Cryptography and Embedded System Security
CRAESS_I

Xiaolu Hou

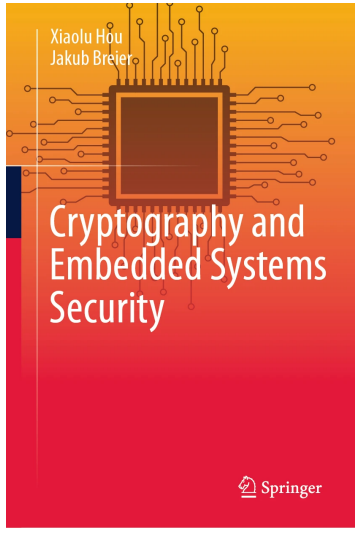
FIIT, STU
xiaolu.hou @ stuba.sk

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
    - 4.2.4;
    - 4.3.1,
    - 4.4.1



Lecture Outline

- Simple Power Analysis (SPA)
- Sample Correlation Coefficient
- Non-profiled DPA Attacks on Symmetric Block Ciphers
- Signal-to-Noise Ratio

Classical power analysis attack methods

- *Simple power analysis (SPA)* assumes the attacker has access to only one or a few measurements corresponding to some fixed inputs.
- *Differential power analysis (DPA)* assumes the attacker can take measurements for a potentially unlimited number of different inputs.
- We will see
  - SPA on RSA
  - DPA on symmetric block ciphers (example: PRESENT)

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 traces 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.

SPA and non-profiled DPA

- Simple Power Analysis (SPA)
- Sample Correlation Coefficient
- Non-profiled DPA Attacks on Symmetric Block Ciphers
- Signal-to-Noise Ratio

Recall – RSA

Definition (RSA)

Let $n = pq$, where p, q are distinct prime numbers. Let $\mathcal{P} = \mathcal{C} = \mathbb{Z}_n$, $\mathcal{K} = \mathbb{Z}^*_{\varphi(n)} - \{ 1 \}$. For any $e \in \mathcal{K}$, define encryption

$$E_e : \mathbb{Z}_n \to \mathbb{Z}_n, \quad m \mapsto m^e \bmod n,$$

and the corresponding decryption

$$D_d : \mathbb{Z}_n \to \mathbb{Z}_n, \quad c \mapsto c^d \bmod n,$$

where $d = e^{-1} \bmod \varphi(n)$. The cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, where $\mathcal{E} = \{ E_e : e \in \mathcal{K} \}$, $\mathcal{D} = \{ D_d : d \in \mathcal{K} \}$, is called *RSA*.

- $\varphi(n) = (p-1)(q-1)$
- Public key: n, e, RSA modulus, encryption exponent
- Private key: d, decryption exponent, is d even or odd?

Remark

- Since $\varphi(n) = (p-1)(q-1)$ is even and $\gcd(d, \varphi(n)) = 1$, d is odd.
- In particular $d_0 = 1$.

Square and multiply algorithm

- Let $n \geq 2$ be an integer, $d \in \mathbb{Z}^*_{\varphi(n)}$, $a \in \mathbb{Z}_n$
- Binary representation of $d = d_{\ell_d - 1} \dots d_2 d_1 d_0$, where $d_i = 0, 1$ and

$$d = \sum_{i=0}^{\ell_d - 1} d_i 2^i,$$

- Then we have

$$a^d = a^{\sum_{i=0}^{\ell_d - 1} d_i 2^i} = \prod_{i=0}^{\ell_d - 1} (a^{2^i})^{d_i} = \prod_{0 \leq i < \ell_d, d_i = 1} a^{2^i}.$$

Thus, to compute $a^d \bmod n$, we can

- First compute a^{2^i} for $0 \leq i < \ell_d$
- Then a^d is a product of a^{2^i} for which $d_i = 1$

Left-to-right square and multiply algorithm

Algorithm 1: Left-to-right square and multiply algorithm for computing modular exponentiation.

Input: $n,\ a,\ d$// $n \in \mathbb{Z}, n \geq 2;\ a \in \mathbb{Z}_n;\ d \in \mathbb{Z}_{\varphi(n)}$
Output: $a^d \bmod n$

```
1 t = 1
2 for i = ℓ_d − 1, i ≥ 0, i − − do
3     t = t ∗ t mod n
      // ith bit of d is 1
4     if d_i = 1 then
5         t = a ∗ t mod n
6 return t
```

Left-to-right square and multiply algorithm

```
1 t = 1
2 for i = ℓ_d − 1, i ≥ 0, i − − do
3     t = t ∗ t mod n
      // ith bit of d is 1
4     if d_i = 1 then
5         t = a ∗ t mod n
6 return t
```

Example

Let $n = 15,\ d = 3 = 11_2,\ a = 2$. Then

$$a^d \bmod n = 2^3 \bmod 15 = 8 \bmod 15 = 8$$

i	d_i	t
1	1	2
0	1	8

Left-to-right square and multiply algorithm

For our experiment, we have set

$$p = 29, \quad q = 41, \quad n = 1189, \quad \varphi(n) = 1120, \quad e = 3, \quad d = 747$$

Algorithm 2: Left-to-right square and multiply algorithm for computing modular exponentiation with parameters from above.

Input: a// $a \in \mathbb{Z}_{1189}$
Output: $a^{747} \bmod 1189$

```
1 n = 1189
2 dbin = [1,1,0,1,0,1,1,1,0,1]// binary representation of d = 747, d_0 = 1, d_1 = 1
3 ℓ_d = length of dbin// bit length of d
4 t = 1
5 for i = ℓ_d − 1, i ≥ 0, i − − do
6     t = t ∗ t mod n
      // ith bit of d is 1
7     if d_i = 1 then
8         t = a ∗ t mod n
9 return t
```

Recall – leakage is dependent on the operations being executed

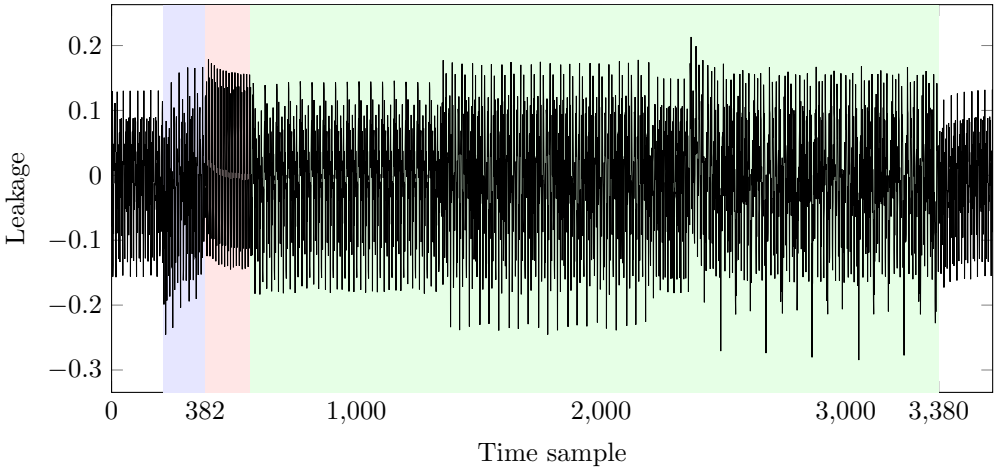


Figure: The averaged trace for 5000 traces from the *Fixed dataset A*. The blue, pink, and green parts of the trace correspond to addRoundKey, sBoxLayer, and pLayer respectively.

SPA on left-to-right square and multiply algorithm

- An SPA attack on the square and multiply algorithm exploits the leakage dependence on the performed operations – we examine the traces to figure out if both square and multiplication are executed in one loop (the corresponding bit of d is 1) or not (the corresponding bit of d is 0).
- Following Kerckhoffs' principle, we assume the attacker has the knowledge of our algorithm except for line 2, which specifies the value of d

Definition (Kerckhoffs' principle)

The security of a cryptosystem should depend only on the secrecy of the key.

In other words, everything is public knowledge except for the secret key.

SPA on left-to-right square and multiply algorithm

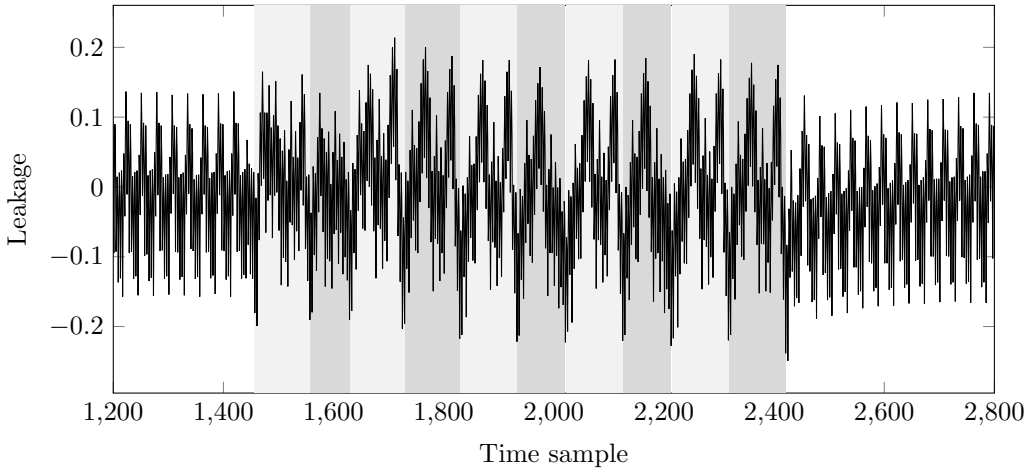


Figure: One trace. We can see 10 similar patterns.

What are those patterns

Input: a// $a \in \mathbb{Z}_{1189}$
Output: $a^{747} \bmod 1189$

```
1 n = 1189
2 dbin = [1, 1, 0, 1, 0, 1, 1, 1, 0, 1]
3 ℓ_d = length of dbin
4 t = 1
5 for i = ℓ_d − 1, i ≥ 0, i − − do
6 |   t = t ∗ t mod n
7 |   if d_i = 1 then
8 |   |   t = a ∗ t mod n
9 return t
```

We have two guesses

Guess a Each pattern corresponds to one modular operation (modular square from line 6 or modular multiplication from line 8);

Guess b Each pattern corresponds to one loop from line 5.

- S: modular square operation from line 6
- M: modular multiplication from line 8.
- Loop in line 5 contains either one square operation (S) or one square followed by a multiplication operation (SM).

$$\text{S} \longleftrightarrow d_i = 0, \quad \text{SM} \longleftrightarrow d_i = 1.$$

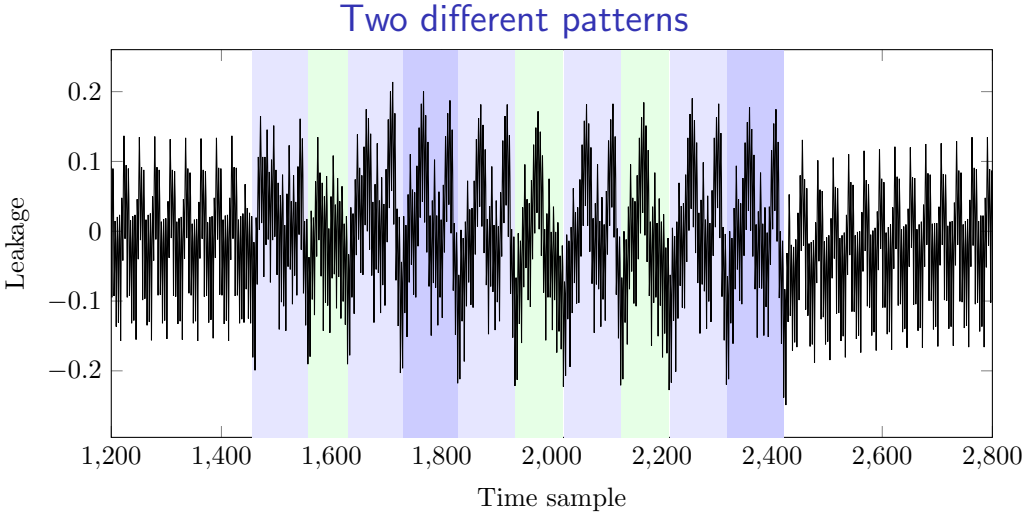


Figure: Highlighted two types of patterns from the previous figure. One pattern with a single cluster of peaks (colored in green) and one with more than one cluster of peaks (colored in blue).

Assume Guess a is correct

Input: a// $a \in \mathbb{Z}_{1189}$
Output: $a^{747} \bmod 1189$

1 $n = 1189$
2 $dbin = [1, 1, 0, 1, 0, 1, 1, 1, 0, 1]$
3 $\ell_d =$ length of $dbin$
4 $t = 1$
5 **for** $i = \ell_d - 1$, $i \geq 0$, $i - -$ **do**
6 $\quad t = t * t \bmod n$
7 $\quad$ **if** $d_i = 1$ **then**
8 $\quad\quad t = a * t \bmod n$
9 **return** t

Guess a

Each pattern corresponds to one modular operation (modular square from line 6 or modular multiplication from line 8);

- There are two possibilities:
  - Green patterns → S; blue patterns → M
  - Green patterns → M; blue patterns → S
- We know that $d_0 = 1$, hence the last blue-colored pattern does not represent S.
- The start of the computation will always be a modular square operation, which then indicates that the first blue-colored pattern corresponds to S.
- We have reached a contradiction

Assume Guess b is correct

Input: a// $a \in \mathbb{Z}_{1189}$
Output: $a^{747} \bmod 1189$

```
1 n = 1189
2 dbin = [1, 1, 0, 1, 0, 1, 1, 1, 0, 1]
3 ℓ_d = length of dbin
4 t = 1
5 for i = ℓ_d − 1, i ≥ 0, i − − do
6 |   t = t ∗ t mod n
7 |   if d_i = 1 then
8 |   |   t = a ∗ t mod n
9 return t
```

Guess b Each pattern corresponds to one loop from line 5.

- There are two possibilities:
  - Green patterns $\rightarrow d_i = 0$; blue patterns $\rightarrow d_i = 1$
  - Green patterns $\rightarrow d_i = 1$; blue patterns $\rightarrow d_i = 0$
- We know that $d_0 = 1$, hence the last blue-colored pattern does not represent $d_i = 0$.
- Green patterns $\rightarrow d_i = 0$; blue patterns $\rightarrow d_i = 1$

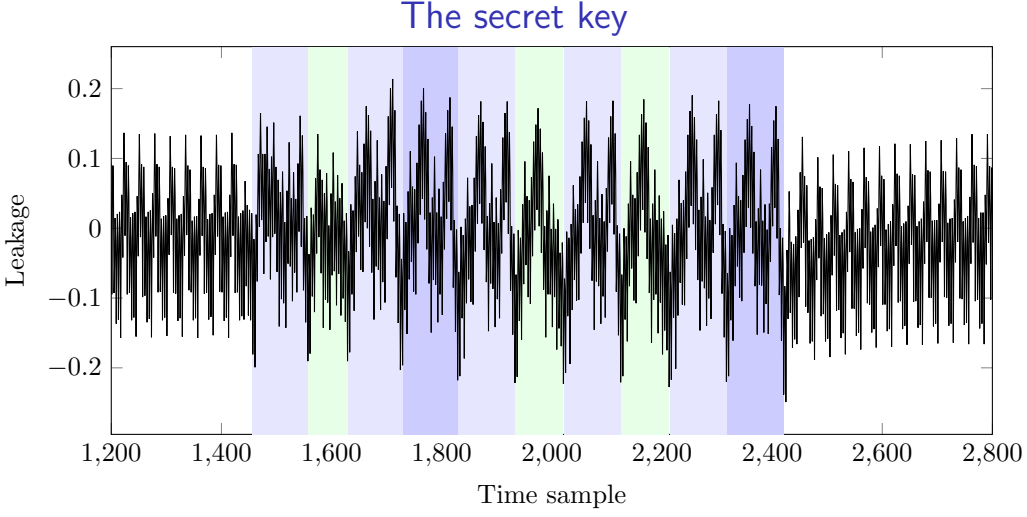
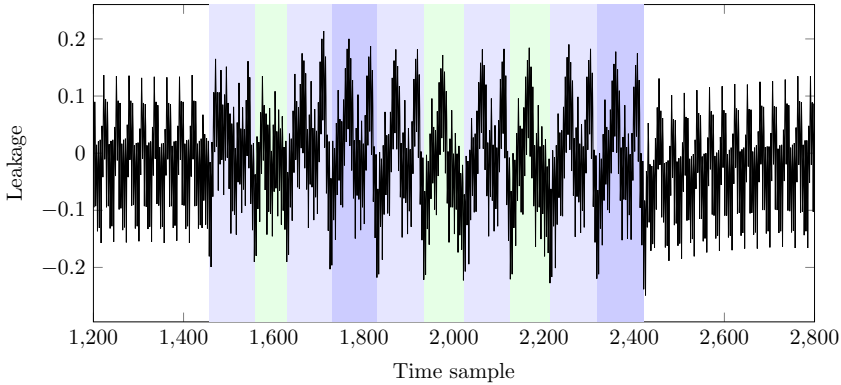


Figure: Green patterns $\rightarrow d_i = 0$; blue patterns $\rightarrow d_i = 1$

We can then read out the value of bits d_i $(i = \ell_d - 1, \ldots, 0, 1)$

The secret key



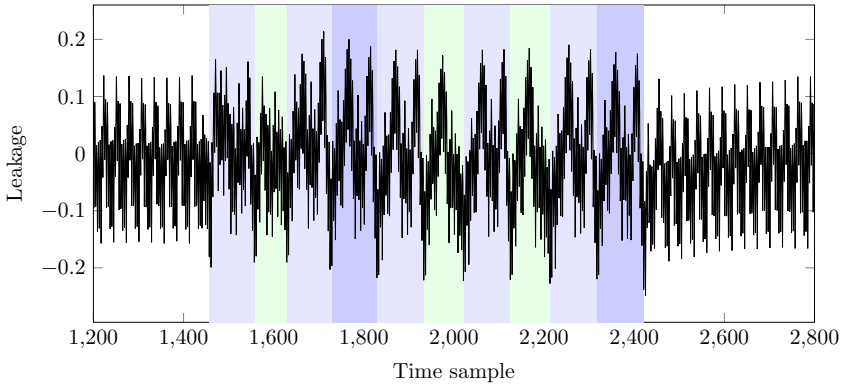
We can then read out the value of bits d_i $(i = \ell_d - 1, \ldots, 0, 1)$:

$$1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1.$$

Finally, we recover the secret key

$$d = 1011101011_2 = 747.$$

The second pattern



- It is not clear whether the second pattern should be green or blue
- Shorter than blue patterns
- In reality, brute force

Remark

- A similar SPA attack can also be applied to the right-to-left square and multiply algorithm.
- The attack can be carried out during either the decryption of RSA or the signature signing procedure of RSA signatures.

SPA and non-profiled DPA

- Simple Power Analysis (SPA)
- Sample Correlation Coefficient
- Non-profiled DPA Attacks on Symmetric Block Ciphers
- Signal-to-Noise Ratio

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 $X : \Omega \to \mathbb{R}$.
- 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.

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 $X - \overline{X}$ and $Y - \overline{Y}$ will tend to be the same. And the 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 $\rho = 0$

Smaple for a pair of random variable

- X and Y: two random variables
- $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$: sample for this pair of random variable (X, Y)
- $\overline{X}$: sample mean for $\{X_1, X_2, \ldots, X_n\}$
- S_x^2: sample variance for $\{X_1, X_2, \ldots, X_n\}$
- $\overline{Y}$: sample mean for $\{Y_1, Y_2, \ldots, Y_n\}$
- S_y^2: sample variance for $\{Y_1, Y_2, \ldots, Y_n\}$

Remark

We note that since the correlation coefficient analyzes the relations between X and Y, we collect samples in pairs (X_i, Y_i).

Sample correlation coefficient – Definition

Correlation coefficient:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Definition

We define the *sample correlation coefficient*, denoted by r, as follows:

$$\begin{aligned} r &= \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\sqrt{S_x^2 S_y^2}} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2\right)\left(\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2\right)}} \\ &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2}}. \end{aligned}$$

Example

Example

Suppose we obtained the following sample

$$\{(1, 11), (0, 9), (1, 12), (1, 14), (0, 9)\}$$

for (X, Y). Then the sample mean for X is given by

$$\overline{x} = ?$$

And the sample mean for Y is given by

$$\overline{y} = ?$$

Example

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}.$$

Example

Sample for (X, Y):

$$\{(1, 11), (0, 9), (1, 12), (1, 14), (0, 9)\}$$

Sample means

$$\overline{x} = \frac{1+0+1+1+0}{5} = \frac{3}{5}, \quad \overline{y} = \frac{11+9+12+14+9}{5} = \frac{55}{5} = 11.$$

The sample correlation coefficient for X and Y is given by

$$r \quad = \quad ?$$

Example

Example

Sample for (X, Y):

$$\{(1, 11), (0, 9), (1, 12), (1, 14), (0, 9)\}$$

Sample means:

$$\overline{x} = \frac{3}{5}, \quad \overline{y} = 11.$$

The sample correlation coefficient for X and Y is given by

$$\begin{aligned} r &= \frac{\sum_{i=1}^{5}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{5}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{5}(y_i - \overline{y})^2}} \\ &= \frac{0.4 \times 0 + (-0.6) \times (-2) + 0.4 \times 1 + 0.4 \times 3 + (-0.6) \times (-2)}{\sqrt{0.4^2 \times 3 + 0.6^2 \times 2}\sqrt{2^2 + 1 + 3^2 + 2^2}} \approx 0.861. \end{aligned}$$

SPA and non-profiled DPA

- Simple Power Analysis (SPA)
- Sample Correlation Coefficient
- Non-profiled DPA Attacks on Symmetric Block Ciphers
- Signal-to-Noise Ratio

DPA vs. SPA

- We have seen that SPA analyzes leakages along the time axis, exploiting relationships between leakages and operations.
- DPA, on the other hand, exploits the relationship between leakages at specific time samples and the data being processed in the DUT.
- Compared to SPA, DPA does not require detailed knowledge about the implementation.
- Normally it suffices to know what cryptographic algorithm is being executed in the DUT.

Attack assumption

- In our DPA attacks, we assume the attacker has the knowledge of the plaintext and the goal is to recover the very first round key of a symmetric block cipher – for some ciphers, e.g. PRESENT-80, this is the first round key; for some ciphers, e.g. AES-128, this is the whitening key, which is equal to the master key.
- Similar attack strategies apply if we assume the attacker has the knowledge of the ciphertext and aims to recover the last round key.

PRESENT

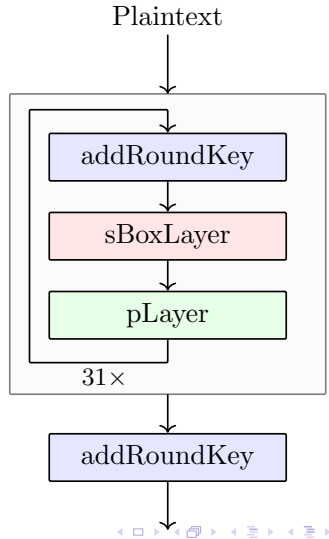
- Proposed in 2007 as a symmetric block cipher optimized for hardware implementation.
- Block length: $n = 64$
- Number of rounds: $\texttt{Nr} = 31$
- Key length: either 80 or 128.

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.



PRESENT – addRoundKey

- Round key $K_i = \kappa^i_{63} \dots \kappa^i_0$, $(1 \leq i \leq 32)$
- Current state $b_{63} b_{62} \dots b_0$
- For $0 \leq j \leq 63$

$$b_j \rightarrow b_j \oplus \kappa^i_j$$

PRESENT – sBoxLayer

- sBoxLayer applies sixteen $4-$bit Sboxes to each nibble of the current cipher state.
- For example, if the input is 0, the output is C.

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

PRESENT – pLayer

pLayer permutes the 64 bits using the following formula:

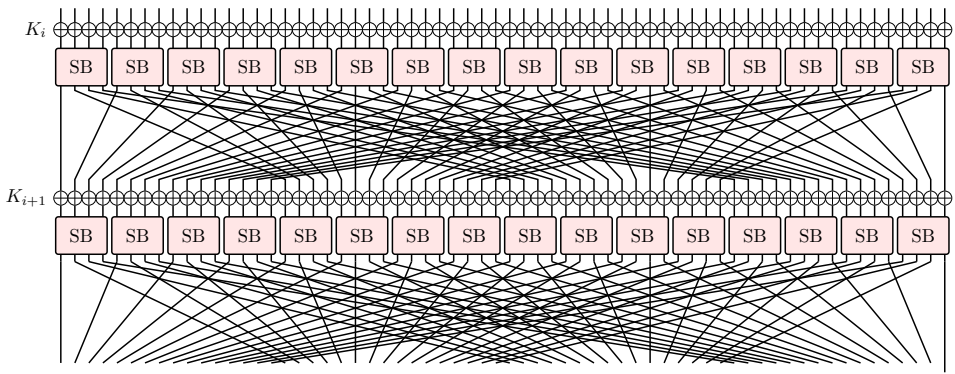
$$\text{pLayer}(j) = \left\lfloor \frac{j}{4} \right\rfloor + (j \bmod 4) \times 16,$$

where j denotes the bit position.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	16	32	48	1	17	33	49	2	18	34	50	3	19	35	51
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
4	20	36	52	5	21	37	53	6	22	38	54	7	23	39	55
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
8	24	40	56	9	25	41	57	10	26	42	58	11	27	43	59
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
12	28	44	60	13	29	45	61	14	30	46	62	15	31	47	63

Two rounds of PRESENT

- For our DPA attacks, we will attack the 0th Sbox and try to recover one nibble of the first round key



DPA attack – step 1

Identify the target cryptographic implementation

- DPA attacks can be applied to unprotected implementations of any symmetric block cipher that has been proposed so far.

Example

As a running example, we will look at the computation of PRESENT.

DPA attack – step 2

Experimental setup and measure leakages

- The efficiency and success of the attack are highly dependent on the measurement devices the attacker has access to.
- Suppose we have taken measurements of the target implementation with M_p plaintexts.
- For $j = 1, \ldots, M_p$, let $\boldsymbol{\ell}_j = (l_1^j, l_2^j, \ldots, l_q^j)$ denote the corresponding power trace, where q is the total number of time samples in one trace.

Example

We will use the *Random plaintext dataset* as illustrations.

- Each trace has 3600 time samples
- Contains 5000 traces with a fixed round key `0xFEDCBA0123456789` and a random plaintext for each trace.
- In particular, we have $M_p = 5000$, $q = 3600$.

DPA attack – step 3

Choose the part of the key to recover

- DPA attack is normally carried out in a divide-and-conquer manner.
- We focus on a small part (e.g. a nibble, a byte) of a round key in each attack and each part of the round key can be recovered independently.
- With the inverse key schedule, one (e.g. AES) or two round keys (e.g. PRESENT, DES) will reveal the master key
- Let k denote the target part of the key and let M_k denote the number of possible values of k.

Example

For our attack example, we will focus on the 0th nibble of the first round key for PRESENT and $M_k = 16$.

DPA attack – step 4

Choose the target intermediate value

- To recover the key, we exploit relationships between leakages and a certain intermediate value being processed in the DUT.
- The goal is to gain information about this intermediate value, which reveals information about our chosen part of the key.
- Let $\boldsymbol{v}$ denote the target intermediate value.
- We require that there is a function φ, such that

$$\boldsymbol{v} = \varphi(k, p),$$

where p denotes (part of) the plaintext.

Example

- k: 0th nibble of the first round key
- $\boldsymbol{v}$: 0th Sbox output of the first round
- p: 0th nibble of the plaintext
- What is φ?

DPA attack – step 4

Choose the target intermediate value

- To recover the key, we exploit relationships between leakages and a certain intermediate value being processed in the DUT.
- The goal is to gain information about this intermediate value, which reveals information about our chosen part of the key.
- Let $\boldsymbol{v}$ denote the target intermediate value.
- We require that there is a function φ, such that

$$\boldsymbol{v} = \varphi(k, p),$$

where p denotes (part of) the plaintext.

Example

- k: 0th nibble of the first round key
- $\boldsymbol{v}$: 0th Sbox output of the first round
- p: 0th nibble of the plaintext

$$\boldsymbol{v} = \mathsf{SB}_{\mathsf{PRESENT}}(k \oplus p),$$

DPA attack – step 5

Compute hypothetical target intermediate values

- A small part of the key is related to our target intermediate value.
- We can make a guess of this part of the key and obtain a hypothetical value for our target intermediate value.
- In particular, for each key hypothesis $\hat{k}_i$ of k, and each (part of the) plaintext p_j, we can compute a hypothesis for $\boldsymbol{v}$, which is given by

$$\hat{\boldsymbol{v}}_{ij} = \varphi(\hat{k}_i, p_j), \quad i = 1, 2, \ldots, M_k, \quad j = 1, 2, \ldots, M_p.$$

Example

- For our attacks, with each key hypothesis of the 0th nibble of the first round key, we have a hypothetical value for the 0th Sbox output:

$$\hat{\boldsymbol{v}}_{ij} = \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_i \oplus p_j), \quad i = 1, 2, \ldots, 16, \quad j = 1, 2, \ldots, 5000.$$

- p_j is the 0th nibble of the plaintext corresponding to the attack trace $\boldsymbol{\ell}_j$.
- We set $\hat{k}_i = i - 1, \quad i = 1, 2, \ldots, 16.$

DPA attack – step 5

Compute hypothetical target intermediate values

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

Example

For our attacks, with each key hypothesis of the 0th nibble of the first round key, we have a hypothetical value for the 0th Sbox output:

$$\hat{\boldsymbol{v}}_{ij} = \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_i \oplus p_j), \quad i = 1, 2, \ldots, 16, \quad j = 1, 2, \ldots, 5000.$$

We set $\hat{k}_i = i - 1, \quad i = 1, 2, \ldots, 16.$

- $\hat{k}_1 = 0$, $\hat{k}_2 = 1$.
- For *Random plaintext dataset*, we have $p_1 = 9, \quad p_2 = \mathtt{C}$.
- $\hat{\boldsymbol{v}}_{ij} =? \; i = 1, 2, \; j = 1, 2$

DPA attack – step 5

Compute hypothetical target intermediate values

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
C	5	6	B	9	0	A	D	3	E	F	8	4	7	1	2

Example

$$\hat{\boldsymbol{v}}_{ij} = \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_i \oplus p_j), \quad i = 1, 2, \ldots, 16, \quad j = 1, 2, \ldots, 5000.$$

- $\hat{k}_1 = \mathtt{0}$, $\hat{k}_2 = \mathtt{1}$.
- $p_1 = \mathtt{9}, \quad p_2 = \mathtt{C}$.

$$\begin{aligned}
\hat{\boldsymbol{v}}_{11} &= \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_1 \oplus p_1) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{0} \oplus \mathtt{9}) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{9}) = \mathtt{E}, \\
\hat{\boldsymbol{v}}_{12} &= \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_1 \oplus p_2) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{0} \oplus \mathtt{C}) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{C}) = \mathtt{4}, \\
\hat{\boldsymbol{v}}_{21} &= \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_2 \oplus p_1) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{1} \oplus \mathtt{9}) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{8}) = \mathtt{3}, \\
\hat{\boldsymbol{v}}_{22} &= \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_2 \oplus p_2) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{1} \oplus \mathtt{C}) = \mathsf{SB}_{\mathsf{PRESENT}}(\mathtt{D}) = \mathtt{7}.
\end{aligned}$$

DPA attack – step 6

Choose the leakage model

- For each hypothetical target intermediate value, we can compute the hypothetical signal depending on our leakage model

$$\mathcal{H}_{ij} := \mathcal{L}(\hat{\boldsymbol{v}}_{ij}) - \text{noise}, \quad i = 1, 2, \ldots, M_k, \quad j = 1, 2, \ldots, M_p,$$

- We subtract the noise component from the leakage model.

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 $\boldsymbol{v}$

DPA attack – step 6

Choose the leakage model

- For each hypothetical target intermediate value, we can compute the hypothetical signal depending on our leakage model

$$\mathcal{H}_{ij} := \mathcal{L}(\hat{\boldsymbol{v}}_{ij}) - \text{noise}, \quad i = 1, 2, \ldots, M_k, \quad j = 1, 2, \ldots, M_p,$$

- We subtract the noise component from the leakage model.
- Hamming weight leakage model

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

- If we choose the Hamming weight leakage model, we have

$$\mathcal{H}_{ij} = \mathrm{wt}\,(\hat{\boldsymbol{v}}_{ij}), \quad i = 1, 2, \ldots, M_k, \quad j = 1, 2, \ldots, M_p.$$

DPA attack – step 6

Choose the leakage model

- If we choose the Hamming weight leakage model, we have

$$\mathcal{H}_{ij} = \mathrm{wt}\left(\hat{\boldsymbol{v}}_{ij}\right), \quad i = 1, 2, \ldots, M_k, \quad j = 1, 2, \ldots, M_p.$$

Example

For our attacks, with each key hypothesis of the 0th nibble of the first round key, we have a hypothetical value for the 0th Sbox output:

$$\hat{\boldsymbol{v}}_{ij} = \mathrm{SB}_{\mathrm{PRESENT}}(\hat{k}_i \oplus p_j), \quad i = 1, 2, \ldots, 16, \quad j = 1, 2, \ldots, 5000.$$

We have computed that

$$\hat{\boldsymbol{v}}_{11} = \mathtt{E}, \quad \hat{\boldsymbol{v}}_{12} = 4, \quad \hat{\boldsymbol{v}}_{21} = 3, \quad \hat{\boldsymbol{v}}_{22} = 7.$$

What are $\mathcal{H}_{ij}$ according to the Hamming weight model? How about the identity leakage model?

DPA attack – step 6

Choose the leakage model

- For each hypothetical target intermediate value, we can compute the hypothetical signal depending on our leakage model

$$\mathcal{H}_{ij} := \mathcal{L}(\hat{\boldsymbol{v}}_{ij}) - \text{noise}, \quad i = 1, 2, \ldots, M_k, \quad j = 1, 2, \ldots, M_p,$$

- We subtract the noise component from the leakage model.

Example

$$\hat{\boldsymbol{v}}_{11} = \mathtt{E}, \quad \hat{\boldsymbol{v}}_{12} = 4, \quad \hat{\boldsymbol{v}}_{21} = 3, \quad \hat{\boldsymbol{v}}_{22} = 7.$$

According to the Hamming weight leakage model:

$$\mathcal{H}_{11} = \mathrm{wt}\,(\mathtt{E}) = 3, \quad \mathcal{H}_{12} = \mathrm{wt}\,(4) = 1, \quad \mathcal{H}_{21} = \mathrm{wt}\,(3) = 2, \quad \mathcal{H}_{22} = \mathrm{wt}\,(7) = 3$$

According to the identity leakage model:

$$\mathcal{H}_{11} = \mathtt{E} = 14, \quad \mathcal{H}_{12} = 4 = 4, \quad \mathcal{H}_{21} = 3 = 3, \quad \mathcal{H}_{22} = 7 = 7$$

What we have done so far

Example

- For our attacks, we aim to recover the 0th nibble of the first round key for PRESENT encryption, denoted k
- The target intermediate value is the 0th Sbox output
- 5000 measurements, corresponding 0th nibble of plaintext is p_j
- With each key hypothesis of k, we have a hypothetical value for $\boldsymbol{v}$:

$$\hat{\boldsymbol{v}}_{ij} = \mathsf{SB}_{\mathsf{PRESENT}}(\hat{k}_i \oplus p_j), \quad i = 1, 2, \ldots, 16, \quad j = 1, 2, \ldots, 5000.$$

- With a chosen leakage model, we have a hypothetical signal
  - Hamming weight leakage model: $\mathcal{H}_{ij} = \mathrm{wt}\,(\hat{\boldsymbol{v}}_{ij})$
  - Identity leakage mode: $\mathcal{H}_{ij} = \hat{\boldsymbol{v}}_{ij}$

DPA attack – step 7

Statistical analysis

- For a fixed key hypothesis $\hat{k}_i$, we view the modeled signal as a random variable $\mathcal{H}_i$ that varies when the plaintext changes.
- If we fix a time sample t, we also consider the leakage at this time as a random variable L_t.
- Then a sample for this pair of random variable $(\mathcal{H}_i, L_t)$ is given by

$$\left\{ (\mathcal{H}_{ij}, l_t^j) \;\middle|\; j = 1, 2, \ldots, M_p \right\}.$$

- We would like to know how good the modeled signals are compared to the actual leakages for each key hypothesis.
- For the correct key hypothesis and the time samples corresponding to POIs, we expect the modeled signals to be correlated to the real leakages.

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 $X_i - \overline{X}$ and $Y_i - \overline{Y}$ will tend to be the same. And the value of ρ will be bigger.
- If when X is large (or small), Y is small (or large), then the signs of $X_i - \overline{X}$ and $Y_i - \overline{Y}$ 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 $\rho = 0$

Sample correlation coefficient – Definition

Correlation coefficient:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Definition

We define the *sample correlation coefficient*, denoted by r, as follows:

$$\begin{aligned} r &= \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\sqrt{S_x^2 S_y^2}} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2\right)\left(\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2\right)}} \\ &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2}}. \end{aligned}$$

DPA attack – step 7

Statistical analysis

- We adopt the notion of correlation coefficient.
- For each key hypothesis $\hat{k}_i$ $(i = 1, 2, \ldots, M_k)$ and each time sample t $(t = 1, 2, \ldots, q)$, we compute the sample correlation coefficient, denoted by $r_{i,t}$, of $\mathcal{H}_i$ and L_t:

$$r_{i,t} := \frac{\sum_{j=1}^{M_p} (\mathcal{H}_{ij} - \overline{\mathcal{H}_i})(l_t^j - \overline{l_t})}{\sqrt{\sum_{j=1}^{M_p} (\mathcal{H}_{ij} - \overline{\mathcal{H}_i})^2} \sqrt{\sum_{j=1}^{M_p} (l_t^j - \overline{l_t})^2}},$$

  where

$$\overline{\mathcal{H}_i} = \frac{1}{M_p} \sum_{j=1}^{M_p} \mathcal{H}_{ij}.$$

  is the average of all hypothetical leakages for the same key hypothesis $\hat{k}_i$
- $\overline{l_t}$ can be obtained by taking all the leakages at time sample t from all traces and compute the average

DPA attack – step 7

Statistical analysis

- We adopt the notion of correlation coefficient.
- For each key hypothesis $\hat{k}_i$ $(i = 1, 2, \ldots, M_k)$ and each time sample t $(t = 1, 2, \ldots, q)$, we compute the sample correlation coefficient, denoted by $r_{i,t}$, of $\mathcal{H}_i$ and L_t:

$$r_{i,t} := \frac{\sum_{j=1}^{M_p} (\mathcal{H}_{ij} - \overline{\mathcal{H}_i})(l_t^j - \overline{l_t})}{\sqrt{\sum_{j=1}^{M_p} (\mathcal{H}_{ij} - \overline{\mathcal{H}_i})^2} \sqrt{\sum_{j=1}^{M_p} (l_t^j - \overline{l_t})^2}}.$$

- In our case,

$$r_{i,t} = \frac{\sum_{j=1}^{5000} (\mathcal{H}_{ij} - \overline{\mathcal{H}_i})(l_t^j - \overline{l_t})}{\sqrt{\sum_{j=1}^{5000} (\mathcal{H}_{ij} - \overline{\mathcal{H}_i})^2} \sqrt{\sum_{j=1}^{5000} (l_t^j - \overline{l_t})^2}}, \quad i = 1, 2, \ldots, 16, \quad t = 1, 2, \ldots, 3600.$$

DPA attack – step 7

- The target intermediate value $\boldsymbol{v}$ will be processed in our DUT at certain points in time – the leakages at those corresponding time samples (POIs) should be correlated to $\boldsymbol{v}$.
- If a good leakage model (i.e. a model that is close to the actual leakage of the DUT) is chosen, we expect $\mathcal{H}_i$ and L_t to be correlated for the correct key hypothesis $\hat{k}_i$ and POIs t.
- The key hypothesis that achieves the highest *absolute value* of $r_{i,t}$ is expected to be the correct key.
- The time samples that achieve higher absolute values of $r_{i,t}$ will be our POIs in the attack.

Remark

In practice, if all r_{it}s are low, we will need more traces for the attack.

DPA attack results – identify leakage model

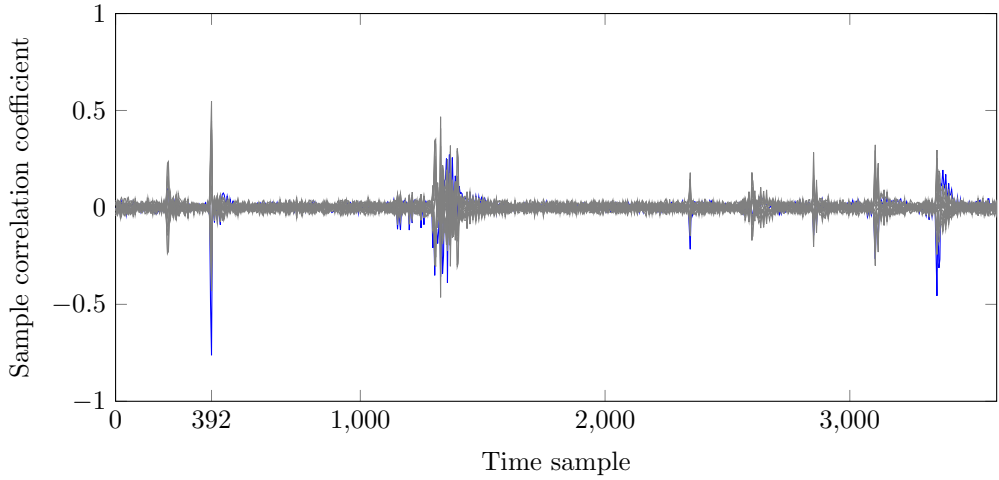


Figure: Sample correlation coefficients $r_{i,t}$ $(i = 1, 2, \ldots, 16)$ for all time samples $t = 1, 2, \ldots, 3600$. Computed with the identity leakage model and *Random plaintext dataset*. The blue line corresponds to the correct key hypothesis $\hat{k}_{10} = 9$.

DPA attack results – Hamming weight leakage model

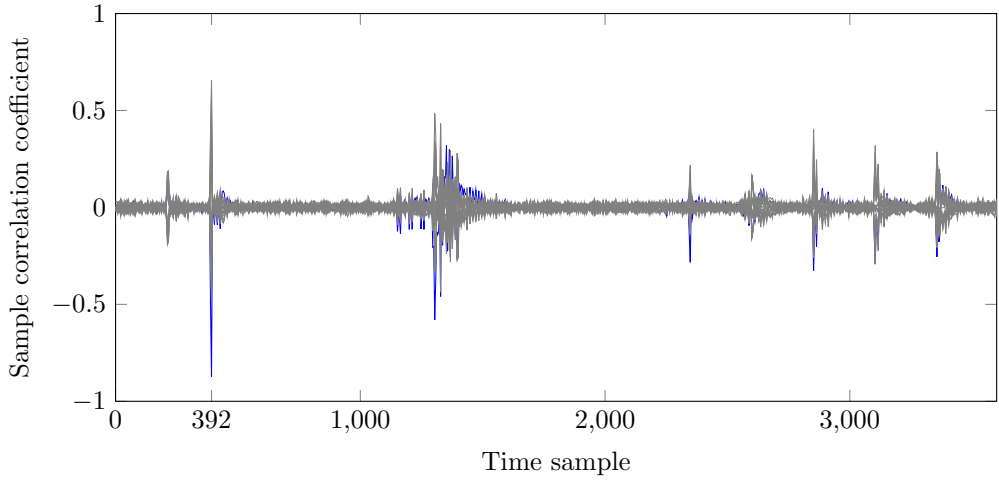


Figure: Sample correlation coefficients $r_{i,t}$ $(i = 1, 2, \ldots, 16)$ for all time samples $t = 1, 2, \ldots, 3600$. Computed with the Hamming leakage model and *Random plaintext dataset*. The blue line corresponds to the correct key hypothesis $\hat{k}_{10} = 9$.

Remarks

- The maximum absolute value of the correlation coefficient calculated using the identity leakage model is 0.7624 for the correct key hypothesis, obtained at time sample 392
- The maximum absolute value of the correlation coefficient calculated using the Hamming weight leakage model is 0.8747 for the correct key hypothesis, obtained at time sample 392
- The higher absolute correlation coefficient value for the Hamming weight leakage model indicates it is probably a better leakage model for our DUT compared to the identity leakage model
- We have seen last week that 392 is the time sample that "leaks" the most according to TVLA

Remarks

- Some peaks of $r_{i,t}$ corresponding to a wrong key hypothesis and the correct key hypothesis appear at similar time samples – $\mathcal{H}_i$s are not independent random variables and for those time samples t, $\mathcal{H}_i$s also correlated with the actual leakage
- The correlation between $\mathcal{H}_i$s also influences the magnitude of the correlation coefficients for the wrong key hypotheses. If the correlation between $\mathcal{H}_i$s is higher, we would also see higher peaks in some wrong key hypotheses. For the PRESENT cipher, correlations among outputs from the initial `addRoundKey` operation are stronger than those between outputs of the initial `sBoxLayer`. Therefore, in step 4, we chose the target intermediate value to an Sbox output.
- The attacks we have seen recover one nibble of the first-round key. The other nibbles can be recovered independently with a similar method using the same traces.

Code

Implementation for attack using the identity leakage model can be found here

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

SPA and non-profiled DPA

- Simple Power Analysis (SPA)
- Sample Correlation Coefficient
- Non-profiled DPA Attacks on Symmetric Block Ciphers
- Signal-to-Noise Ratio

Recall – leakages

- Since by analyzing the power consumption, we can deduce the secret key, 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.
- 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

Definition

- Signal-to-noise ratio (SNR) is commonly used in electrical engineering and signal processing, the general definition is

$$\text{SNR} = \frac{\text{Var(signal)}}{\text{Var(noise)}},$$

- In our case, for a fixed time sample t, X_t is the signal, which is part of the leakage relevant to our attack.
- And we define

$$\text{SNR} = \frac{\text{Var}(X_t)}{\text{Var}(N_t)}.$$

- $\text{Var}(X_t)$ measures how much the leakage varies at time sample t due to the signal.
- $\text{Var}(N_t)$ measures how much the leakage varies due to the noise.
- SNR quantifies how much information is leaked at time sample t from the measurements.
- The higher the SNR, the lower the noise.

SNR – Example

Example

- Suppose we are interested in the Hamming weight of an $8-$bit intermediate value at time sample t.
- In particular, the intermediate value we would like to analyze is from $\mathbb{F}_2^8$.
- We further assume that the leakage L_t is given by the Hamming weight model
- Thus $X_t = \mathrm{wt}(\boldsymbol{v})$ for $\boldsymbol{v} \in \mathbb{F}_2^8$
- Then the variance of the signal is given by $\mathrm{Var}(\mathrm{wt}(\boldsymbol{v}))$ for $\boldsymbol{v} \in \mathbb{F}_2^8$.
- Recall that 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, $\mathrm{E}[X]$.

SNR – Example

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, $\mathrm{E}[X]$.

Example

- The variance of the signal is given by $\mathrm{Var}(\mathrm{wt}(\boldsymbol{v}))$ for all $\boldsymbol{v} \in \mathbb{F}_2^8$

$$\mathrm{E}\left[\mathrm{wt}(\boldsymbol{v})\right] = \frac{1}{|\mathbb{F}_2^8|} \sum_{\boldsymbol{v} \in \mathbb{F}_2^8} \mathrm{wt}(\boldsymbol{v}) = \frac{1}{2^8} \sum_{i=1}^{8} i \binom{8}{i} = ?$$

SNR – Example

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, $\mathrm{E}[X]$.

Example

$$\begin{aligned}
\mathrm{E}\left[\mathrm{wt}\left(\boldsymbol{v}\right)\right] &= \frac{1}{|\mathbb{F}_2^8|}\sum_{\boldsymbol{v}\in\mathbb{F}_2^8}\mathrm{wt}\left(\boldsymbol{v}\right) = \frac{1}{2^8}\sum_{i=1}^{8} i\binom{8}{i} = \frac{1}{2^8}\sum_{i=1}^{8}\frac{8!}{(i-1)!(8-i)!} \\
&= \frac{8}{2^8}\sum_{i=1}^{8}\frac{7!}{(i-1)!(7-(i-1))!} = \frac{8}{2^8}\sum_{j=0}^{7}\binom{7}{j} = \frac{8\times 2^7}{2^8} = 4.
\end{aligned}$$

$$\mathrm{E}\left[\mathrm{wt}\left(\boldsymbol{v}\right)^2\right] = ?$$

SNR – Example

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

Example

$$\begin{aligned}
\mathrm{E}\left[\mathrm{wt}\left(\boldsymbol{v}\right)^2\right] &= \frac{1}{|\mathbb{F}_2^8|}\sum_{\boldsymbol{v}\in\mathbb{F}_2^8}\mathrm{wt}\left(\boldsymbol{v}\right)^2 = \frac{1}{2^8}\sum_{i=1}^{8} i^2\binom{8}{i} = \frac{1}{2^8}\sum_{i=1}^{8} i\frac{8!}{(i-1)!(8-i)!} \\
&= \frac{8}{2^8}\left(\sum_{i=1}^{8}(i-1)\frac{7!}{(i-1)!(8-i)!} + \sum_{i=1}^{8}\frac{7!}{(i-1)!(7-(i-1))!}\right) \\
&= \frac{1}{2^5}\left(7\sum_{i=2}^{8}\frac{6!}{(i-2)!(6-(i-2))!} + \sum_{j=0}^{7}\binom{7}{j}\right) = \frac{1}{2^5}\left(7\sum_{j=0}^{6}\binom{6}{j} + 2^7\right) \\
&= \frac{1}{2^5}(7\times 2^6 + 2^7) = 7\times 2 + 2^2 = 18.
\end{aligned}$$

$\mathrm{E}\left[\mathrm{wt}\left(\boldsymbol{v}\right)\right] = 4.$, $\mathrm{Var}(X_t) = \mathrm{Var}(\mathrm{wt}\left(\boldsymbol{v}\right)) = ?$

SNR – Example

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, $\mathrm{E}[X]$.

Example

- Assume the leakage L_t is given by the Hamming weight model
- Thus $X_t = \mathrm{wt}(\boldsymbol{v})$ for some $\boldsymbol{v} \in \mathbb{F}_2^8$

$$\mathrm{E}[\mathrm{wt}(\boldsymbol{v})] = 4, \quad \mathrm{E}\left[\mathrm{wt}(\boldsymbol{v})^2\right] = 18$$

$$\mathrm{Var}(X_t) = \mathrm{Var}(\mathrm{wt}(\boldsymbol{v})) = 18 - 4^2 = 2.$$

- Let σ^2 denote the variance of the noise N_t. We have

$$\mathsf{SNR} = \frac{\mathrm{Var}(X_t)}{\mathrm{Var}(N_t)} = \frac{2}{\sigma^2}.$$

SNR – Example

Example

- We are interested in the Hamming weight of an $8-$bit intermediate value at time sample t.
- In particular, the intermediate value we would like to analyze is from $\mathbb{F}_2^8$.
- Assume the leakage L_t is given by the Hamming weight model
- Thus $X_t = \mathrm{wt}(\boldsymbol{v})$ for some $\boldsymbol{v} \in \mathbb{F}_2^8$

$$\mathrm{E}[\mathrm{wt}(\boldsymbol{v})] = 4, \quad \mathrm{E}\left[\mathrm{wt}(\boldsymbol{v})^2\right] = 18$$

$$\mathrm{Var}(X_t) = \mathrm{Var}(\mathrm{wt}(\boldsymbol{v})) = 18 - 4^2 = 2.$$

- Let σ^2 denote the variance of the noise N_t. We have

$$\mathsf{SNR} = \frac{\mathrm{Var}(X_t)}{\mathrm{Var}(N_t)} = \frac{2}{\sigma^2}.$$

SNR – Example

Example

- *Random dataset*: This dataset contains 10000 traces with a random round key and a random plaintext for each trace.
- Suppose we are interested in the exact value of the 0th Sbox output in the first round of PRESENT. Let us denote this intermediate value by $\boldsymbol{v}$.
- Fix a time sample t
- Then signal X_t is given by the part of the leakage related to $\boldsymbol{v}$.
- To compute $\mathrm{Var}(X_t)$, we first divide the traces into 16 sets: $A_1, A_2, \ldots, A_{16}$, where A_i contains traces corresponding to $\boldsymbol{v} = i - 1$
- For a fixed value of $\boldsymbol{v}$, X_t is a constant
- Take $t = 600$.
- We compute the average of the trace leakages at $t = 600$ across each set, which can be considered as approximation of the signals

$$0.08212, \quad 0.08221, \quad 0.08209, \quad \ldots$$

SNR – Example

Example

- Take $t = 600$.
- We compute the average of the trace leakages at $t = 600$ across each set, which can be considered as approximation of the signals

$$0.08212, \quad 0.08221, \quad 0.08209, \quad \dots$$

- Then the variance of X_t is given by the variance of those average values

$$\mathrm{Var}(X_{600}) \approx 1.0088 \times 10^{-8}$$

- The noise in each trace at $t = 600$ is given by the leakage minus the corresponding average

$$\mathrm{Var}(N_t) \approx 6.4184 \times 10^{-6}$$

- $\mathrm{SNR}_{600} \approx 0.00157$
- The same computations can be done for other time samples

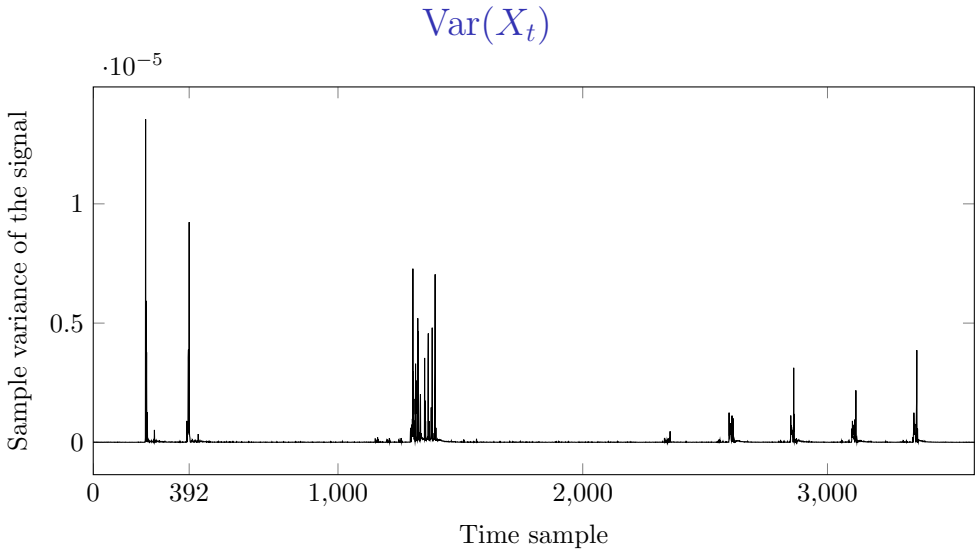


Figure: The signal is given by the exact value of the 0th Sbox output.

SNR

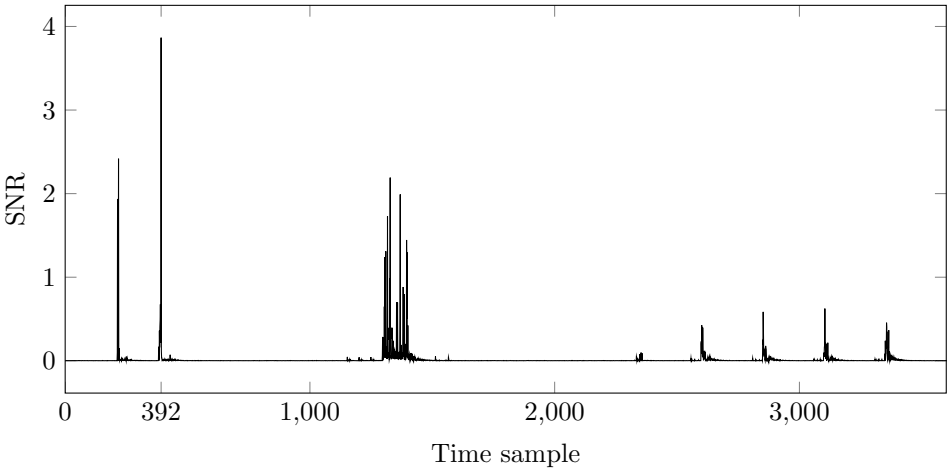
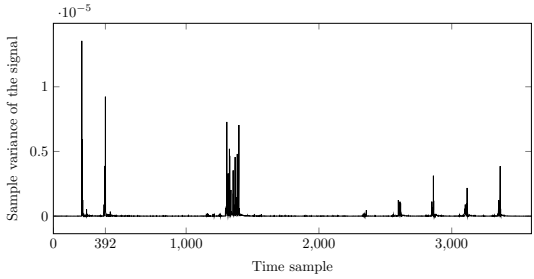
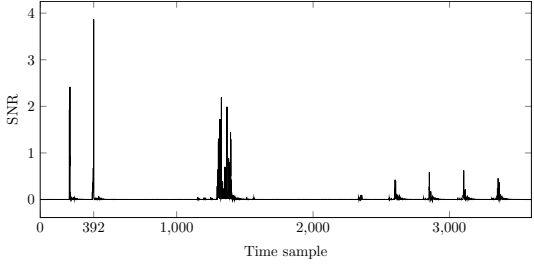


Figure: The signal is given by the exact value of the 0th Sbox output.

Comparison – $\mathrm{Var}(X_t)$ and SNR





Recall – averaged trace

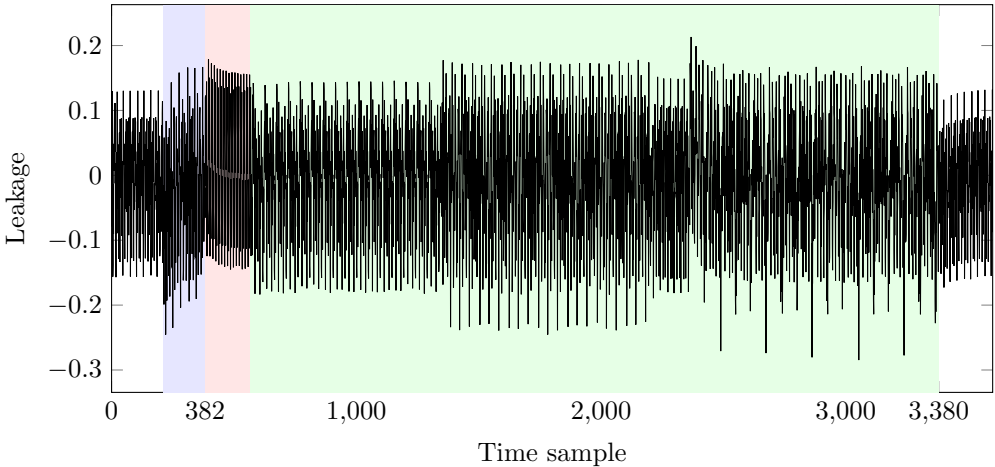


Figure: The averaged trace for 5000 traces from the *Fixed dataset A*. The blue, pink, and green parts of the trace correspond to addRoundKey, sBoxLayer, and pLayer respectively.

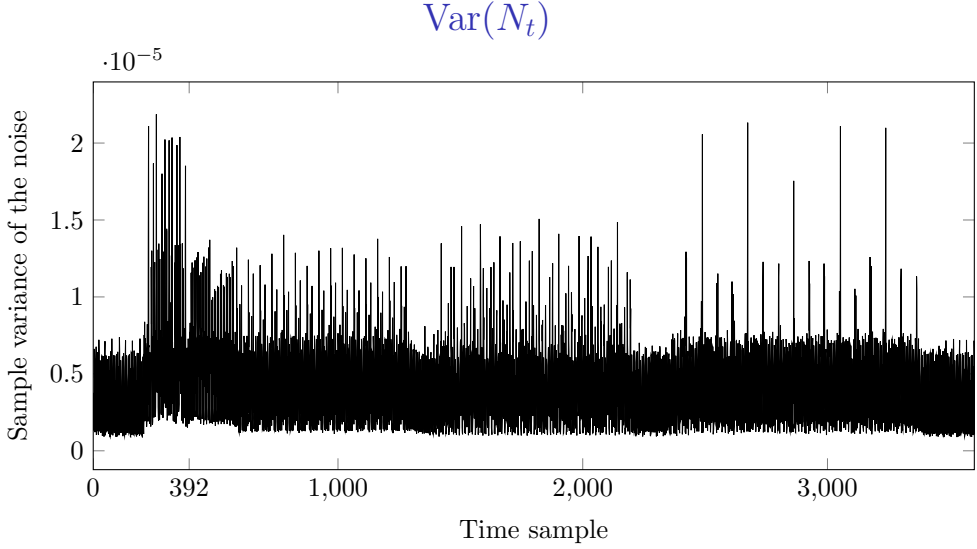
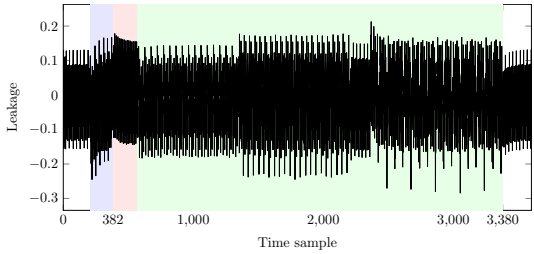
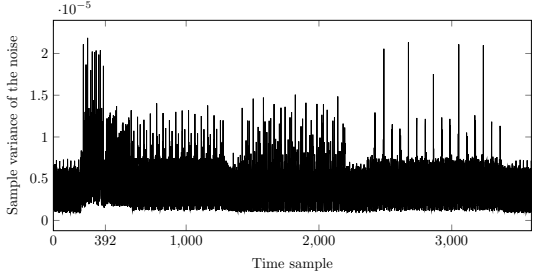


Figure: The signal is given by the exact value of the 0th Sbox output.

Noise and averaged trace





Observations

- The shape of variance of noise has similarities to one round of PRESENT computations – most of the leakage is not related to $\boldsymbol{v}$.
- The peaks for the variance of signal and SNR correspond to each other.
- The first two peaks are likely related to AddRoundKey and sBoxLayer.
- We can deduce that the peak at $t = 392$ is related to the 0th Sbox computation
- The peaks after 1000 are probably caused by the permutation of the 4 bits of $\boldsymbol{v}$, the 0th Sbox output.

SNR – Hamming weight

- With the same dataset, we can also focus on the Hamming weight of the 0th Sbox output.
- Instead of dividing the traces to 16 sets, we now have 5 sets corresponding to Hamming weight $0, 1, 2, 3, 4$.

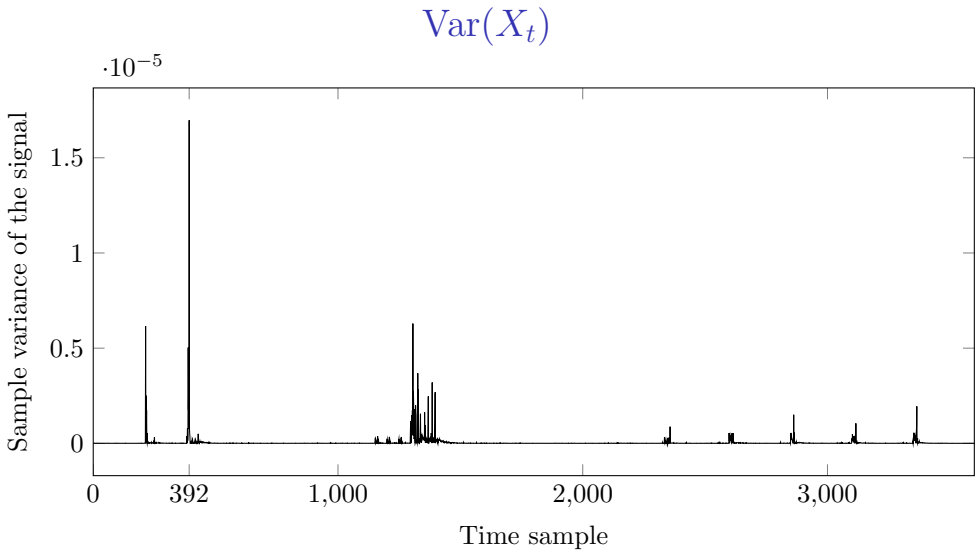


Figure: The signal is given by the Hamming weight of the 0th Sbox output.

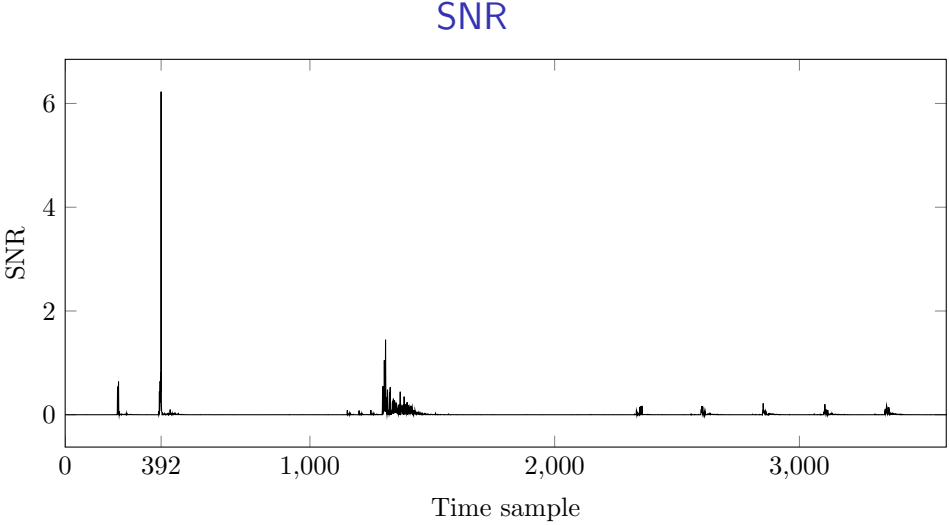


Figure: The signal is given by the Hamming weight of the 0th Sbox output.

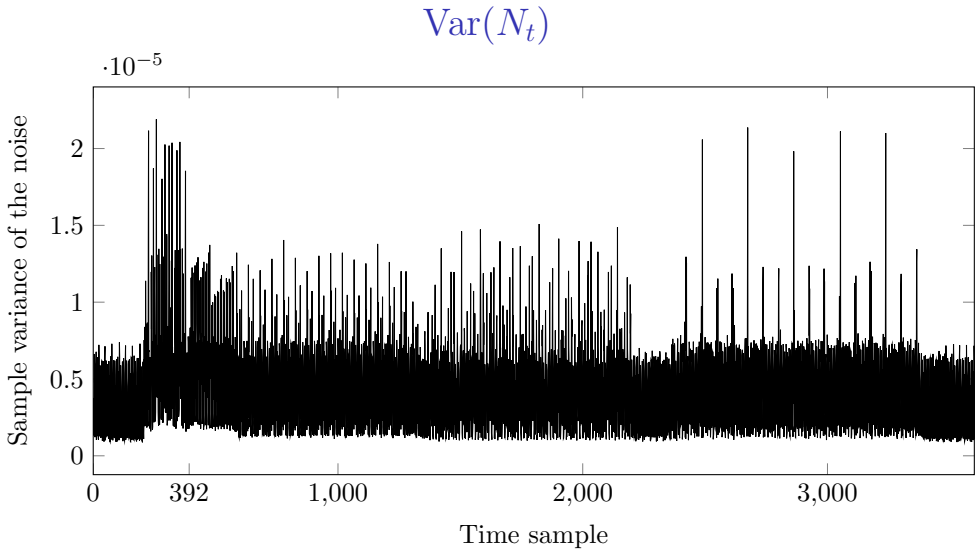
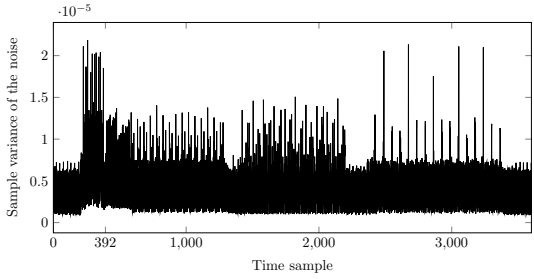
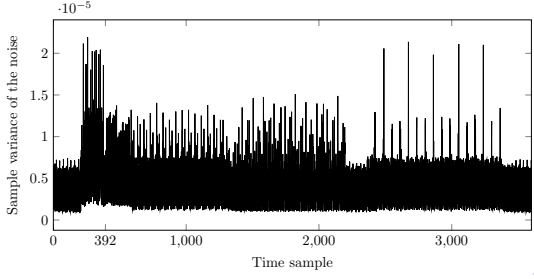


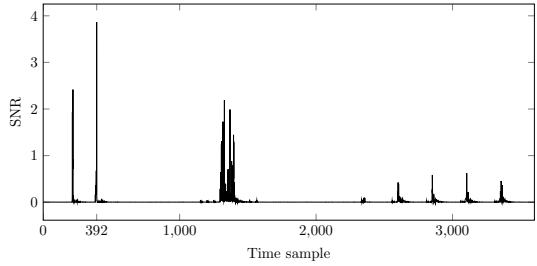
Figure: The signal is given by the Hamming weight of the 0th Sbox output.

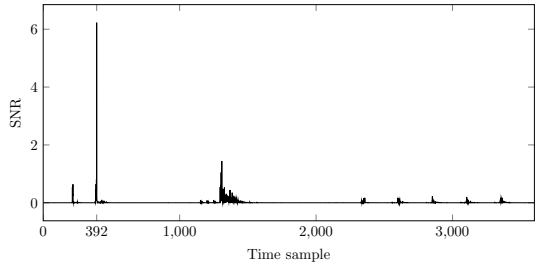
Comparison – variance of the noise





Comparison – SNR





Observations

- The sample variances of the noise are very similar and resemble the leakage of PRESENT computation
- The peaks in the variance of signal and SNR also correspond to each other.
- The locations of the peaks for SNR are similar.
- The highest peak in both SNR figures are at time sample 392.
  - This time sample corresponds to the computation of the 0th Sbox in sBoxLayer.
  - Higher SNR value at this point when we consider Hamming weight – the Hamming weight leakage model is closer to the device leakage than the identity leakage model

POIs

- Normally in DPA attacks, we would like to focus on time samples where the corresponding SNRs are high.
- We refer to those time samples as *points of interest (POIs)*.
- Signal given by exact value of $\boldsymbol{v}$
  - One POI $t = 392$.
  - Three POIs: $t = 392, 218, 1328$
- Signal given by the Hamming weight of $\boldsymbol{v}$
  - One POI $t = 392$.
  - Three POIs: $t = 392, 1309, 1304$

Assignment 4 A

- Grade 0 for late submission
- Start early
- Can utilize code from the Github repo
- Presentation of the code during Week 9 tutorial