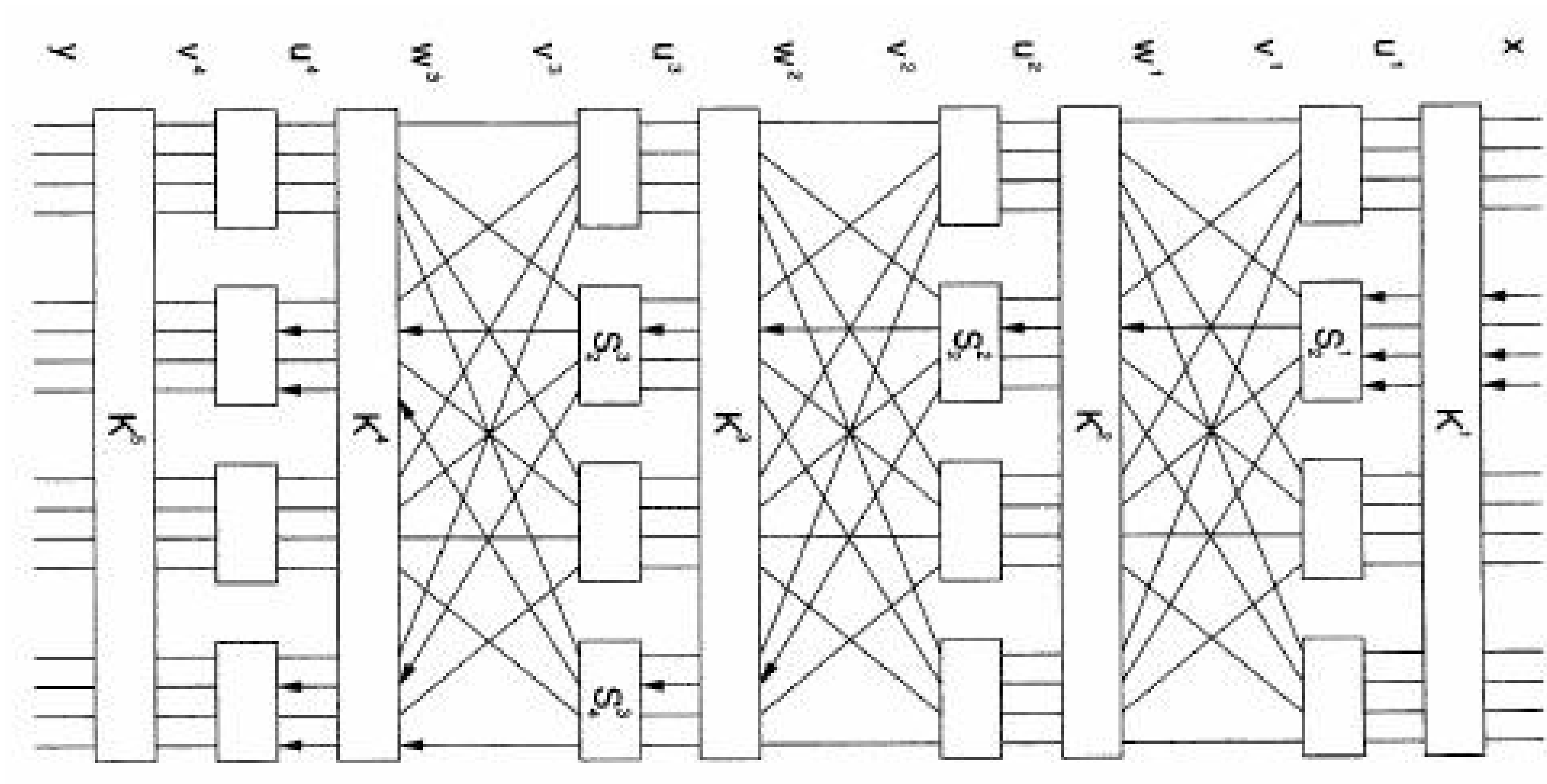
Linear Cryptanalysis

T-79.5501 Cryptology

Lecture 5
February 26, 2008

Kaisa Nyberg

SPN – A Small Example



Linear Approximation of S-boxes

S-boxes

S-box is a function $f : \{0,1\}^n \to \{0,1\}^m$, where m and n are (small) integers.

Example. The S-box S_4 of the DES

7	13	14	3	0	6	9	10	1	2	8	5	11	12	4	15
13	8	11	5	6	15	0	3	4	7	2	12	1	10	14	9
10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4
3	15	0	6	10	1	13	8	9	4	5	11	12	7	2	14

DES S-box S_4 First Row

7	13	14	3	0	6	9	10	1	2	8	5	11	12	4	15

x	y	$x_1 \oplus y_3$	x	y	$x_1 \oplus y_3$
0000	0111	1	1000	0001	1
0001	1101	0	1001	0010	0
0010	1110	1	1010	1000	1
0011	0011	1	1011	0101	1
0100	0000	0	1100	1011	0
0101	0110	1	1101	1100	1
0110	1001	0	1110	0100	1
0111	1010	1	1111	1111	0

The S-box π_S

z	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
$\pi_S(z)$	E	4	D	1	2	F	B	8	3	A	6	C	5	9	0	7

X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3	Y_4
0	0	0	0	1	1	1	0
0	0	0	1	0	1	0	0
0	0	1	0	1	1	0	1
0	0	1	1	0	0	0	1
0	1	0	0	0	0	1	0
0	1	0	1	1	1	1	1
0	1	1	0	1	0	1	1
0	1	1	1	1	0	0	0
1	0	0	0	0	0	1	1
1	0	0	1	1	0	1	0
1	0	1	0	0	1	1	0
1	0	1	1	1	1	0	0
1	1	0	0	0	1	0	1
1	1	0	1	1	0	0	1
1	1	1	0	0	0	0	0
1	1	1	1	0	1	1	1

Linearity of S-box

- **Definition** Suppose $f:\{0,1\}^n \rightarrow \{0,1\}^m$ is an S-box and $a=(a_1,\ldots,a_n)\in\{0,1\}^n$ and $b=(b_1,\ldots,b_n)\in\{0,1\}^m$. We use $N_L(a,b)$ to denote the number of $x\in\{0,1\}^n$ such that $f(x)=y$ and

$$a_1x_1 \oplus a_2x_2 \oplus \ldots \oplus a_nx_n = b_1y_1 \oplus b_2y_2 \oplus \ldots \oplus b_ny_n.$$

or using the short notation

$$a\cdot x \oplus b\cdot y = 0.$$

- *Remark.* Then the *bias* of the random variable $a\cdot \mathbf{X} \oplus b\cdot \mathbf{Y}$ is equal to $2^{-n}N_L(a,b)-\frac{1}{2}$ (to be defined soon).

The Linear Approximation Table $N_L(a,b)$

a \ b	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
1	8	8	6	6	8	8	6	14	10	10	8	8	10	10	8	8
2	8	8	6	6	8	8	6	6	8	8	10	10	8	8	2	10
3	8	8	8	8	8	8	8	8	10	2	6	6	10	10	6	6
4	8	10	8	6	6	4	6	8	8	6	8	10	10	4	10	8
5	8	6	6	8	6	8	12	10	6	8	4	10	8	6	6	8
6	8	10	6	12	10	8	8	10	8	6	10	12	6	8	8	6
7	8	6	8	10	10	4	10	8	6	8	10	8	12	10	8	10
8	8	8	8	8	8	8	8	8	6	10	10	6	10	6	6	2
9	8	8	6	6	8	8	6	6	4	8	6	10	8	12	10	6
A	8	12	6	10	4	8	10	6	10	10	8	8	10	10	8	8
B	8	12	8	4	12	8	12	8	8	8	8	8	8	8	8	8
C	8	6	12	6	6	8	10	8	10	8	10	12	8	10	8	6
D	8	10	10	8	6	12	8	10	4	6	10	8	10	8	8	10
E	8	10	10	8	6	4	8	10	6	8	8	6	4	10	6	8
F	8	6	4	6	6	8	10	8	8	6	12	6	6	8	10	8

FIGURE 3.2
Linear approximation table: values of $N_L(a,b)$

Piling-Up Lemma

Piling-Up Lemma

- **Definition** Suppose that $\mathbf{T}$ is a discrete random variable that takes values from $\{0,1\}$. Then the quantity

$$\varepsilon = \mathbf{Pr}[\mathbf{T}=0] - \frac{1}{2}$$

is called the bias of $\mathbf{T}$.

- **Lemma 3.1** Suppose $\mathbf{T_j}$ are independent discrete random variables with biases ε_j, $j = 1, 2, \ldots, k$. Then the bias ε of $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2 \oplus \ldots \oplus \mathbf{T}_k$ can be calculated as

$$\varepsilon = 2^{k-1}\varepsilon_1\varepsilon_2\cdots\varepsilon_k.$$

Proof of Piling-Up Lemma

- *Proof*. We will give the proof for $k = 2$. The general case follows by induction. By independency

$$
\begin{aligned}
&\mathbf{Pr}[\mathbf{T} = 0] = \mathbf{Pr}[\mathbf{T_1} = 0]\mathbf{Pr}[\mathbf{T_2} = 0] + \mathbf{Pr}[\mathbf{T_1} = 1]\mathbf{Pr}[\mathbf{T_2} = 1] \\
&= \mathbf{Pr}[\mathbf{T_1} = 0]\mathbf{Pr}[\mathbf{T_2} = 0] + (1 - \mathbf{Pr}[\mathbf{T_1} = 0])(1 - \mathbf{Pr}[\mathbf{T_2} = 0]) \\
&= 2\mathbf{Pr}[\mathbf{T_1} = 0]\mathbf{Pr}[\mathbf{T_2} = 0] - \mathbf{Pr}[\mathbf{T_1} = 0] - \mathbf{Pr}[\mathbf{T_2} = 0] + 1
\end{aligned}
$$

- From this we get

$$
\begin{aligned}
&\varepsilon = \mathbf{Pr}[\mathbf{T} = 0] - 1/2 \\
&= 2(\mathbf{Pr}[\mathbf{T_1} = 0]\mathbf{Pr}[\mathbf{T_2} = 0] - \frac{1}{2}\mathbf{Pr}[\mathbf{T_1} = 0] - \frac{1}{2}\mathbf{Pr}[\mathbf{T_2} = 0] + \frac{1}{4}) \\
&= 2(\mathbf{Pr}[\mathbf{T_1} = 0] - \frac{1}{2})(\mathbf{Pr}[\mathbf{T_2} = 0] - \frac{1}{2}) = 2\varepsilon_1\varepsilon_2.
\end{aligned}
$$

Piling-Up Lemma and Independence

- **Example** Let $\mathbf{T}_1$, $\mathbf{T}_2$ and $\mathbf{T}_3$ be independent random variables with biases $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 1/4$. Denote

$$
\begin{aligned}
\mathbf{T}_{12} &= \mathbf{T}_1 \oplus \mathbf{T}_2 \text{ with bias } \varepsilon_{12} = 2\varepsilon_1\varepsilon_2 = \frac{1}{8}, \\
\mathbf{T}_{23} &= \mathbf{T}_2 \oplus \mathbf{T}_3 \text{ with bias } \varepsilon_{23} = 2\varepsilon_2\varepsilon_3 = \frac{1}{8}, \\
\mathbf{T}_{13} &= \mathbf{T}_1 \oplus \mathbf{T}_3 \text{ with bias } \varepsilon_{13} = 2\varepsilon_1\varepsilon_3 = \frac{1}{8}.
\end{aligned}
$$

Then $\mathbf{T}_{12}$ and $\mathbf{T}_{23}$ cannot be independent. If they were independent, then by the Piling-Up Lemma the bias of $\mathbf{T}_{13} = \mathbf{T}_{12} \oplus \mathbf{T}_{23}$ would be equal to $2 \cdot \frac{1}{8} \cdot \frac{1}{8} = \frac{1}{32}$ which is not the case.

Converse of the Piling-Up Lemma

- It can be shown that the converse of the Piling-Up Lemma also holds. We state it here for two random variables.
- **Converse of the Piling-Up Lemma.** Suppose $\mathbf{T}_1$ and $\mathbf{T}_2$ are discrete random variables with biases ε_1 and ε_2. If the bias ε of $\mathbf{T} = \mathbf{T}_1 \oplus \mathbf{T}_2$ satisfies

  $$\varepsilon = 2\varepsilon_1\varepsilon_2,$$

  then $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent.
- To give the proof we introduce first the Walsh-Hadamard transform.

Walsh-Hadamard Transform

- **Definition** Suppose $f : \{0,1\}^n \to \mathbb{Z}$ is any integer-valued function of bit strings of length n. The Walsh-Hadamard transform transforms f to a function $F : \{0,1\}^n \to \mathbb{Z}$ defined as

$$F(w) = \sum_{x \in \{0,1\}^n} f(x)(-1)^{w \cdot x}, w \in \{0,1\}^n,$$

  where the sum is taken over integers.

- The Walsh-Hadamard Transform can also be inverted. Actually, it is its own inverse upto a constant multiplier (see exercises):

$$f(x) \ = \ 2^{-n} \sum_{w \in \{0,1\}^n} F(w)(-1)^{w \cdot x} \text{, for all } x \in \{0,1\}^n.$$

Probability Distribution and Bias of $(\mathbf{T}_1, \mathbf{T}_2)$

- Suppose $\mathbf{Z} = (\mathbf{T}_1, \mathbf{T}_2)$ is a pair of binary random variables, $a = (a_1, a_2)$ be a pair of bits and ε_a be the bias of $a \cdot \mathbf{Z} = a_1 \mathbf{T}_1 \oplus a_2 \mathbf{T}_2$.

- **Lemma**

$$\varepsilon_a = \frac{1}{2} \sum_{(t_1,t_2)} \mathbf{Pr}[\mathbf{Z} = (t_1, t_2)](-1)^{a_1 t_1 \oplus a_2 t_2}$$

- *Proof*. Denote $t = (t_1, t_2)$ and $a \cdot t = a_1 t_1 \oplus a_2 t_2$. Then

$$2\varepsilon_a = 2\mathbf{Pr}[a \cdot \mathbf{Z} = 0] - 1 = \mathbf{Pr}[a \cdot \mathbf{Z} = 0] - \mathbf{Pr}[a \cdot \mathbf{Z} = 1]$$
$$= \sum_{t, a \cdot t = 0} \mathbf{Pr}[\mathbf{Z} = t] - \sum_{t, a \cdot t = 1} \mathbf{Pr}[\mathbf{Z} = t] = \sum_{t} \mathbf{Pr}[\mathbf{Z} = t](-1)^{a \cdot t}.$$

Probability Distribution and Bias of $(\mathbf{T}_1, \mathbf{T}_2)$

- Indeed, $\varepsilon_a = F(a)$ is the Walsh-Hadamard transform of $f(t) = \Pr[\mathbf{Z} = t]$.
- Using the inverse Walsh-Hadamard transform we get the following

$$\mathbf{Pr}[\mathbf{Z} = t] = \frac{1}{2} \sum_{(a_1, a_2)} \varepsilon_a (-1)^{a_1 t_1 \oplus a_2 t_2}.$$

Proof of the Converse of the Piling-Up Lemma, $k = 2$

- **Claim.** If the bias of $\mathbf{T}_1 \oplus \mathbf{T}_2$ is equal to $2\varepsilon_1\varepsilon_2$ then $\mathbf{T}_1$ and $\mathbf{T}_2$ are independent.

- *Proof.* For $a = (a_1, a_2) \in \{0,1\}^2$, we use ε_a to denote the bias of $a \cdot \mathbf{Z} = a_1\mathbf{T}_1 \oplus a_2\mathbf{T}_2$. Then

$$\begin{aligned}
\mathbf{Pr}[\mathbf{T}_1 = t_1, \mathbf{T}_2 = t_2] &= \sum_a \varepsilon_a (-1)^{a_1t_1 \oplus a_2t_2} \\
&= \varepsilon_{(0,0)} + \varepsilon_{(1,0)}(-1)^{t_1} + \varepsilon_{(0,1)}(-1)^{t_2} + \varepsilon_{(1,1)}(-1)^{t_1 \oplus t_2} \\
&= \frac{1}{2} + \varepsilon_1(-1)^{t_1} + \varepsilon_2(-1)^{t_2} + 2\varepsilon_1\varepsilon_2(-1)^{t_1}(-1)^{t_2} \\
&= (\varepsilon_1(-1)^{t_1} + \frac{1}{2})(\varepsilon_2(-1)^{t_2} + \frac{1}{2}) \\
&= \mathbf{Pr}[\mathbf{T}_1 = t_1]\mathbf{Pr}[\mathbf{T}_2 = t_2]
\end{aligned}$$

Linear Attack on the SPN

Related Random Variables

$$
\begin{aligned}
\mathbf{T}_1 &= \mathbf{U}_5^1 \oplus \mathbf{U}_7^1 \oplus \mathbf{U}_8^1 \oplus \mathbf{V}_6^1 \text{ has bias } \frac{1}{4}, \text{ as } N_L(B,4) = 12 \\
\mathbf{T}_2 &= \mathbf{U}_6^2 \oplus \mathbf{V}_6^2 \oplus \mathbf{V}_8^2 \text{ has bias } -\frac{1}{4}, \text{ as } N_L(4,5) = 4 \\
\mathbf{T}_3 &= \mathbf{U}_6^3 \oplus \mathbf{V}_6^3 \oplus \mathbf{V}_8^3 \text{ has bias } -\frac{1}{4}, \text{ as } N_L(4,5) = 4 \\
\mathbf{T}_4 &= \mathbf{U}_{14}^3 \oplus \mathbf{V}_{14}^3 \oplus \mathbf{V}_{16}^3 \text{ has bias } -\frac{1}{4}, \text{ as } N_L(4,5) = 4
\end{aligned}
$$

The four random variables have biases that are high in absolute value. By the Piling-Up Lemma we get the linear approximation

$$
\mathbf{T} = \mathbf{X}_5 \oplus \mathbf{X}_7 \oplus \mathbf{X}_8 \oplus \mathbf{U}_6^4 \oplus \mathbf{U}_8^4 \oplus \mathbf{U}_{14}^4 \oplus \mathbf{U}_{16}^4 \quad (3.3)
$$

with bias $|2^3(\frac{1}{4})^4| = \frac{1}{32}$ in absolute value.

The Last-Round Trick

- Matsui's Algorithm 2 is based on the following assumption:
  Wrong Key Assumption. If on the last round a wrong key is used to decrypt the ciphertext then the random variable of the linear approximation is much more uniformly distributed as indicated by the bias.
- In the example of the textbook, if wrong partial keys K_i^5, $i = 5,6,7,8,13,14,15,16$ are used to compute the values of $\mathbf{U}_6^4$, $\mathbf{U}_8^4$, $\mathbf{U}_{14}^4$, and $\mathbf{U}_{16}^4$, then the distribution of $\mathbf{T}$ is almost uniform.
- In this manner, part of the last round key bits can be found. The rest can be found by repeating the attack with a different approximation, or by exhaustive search.
- The required number of plaintext-ciphertext pairs is proportional to the inverse of the squared bias of the linear approximation. In the case of the example the data requirement is about 8000 plaintext-ciphertext pairs obtained using the same key.