The Davies-Murphy Power Attack

Sébastien Kunz-Jacques
Frédéric Muller
Frédéric Valette
DCSSI Crypto Lab

Introduction

- Two approaches for attacking crypto devices
  - “traditional” cryptanalysis
  - Side Channel Attacks (SCA)
- SCA often use techniques from traditional cryptanalysis
- Popular methods (SPA, DPA, CA, …) have limitations

Results in this paper

- A new SCA based on Davies-Murphy Attack against DES
- More flexible and powerful than previously known attacks
  - Apply to inner rounds
  - Avoid DPA countermeasures

Side Channel Attacks

- Fundamental hypothesis : side channel leak secret information
- Power Attacks : power consumption W is correlated with manipulated data D

$$W = \lambda D + \text{Noise}$$

- Other techniques : Fault Injection, Timing, Electromagnetic Radiations, …

Usual Approach

Power Attacks apply if we can predict :

- D depending on plaintext and few key bits

  $\rightarrow$ *Differential Power Attack (DPA)*

- A function of $D_1,\ldots,D_t$ depending on plaintext and few key bits

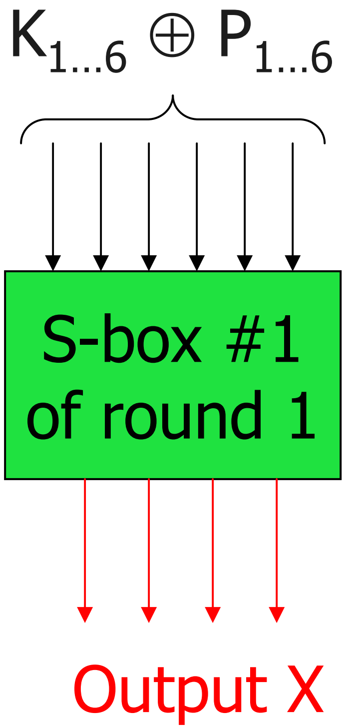
  $\rightarrow$ *Higher-Order DPA (HO-DPA)*

The case of DES

Basic DPA

Predict X with
- plaintext
- guess on 6 key bits ($K_{1...6}$)

Applies to round 1 or 16



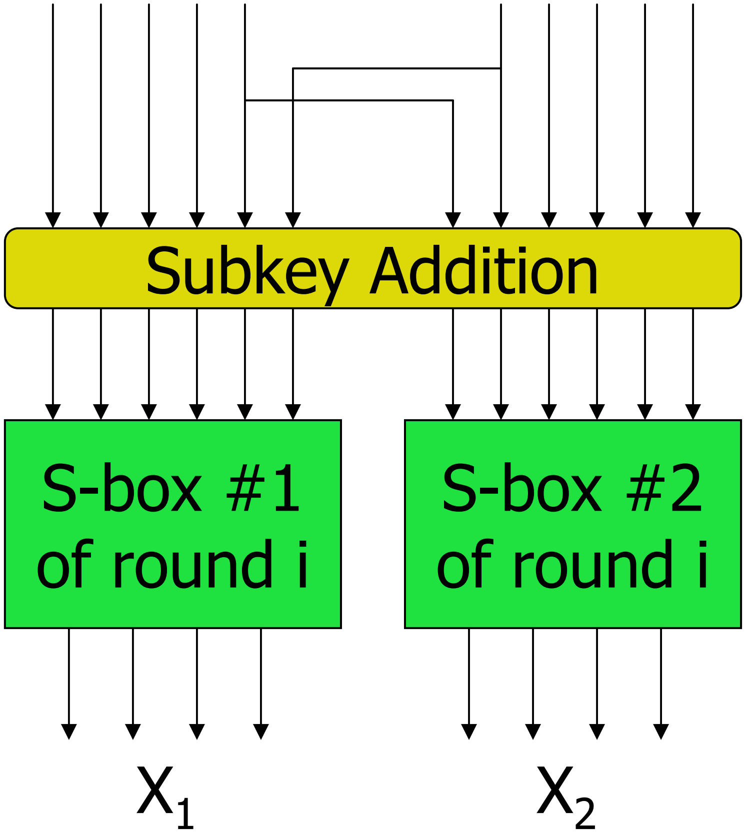
Limitations

- Countermeasures :
  - Protection of the first/last rounds
  - Masking
  - Duplication
- Practical Problems to detect the correct key
  - “Ghost peaks”
  - Hardware Implementation / Parallelism

Motivations of this paper

- Find a better attack
  - Target any inner round of DES
  - Avoid popular countermeasures (masking)
- Techniques from “traditional” cryptanalysis may be a good starting point
  - They demonstrate real weaknesses of the designs
  - Often useful in Side Channel Attacks

Davies-Murphy Attack



Duplication of input bits

Outputs are not balanced

Davies-Murphy Attack

- Observation by Davies and Murphy about pairs of adjacent S-boxes
- Distribution of $(X_1, X_2) \in \{0,1\}^8$ is not uniform
- Two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are possible depending on 1 key bit k

Distributions

y_2 \ y_1	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
00	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
01	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
02	2	2	4	6	4	4	6	4	6	4	0	4	4	2	6	6
03	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
04	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
05	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
06	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
07	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
08	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
09	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
10	6	6	4	2	4	4	2	4	2	4	8	4	4	6	2	2
11	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
12	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
13	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
14	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
15	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

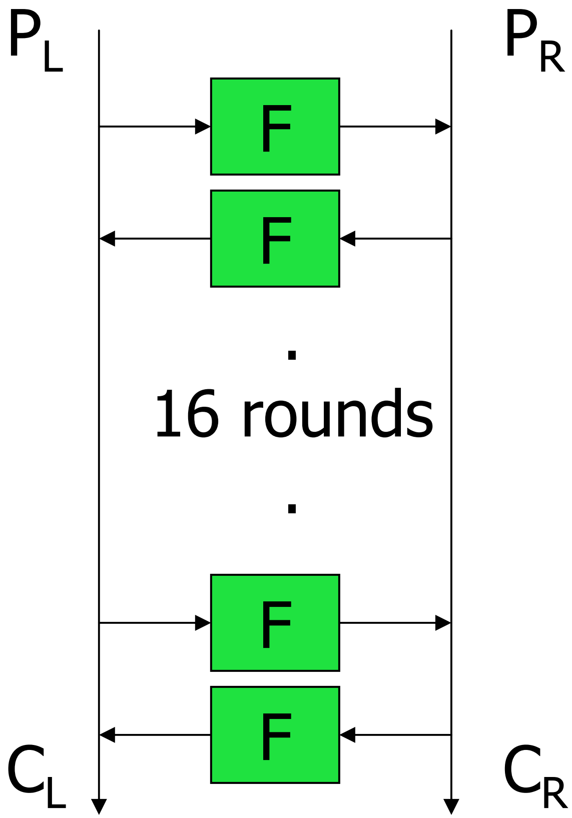
Case $k = 0$

y_2 \ y_1	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
00	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
01	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
02	6	6	4	2	4	4	2	4	2	4	8	4	4	6	2	2
03	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
04	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
05	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
06	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
07	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
08	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
09	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
10	2	2	4	6	4	4	6	4	6	4	0	4	4	2	6	6
11	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
12	3	3	4	5	4	4	5	4	5	4	2	4	4	3	5	5
13	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
14	5	5	4	3	4	4	3	4	3	4	6	4	4	5	3	3
15	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Case $k = 1$

Table 1. Biased Distributions for S_1 and S_2 (all elements in the table should be divided by 2^{10})

Application to 16 rounds



The distribution of $P_L \oplus C_L$ (or $P_R \oplus C_R$) is the XOR of 8 "biased" distributions

Biham and Biruykov (1994) showed how to mount a key-recovery attack based on this property

Impact for Power Attacks

- For a random (even masked) input of any inner round, the output is not balanced
- This imbalance depends on 1 key bit k
- Hence, the power consumption is different in average when k=0 and k=1
- In theory, analysis of power curves → retrieve k

Link with DPA

Power Attacks apply if we can predict :

- D depending on plaintext and few key bits

  → *Differential Power Attack (DPA)*

- Here, we can predict the distribution of intermediate data D from 1 key bit

  → *Davies-Murphy Power Attack (DMPA)*

Consumption Model

Attacks require a consumption model :

- Linear Model

  $W = \lambda \, \text{Linear}(D) + \text{Noise}$

- Hamming Weight Model

  $W = \lambda \, \text{Hamming}(D) + \text{Noise}$

Used in the paper

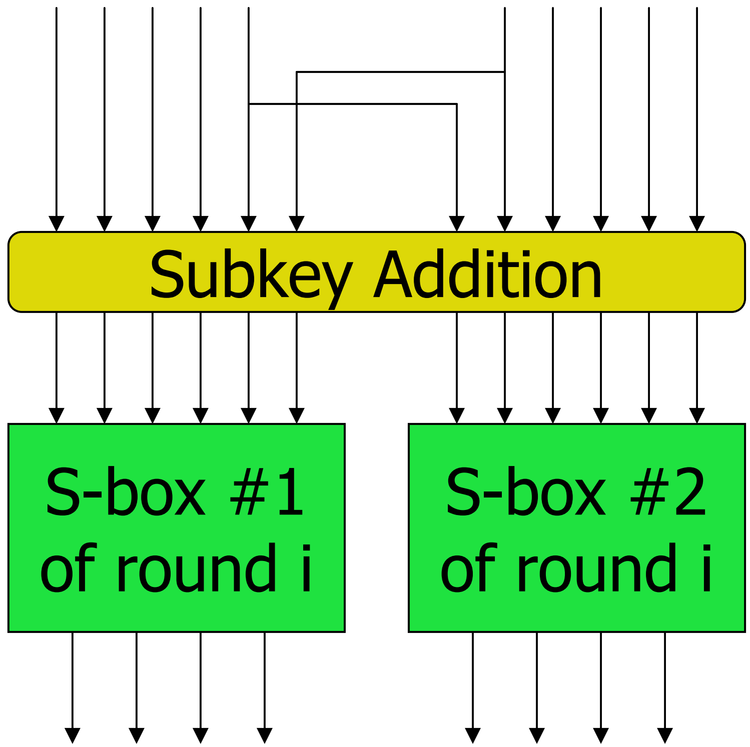
- Hamming Distance Model

  $W = \lambda \, \text{Hamming}(D \oplus R) + \text{Noise}$

What is R ?

- R is a constant value
- Typically, D is stored in a register that contained R previously
- Consumption depends on how many bits are flipped

Average Hamming Distance



Duplication of input bits

Outputs are not balanced w.r.t. hamming distance

h_1=Hamming(R⊕X_1) h_2=Hamming(R⊕X_2)

Distributions with R=0

Random Distribution						Case $k=0$						Case $k=1$					
$h_2 \backslash h_1$	0	1	2	3	4	$h_2 \backslash h_1$	0	1	2	3	4	$h_2 \backslash h_1$	0	1	2	3	4
0	4	16	24	16	4	0	4	16	24	16	4	0	4	16	24	16	4
1	16	64	96	64	16	1	16	64	96	64	16	1	16	64	96	64	16
2	24	96	144	96	24	2	26	96	144	96	22	2	22	96	144	96	26
3	16	64	96	64	16	3	14	64	96	64	18	3	18	64	96	64	14
4	4	16	24	16	4	4	4	16	24	16	4	4	4	16	24	16	4

Table 2. Distributions of output hamming weight for S_1 and S_2 (all elements in the table should be divided by 2^{10})

Statistical Distance

S-boxes	Statistical Distance $\lvert \mathcal{D}_1 - \mathcal{D}_0 \rvert$			
	constant $= 0$	worst constant	best constant	average value
(S_1, S_2)	$\frac{1}{64}$	0	$\frac{5}{32}$	$\frac{1.5}{32}$
(S_2, S_3)	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{7}{32}$	$\frac{3.656}{32}$
(S_3, S_4)	$\frac{1}{128}$	0	$\frac{9}{128}$	$\frac{0.473}{32}$
(S_4, S_5)	$\frac{1}{64}$	0	$\frac{9}{64}$	$\frac{0.984}{32}$
(S_5, S_6)	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{3}{32}$	$\frac{1.195}{32}$
(S_6, S_7)	$\frac{3}{64}$	$\frac{1}{64}$	$\frac{9}{128}$	$\frac{1.262}{32}$
(S_7, S_8)	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{25}{128}$	$\frac{3.094}{32}$
(S_8, S_1)	$\frac{1}{16}$	$\frac{1}{128}$	$\frac{3}{32}$	$\frac{0.711}{32}$

Table 4. Statistical distances with constant R_i's

Summary

- Power consumption is correlated with hamming distance
- Distribution of hamming distance depends on 1 key bit k
- With an appropriate indicator, determine if k=0 or k=1

Summary (2)

- More details in the paper
- Practical problems
  - What indicator to choose ?
  - Parallelism of the architecture ?

Conclusion

- Davies-Murphy Power Attack (DMPA)
  - Predict the distribution of intermediate data
  - Apply to any DES inner round
  - Counter the effect of masking countermeasures
- Extensions and Improvements
  - Find better methods for parallel implementations
  - Extend to other ciphers