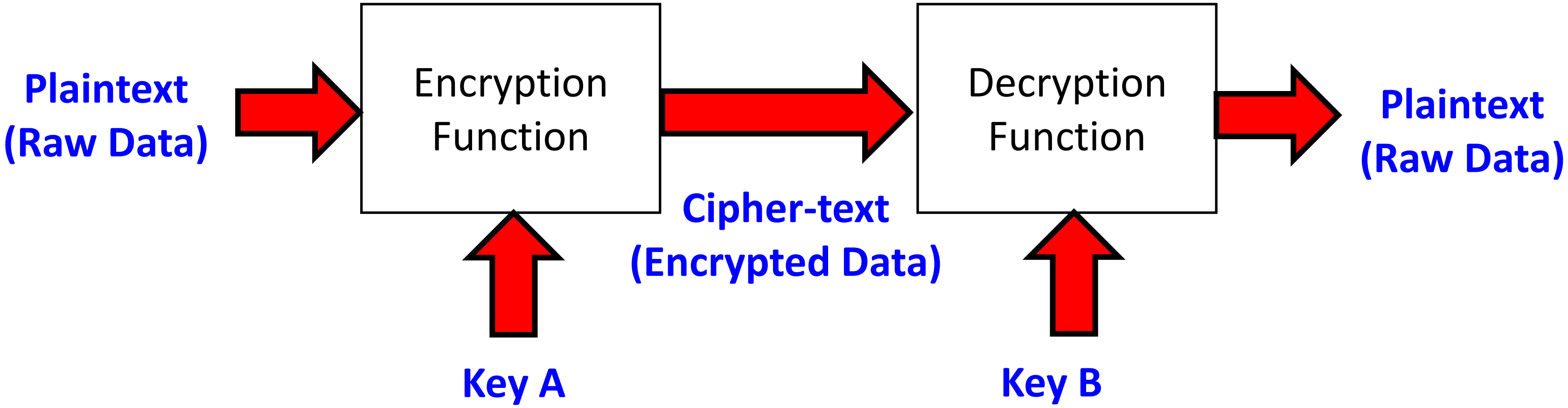
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Side-Channel Attack

A/P Gwee Bah Hwee

Encryption/Decryption

A typical encryption-decryption flow



Symmetrical encryption

When Key A = Key B = Key. Example: Advanced Encryption Standard (AES)

Asymmetrical encryption

When Key A ≠ Key B. Example: Elliptic Curve Cryptography (ECC)

Remember 3 definitions: Plaintext, Ciphertext and Key for encryption/decryption

Simple Encryption - Substitution

Why is the simple encryption not working?

Cipher-text (Encrypted Data)

Key

Plaintext (Raw Data)

C ogugcoah-nitgiunvg psblna sinvgounty, iciycie tgahirlrenacl sinvgounty, Uniecprog (its Uniecprog) hcu 33,000 sidgoeocdsctg cid pruteocdsctg utsdgitu ni thg arllgegu rf Gieniggonie, Bsuniguu, Uangiag, cid Hsmcintngu, Cotu cid Urancl Uangiagu, cid ntu eocdsctg arllgeg. its’u Lgg Krie ahnci Uahrrl rf Mgdnanig wcu gutcblnuhgd jrnitly wnth Nmpgoncl arllgeg Lridri.

Key

A research-intensive public university, Nanyang Technological University, Singapore (NTU Singapore) has 33,000 undergraduate and postgraduate students in the colleges of Engineering, Business, Science, and Humanities, Arts and Social Sciences, and its Graduate College. NTU’s Lee Kong Chian School of Medicine was established jointly with Imperial College London.

Answer: Data would be biased

# of ‘a’ = 25
# of ‘e’ = 37
# of ‘i’ = 28
# of ‘o’ = 17
# of ‘u’ = 12

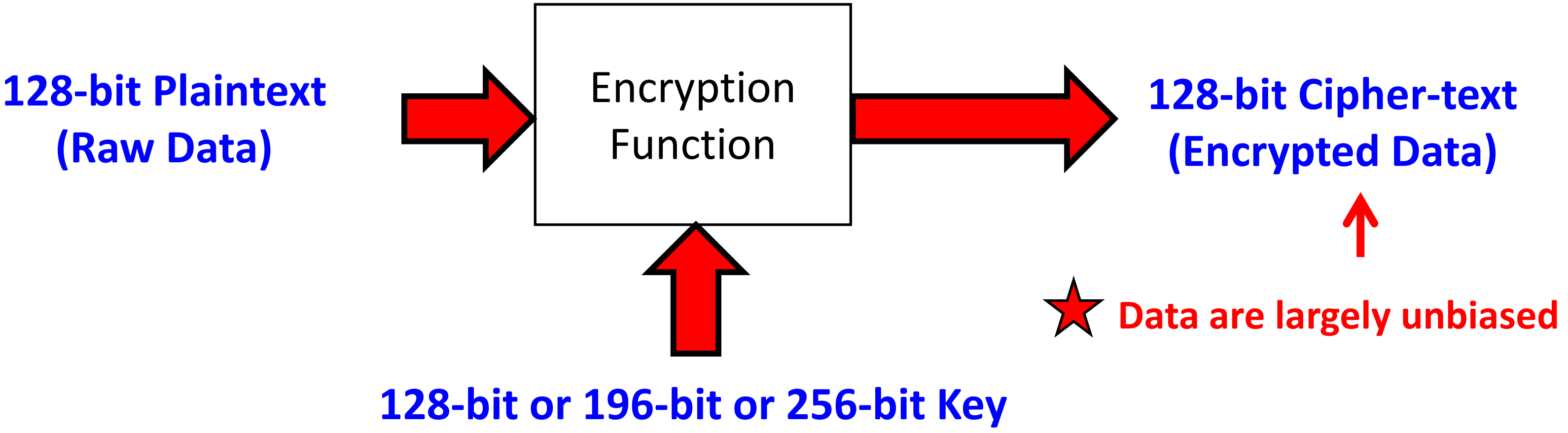
vs

# of ‘x’ = 0
# of ‘z’ = 0
# of ‘q’ = 0
# of ‘j’ = 1
# of ‘k’ = 1

All forms of biaising are bad to encryption/decryption

Advanced Encryption Standard

<u>Advanced Encryption Standard (AES) – a standard established by NIST (National Institute of Standards and Technology in 2001)</u>



128-bit Plaintext (Raw Data)

Encryption Function

128-bit Cipher-text (Encrypted Data)

★ **Data are largely unbiased**

128-bit or 196-bit or 256-bit Key

<u>Key size combinations</u>

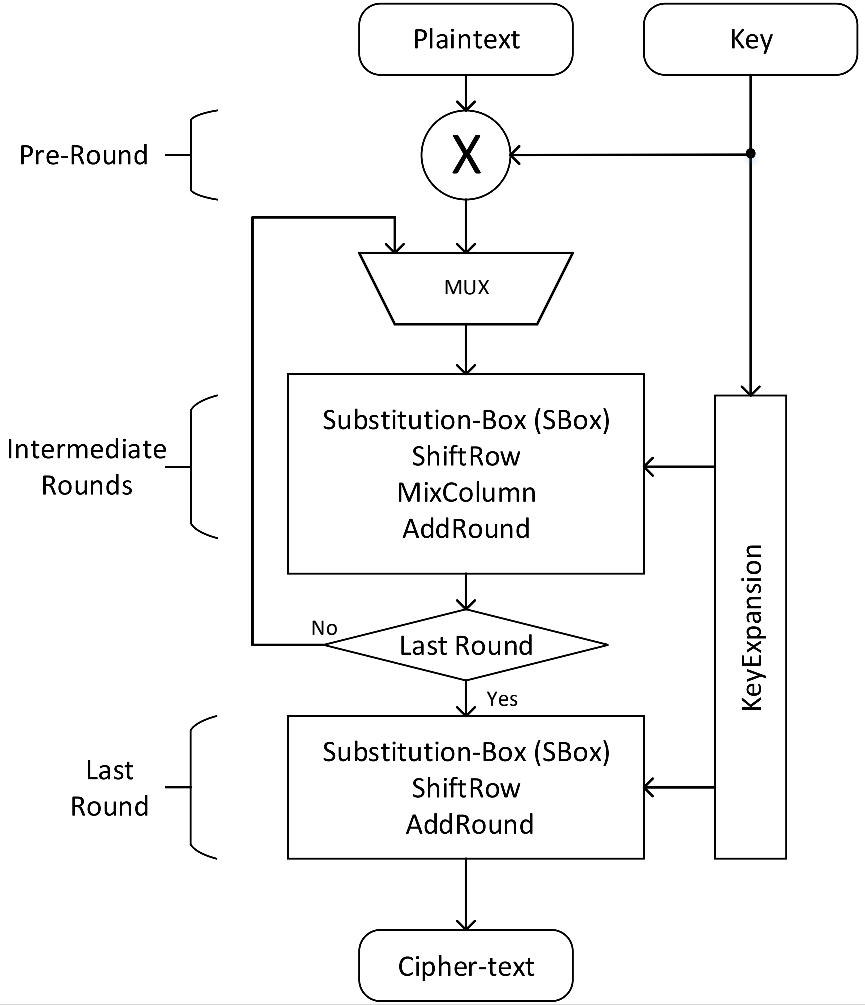
2^{128} = 3.40x10^{30} combinations
2^{196} = 1.00x10^{59} combinations
2^{256} = 1.16x10^{77} combinations

This is difficult to find out all the combinations now! (but not true in the future)

(i) Always follow "Standard" encryption; (ii) Large key size is more secure

AES

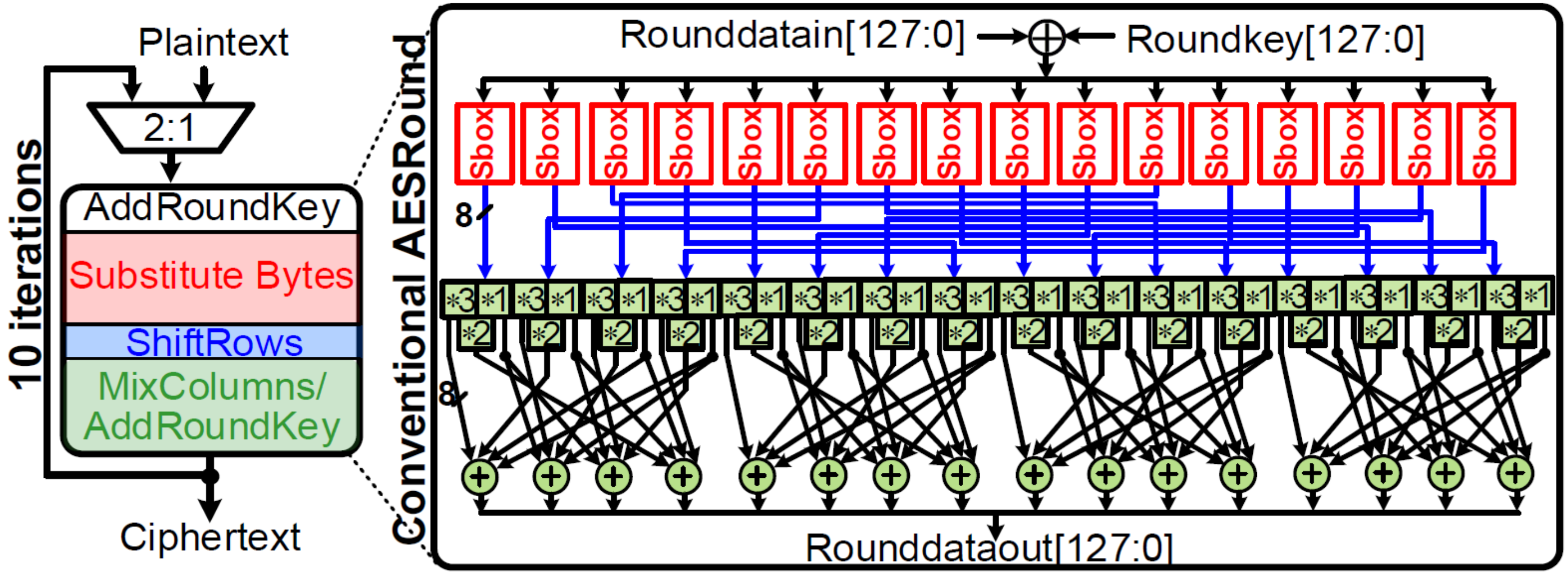
AES – an iterative process of de-biasing the encrypted data by using the substitution-Permutation network



Why SCA is possible for AES?

Simple answer:

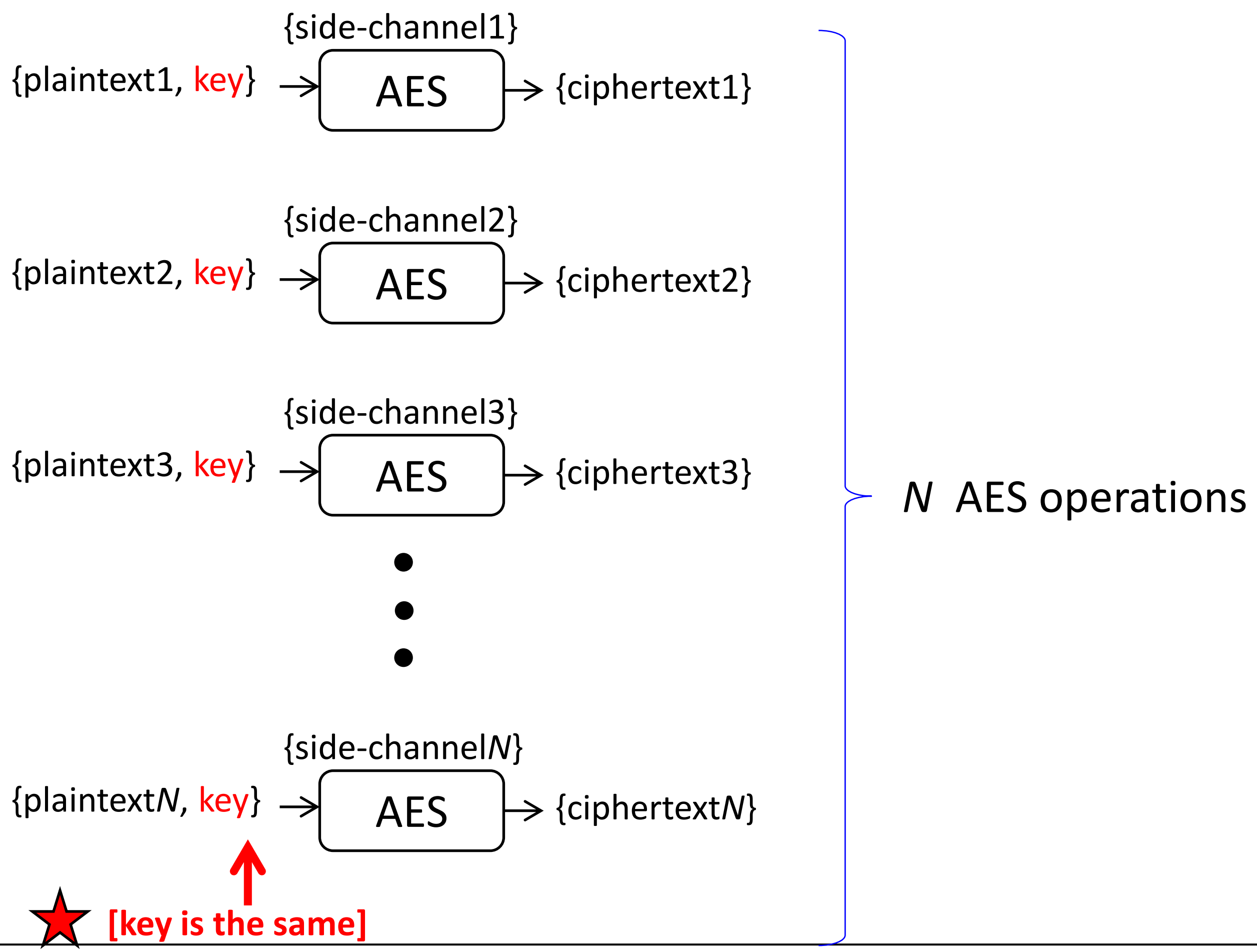
Leveraging on its 8-bit S-Box computations, we can try correlating an 8-bit data of a sub-block at a time (vis-a-vis 128-bit data of the device) by observing the SCA leakages. The search space now is low, i.e. 16×2^8 vs $2^{8\times16}$.



Note: For any guessing approach, try make the search space low with a linear relationship instead of an exponential relationship

SCA – Finding Statistical Properties

Let's look at *N* times of AES operations – any thing in common?



{side-channel1}

{plaintext1, key} → AES → {ciphertext1}

{side-channel2}

{plaintext2, key} → AES → {ciphertext2}

{side-channel3}

{plaintext3, key} → AES → {ciphertext3}

⋮

{side-channel*N*}

{plaintext*N*, key} → AES → {ciphertext*N*}

★ **[key is the same]**

N AES operations

Establish the statistical properties between (plaintext, side-channel) or (cipher-text, side-channel) to guess the correct key

SCA: Find Statistical Properties

Assumptions:

(1) You know/guess the algorithm (e.g. AES) but do not know the key

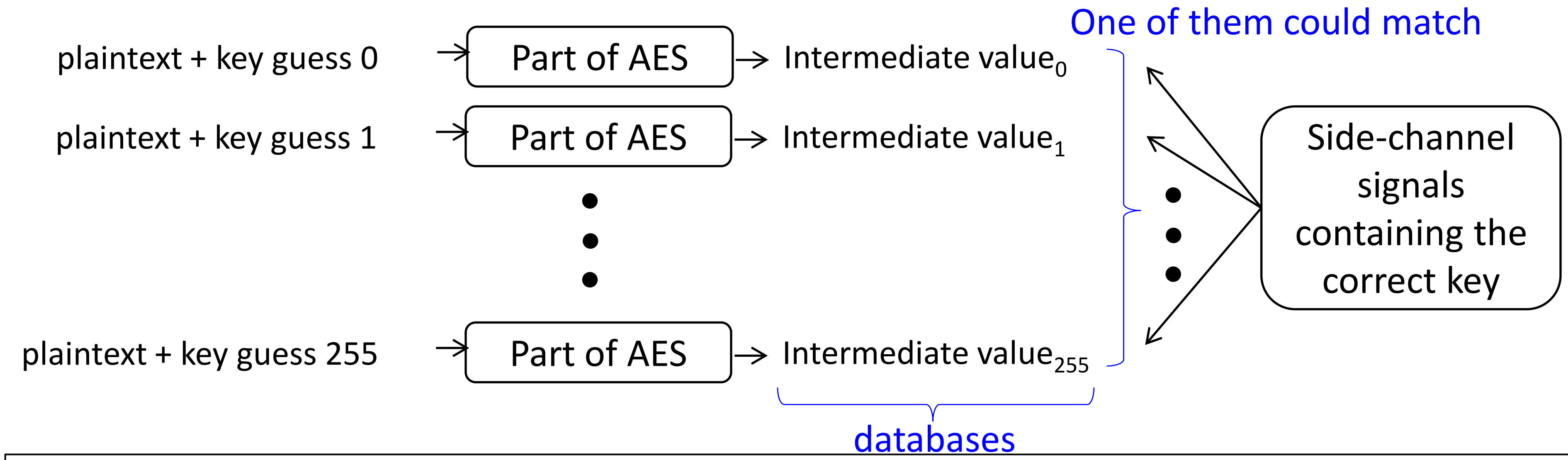
(2) You can interpret ciphertext or plaintext or both

(3) You can measure the side-channel signals

Principles:

(1) Form each database of an intermediate value (within AES) by considering all 256 key guess (i.e. 0 to 255) per sub-key byte

(2) The side-channel signals containing the correct key must match well with one of the databases



Remember the three assumptions (e.g. algorithm, data, and side-channel signals) for an attack

Differential Power Analysis (DPA)

Fundamental Principles

(1) Using one bit position (e.g., after S-box) as a target point

(2) Based on one plaintext, try all guess keys (from 0 to 255 per bloc) to check if the target point is “1” or “0”

(3) Repeat (2) for other plaintexts to build the databases with the target point “1” or “0” for each guess keys

(4) Based on each guess key, sum all the side-channels having the reference point “1” as Sum+, and that having the reference point “0” as Sum-

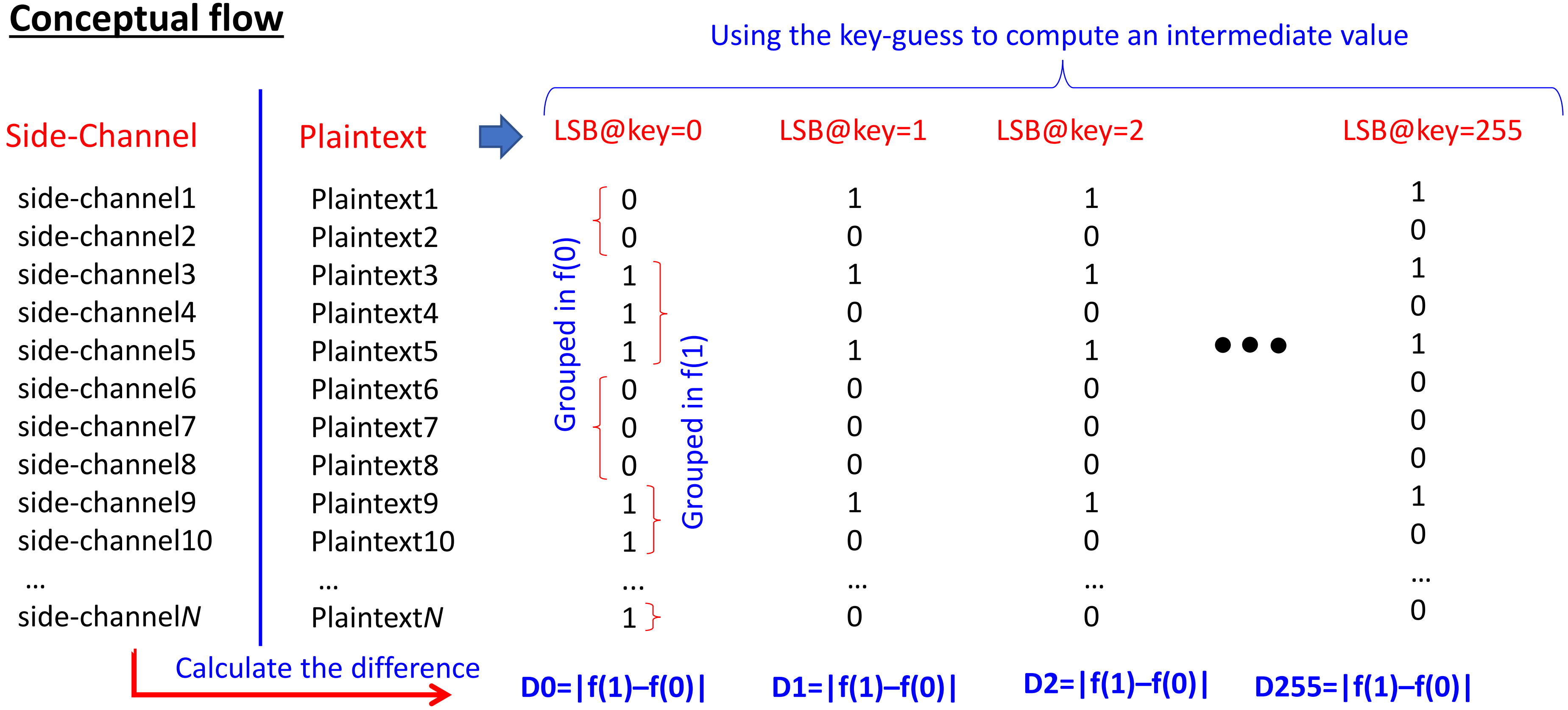
(5) Following (4), subtract between Sum+ and Sum- to be Diff (absolute number)

(6) If the guess key is not correct, Diff would be small (averaging effect)

(7) Compare all Diffs, and the guess key with the highest Diff would likely to the correct key

Differential Power Analysis (DPA)

Conceptual flow



Using the key-guess to compute an intermediate value

Side-Channel	Plaintext	LSB@key=0	LSB@key=1	LSB@key=2		LSB@key=255
side-channel1	Plaintext1	0	1	1		1
side-channel2	Plaintext2	0	0	0		0
side-channel3	Plaintext3	1	1	1		1
side-channel4	Plaintext4	1	0	0		0
side-channel5	Plaintext5	1	1	1	•••	1
side-channel6	Plaintext6	0	0	0		0
side-channel7	Plaintext7	0	0	0		0
side-channel8	Plaintext8	0	0	0		0
side-channel9	Plaintext9	1	1	1		1
side-channel10	Plaintext10	1	0	0		0
…	…	…	…	…		…
side-channel*N*	Plaintext*N*	1	0	0		0
		D0=|f(1)–f(0)|	D1=|f(1)–f(0)|	D2=|f(1)–f(0)|		D255=|f(1)–f(0)|

Calculate the difference

DPA - Discussion

Points to take note

(1) Noise could make the analysis difficult

(2) Traces may not be sufficiently large

(3) One corrupted side-channel signal may corrupt the results

(4) Countermeasures may make the evaluation difficult in view of noise, corrupted side-channel signals, and ghost peaks

(5) Pre-processing techniques may help (see later in Session 5)

If DPA does not work, do we have other choice?
Think about other statistical properties of the intermediate data

Common SCA Hypothesis Models

Power Model

$$P_{Total} = \underbrace{P_{op} + P_{data}} + P_{el.noise} + P_{const}$$

Question: How do we find these dependency?

Answer: Link these dependencies with SCA hypothesis models

No	Common SCA Hypothesis Model	Brief Explanation
1	Hamming Distance – Bus-wise	To count the number of transitions
2	Hamming Distance – Bit-wise	
3	Hamming Weight – Bus-wise	To count the number of non-zero items
4	Hamming Weight – Bit-wise	
5	Zero Value Model	To evaluate the switching activity when the data is equal to zero

Establishing a good SCA hypothesis model to match the data/operation dependency is a key to SCA

Hamming Distance (HD)

The HD model assumes that all cells contribute to the power dissipation equally and that there is no difference between 0→ 1 and 1→0 transitions. The HD between two values v0 and v1 can be calculated as follows.

$$HD(v_0, v_1) = HW(v_0 \oplus v_1)$$

❖ **Example:**

v0= 0011 1100
v1= 0111 0000

$$HD(v_0, v_1) = 3$$

⊕ = 0100 1100

$$HD(v_0 \oplus v_1) = 3$$

Attackers commonly use the HD model to describe the power dissipation of buses and registers.

Hamming Weight (HW)

Hamming-weight model for power simulation

- To assume that the power is proportional to the number of bits that are set in the processed data value

The power consumption that is caused by a bit *v* is directly or inversely proportional to the value of the bit if the cell that processes *v* always stores the same value before or after the processing of *v*.

❖ **If an *n*-bit v0 is set to ‘0’ before transiting to v1**

$$HD(v_0, v_1) = HD(v_0 \oplus v_1) = HW(v_1)$$

❖ **If an *n*-bit v0 is set to ‘1’ before transiting to v1**

$$HD(v_0, v_1) = HD(v_0 \oplus v_1) = n - HW(v_1)$$

(i) A 1-bit HW model can be used if all bits of v0 are constant but not the same
(ii) Attackers only resort to the HW model if the HD model cannot be applied.

Correlation Matrix

Correlation Coefficient

- We can always corelate a set of side-channel signals with a set of statistical parameters of the intermediate values

$$r_{i,j,t} = \frac{\sum_{m=1}^{n} \overset{\text{Power Traces}}{\boxed{\left(X_{i,j,m} - \bar{X}_{i,j}\right)}} \overset{\text{Power Model}}{\boxed{\left(Y_{t,m} - \bar{Y}_{t}\right)}}}{\sqrt{\sum_{m=1}^{n}\left(X_{i,j,m} - \bar{X}_{i,j}\right)^{2}} \cdot \sqrt{\sum_{m=1}^{n}\left(Y_{t,m} - \bar{Y}_{t}\right)^{2}}}$$

Correlation Power Analysis

(1) Correlate side-channels with HW of the intermediate values

(2) Correlate side-channels with HD of the intermediate values

(3) Correlate side-channels with other statistical properties of the intermediate values

Correlation Power Analysis

Fundamental Principles

(1) Choose a target point (e.g., a bus, or a bit of the S-Box)

(2) Choose the statistical property (HW or HD or other) of the target point

(3) Based on one plaintext, try all guess keys (from 0 to 255 per bloc) to check if the statistical properties of the target point

(4) Repeat (3) for other plaintexts to build the databases with the statistical properties of the target point for each guess keys

(5) Based on each guess key, correlate the side-channels with the statistical properties of the target point

(6) If the guess key is right, if the correlation coefficient of the correct key should be the highest compared to that of other guess (wrong) keys

Correlation Power Analysis

Conceptual flow

Using the key guess to compute an intermediate value

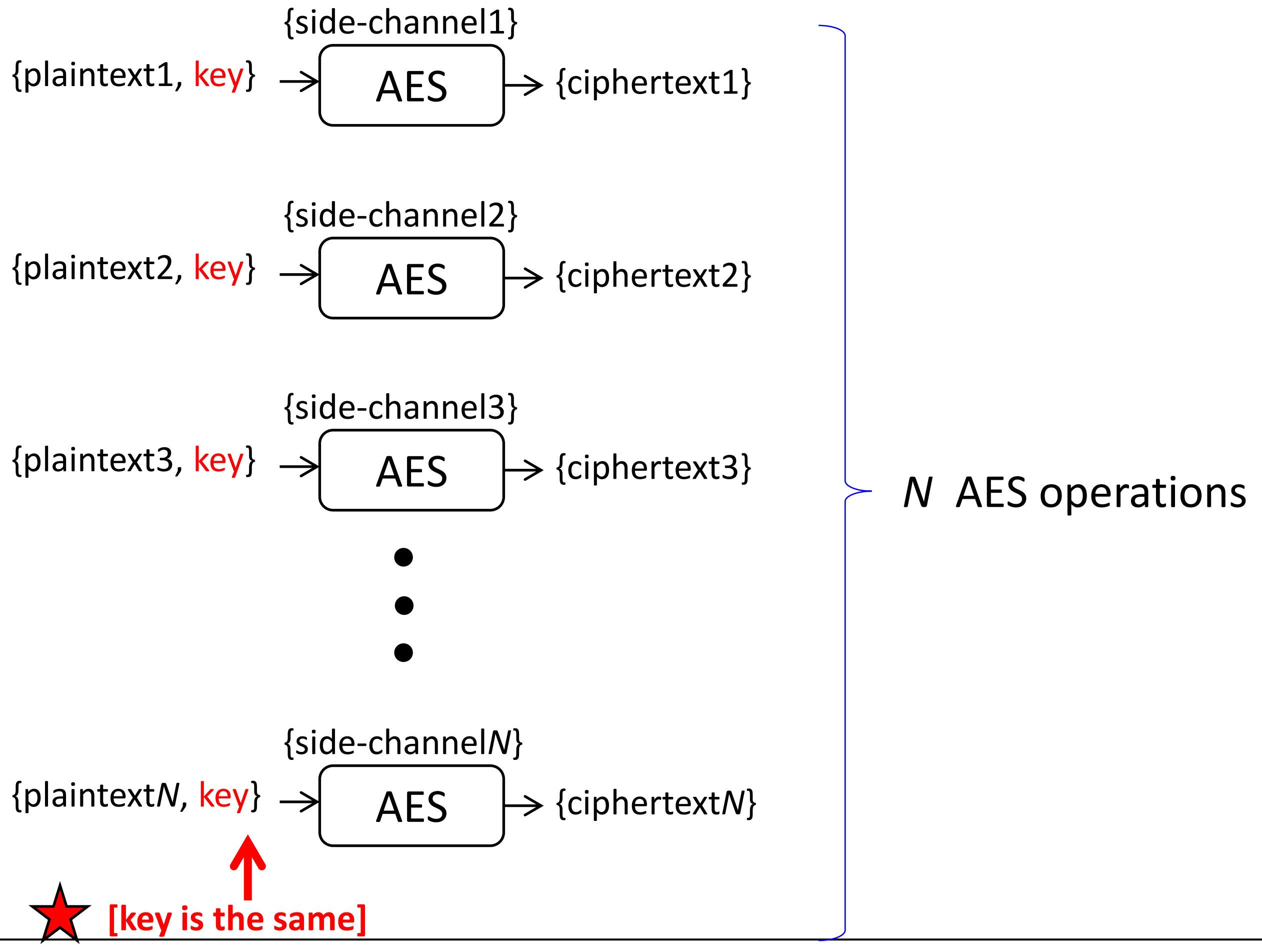
Side-Channel	Plaintext	HD@key=0	HD@key=1	HD@key=2	HD@key=255
side-channel1	Plaintext1	4	3	1	5
side-channel2	Plaintext2	3	2	0	4
side-channel3	Plaintext3	2	1	5	1
side-channel4	Plaintext4	3	0	6	0
side-channel5	Plaintext5	1	1	1	3
side-channel6	Plaintext6	0	3	3	0
side-channel7	Plaintext7	2	6	2	0
side-channel8	Plaintext8	4	7	0	6
side-channel9	Plaintext9	1	8	1	2
side-channel10	Plaintext10	6	1	1	0
…	…	…	…	…	…
side-channel*N*	Plaintext*N*	7	1	0	7
		C0	**C1**	**C2**	**C255**

Calculate the correlation

$$r_{i,j} = \frac{\sum_{d=1}^{D}(h_{d,i}-\bar{h}_i)\cdot(t_{d,j}-\bar{t}_j)}{\sqrt{\sum_{d=1}^{D}(h_{d,i}-\bar{h}_i)^2\cdot(t_{d,j}-\bar{t}_j)^2}}$$

SCA – Finding Statistical Properties

Let's look at *N* times of AES operations – any thing in common?



Establish the statistical properties between (plaintext, side-channel) or (cipher-text, side-channel) to guess the correct key

CPA - Discussion

Points to take note

(1) In CPA, the signals/noise are somewhat normalized, hence usually resulting in better results

(2) HW and HD models disclose a lot of statistical info of the side-channel signals

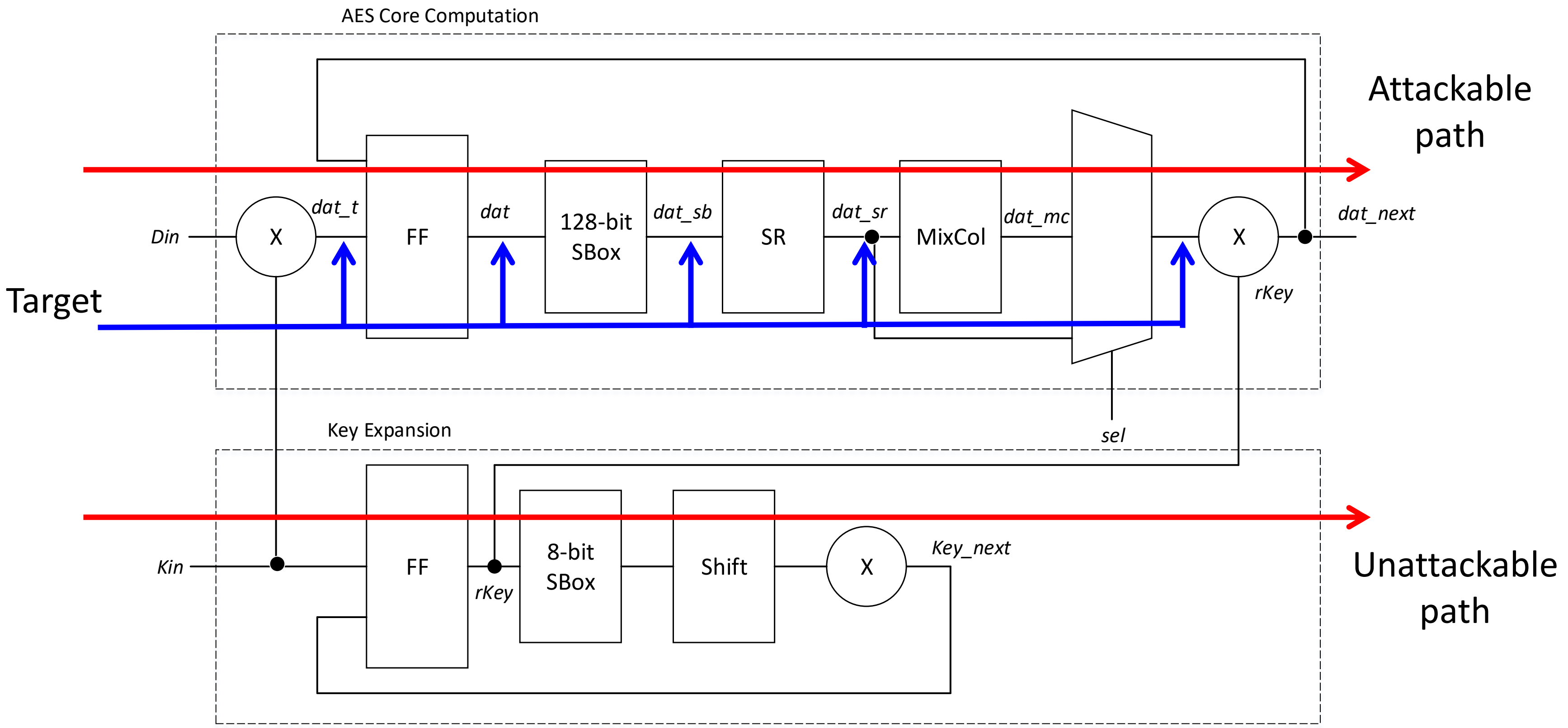
(3) Traces need to be a lot in order to stabilize the correlation coefficient of the correct key

(4) Countermeasures may make the evaluation difficult in view of noise, corrupted side-channel signals and undisclosed leakage regions

(5) Pre-processing techniques may help

Target Point

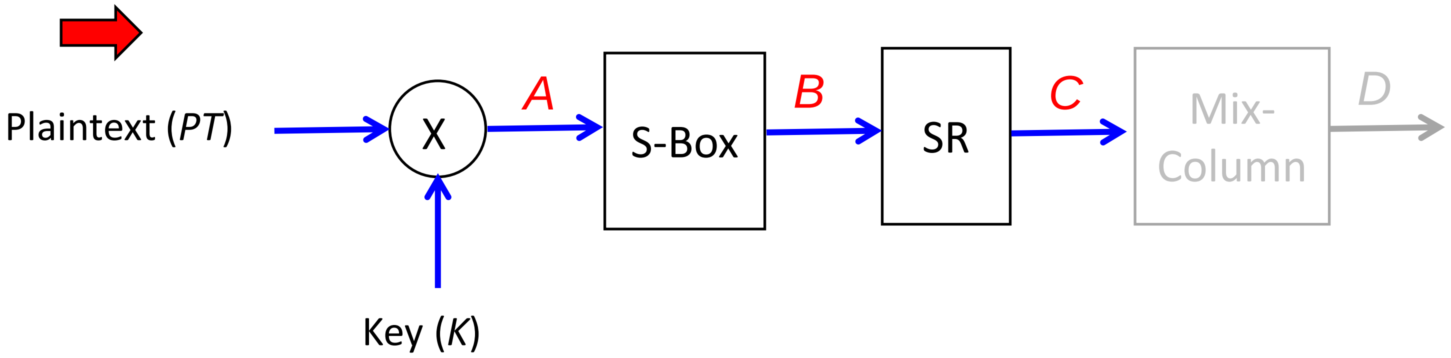
Because we are exploiting the intermediate value of AES which is independent of the AES implementation. We term the intermediate value as ‘target’ which must be is dependent on both the key and data.



Common target: Location involving Sbox (which is a non-linear function) could be good choice.

First Round SCA

- The functions for the target



A = *PT*⊗*K*

B = S-Box(*PT*⊗*K*)

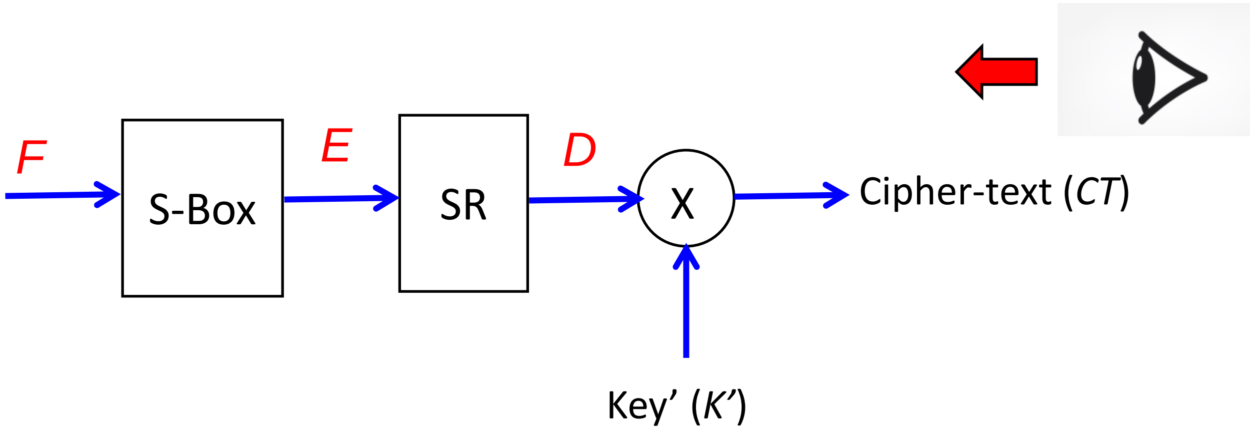
C = SR(S-Box(*PT*⊗*K*))

These intermediate values can generated on the basis of each sub-key (8-bit) – meaning that each attempt only has 256 combinations of key guesses

For the first round attacks, the locations *A*, *B* and *C* are the common targets.

Last Round SCA

❖ The functions for the target



$D = CT \otimes K'$

$E = \text{Inv-SR}(CT \otimes K')$

$F = \text{Inv-S-Box}(\text{Inv-SR}(PT \otimes K))$

Inverse functions are needed. These intermediate values can generated on the basis of each sub-key (8-bit) – meaning that each attempt only has 256 combinations of key guesses

For the last round attacks, the locations *D*, *E* and *F* are the common targets.

END