495-513). Springer Berlin Heidelberg.

Week 6 - Quiz

- Before the lecture
- 1.5 hours: 11 am - 12:30pm
- Do not use “písané písmo” but “paličkové”.
- Write down the answers on the papers given to you, more can be provided upon request - full name should be written on each page of the answer sheet.
- Detailed computation steps are required. 0 mark will be given if only a final answer is provided.
- Four questions
- Weeks 1 – 3