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

<table>
<thead>
<tr>
<th>7</th>
<th>13</th>
<th>14</th>
<th>3</th>
<th>0</th>
<th>6</th>
<th>9</th>
<th>10</th>
<th>1</th>
<th>2</th>
<th>8</th>
<th>5</th>
<th>11</th>
<th>12</th>
<th>4</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>10</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>9</td>
<td>0</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>13</td>
<td>15</td>
<td>1</td>
<td>3</td>
<td>14</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>0</td>
<td>6</td>
<td>10</td>
<td>1</td>
<td>13</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>11</td>
<td>12</td>
<td>7</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>
## DES S-box $S_4$ First Row

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$x_1 \oplus y_3$</th>
<th>$x$</th>
<th>$y$</th>
<th>$x_1 \oplus y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>0111</td>
<td>1</td>
<td>1000</td>
<td>0001</td>
<td>1</td>
</tr>
<tr>
<td>0001</td>
<td>1101</td>
<td>0</td>
<td>1001</td>
<td>0010</td>
<td>0</td>
</tr>
<tr>
<td>0010</td>
<td>1110</td>
<td>1</td>
<td>1010</td>
<td>1000</td>
<td>1</td>
</tr>
<tr>
<td>0011</td>
<td>0011</td>
<td>1</td>
<td>1011</td>
<td>0101</td>
<td>1</td>
</tr>
<tr>
<td>0100</td>
<td>0000</td>
<td>0</td>
<td>1100</td>
<td>1011</td>
<td>0</td>
</tr>
<tr>
<td>0101</td>
<td>0110</td>
<td>1</td>
<td>1101</td>
<td>1100</td>
<td>1</td>
</tr>
<tr>
<td>0110</td>
<td>1001</td>
<td>0</td>
<td>1110</td>
<td>0100</td>
<td>1</td>
</tr>
<tr>
<td>0111</td>
<td>1010</td>
<td>1</td>
<td>1111</td>
<td>1111</td>
<td>0</td>
</tr>
</tbody>
</table>
The S-box $\pi_S$

<table>
<thead>
<tr>
<th>$z$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_S(z)$</td>
<td>E</td>
<td>4</td>
<td>D</td>
<td>1</td>
<td>2</td>
<td>F</td>
<td>B</td>
<td>8</td>
<td>3</td>
<td>A</td>
<td>6</td>
<td>C</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

![S-box Diagram]

Linear Cryptanalysis – 6/36
**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 \( a \cdot X \oplus b \cdot Y \) is equal to \( 2^{-n}N_L(a, b) - \frac{1}{2} \) (to be defined soon).
The Linear Approximation Table $N_L(a, b)$

![Table Image]
Piling-Up Lemma
Piling-Up Lemma

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

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

is called the bias of $T$.

**Lemma 3.1** Suppose $T_j$ are independent discrete random variables with biases $\varepsilon_j$, $j = 1, 2, \ldots, k$. Then the bias $\varepsilon$ of $T = T_1 \oplus T_2 \oplus \ldots \oplus T_k$ 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

\[
\Pr[T = 0] = \Pr[T_1 = 0]\Pr[T_2 = 0] + \Pr[T_1 = 1]\Pr[T_2 = 1]
\]

\[
= \Pr[T_1 = 0]\Pr[T_2 = 0] + (1 - \Pr[T_1 = 0])(1 - \Pr[T_2 = 0])
\]

\[
= 2\Pr[T_1 = 0]\Pr[T_2 = 0] - \Pr[T_1 = 0] - \Pr[T_2 = 0] + 1
\]

From this we get

\[
\varepsilon = \Pr[T = 0] - 1/2
\]

\[
= 2(\Pr[T_1 = 0]\Pr[T_2 = 0] - \frac{1}{2}\Pr[T_1 = 0] - \frac{1}{2}\Pr[T_2 = 0] + \frac{1}{4})
\]

\[
= 2(\Pr[T_1 = 0] - \frac{1}{2})(\Pr[T_2 = 0] - \frac{1}{2}) = 2\varepsilon_1\varepsilon_2.
\]
Piling-Up Lemma and Independence

**Example** Let $T_1$, $T_2$ and $T_3$ be independent random variables with biases $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 1/4$. Denote

$$T_{12} = T_1 \oplus T_2 \text{ with bias } \varepsilon_{12} = 2\varepsilon_1\varepsilon_2 = \frac{1}{8},$$

$$T_{23} = T_2 \oplus T_3 \text{ with bias } \varepsilon_{23} = 2\varepsilon_2\varepsilon_3 = \frac{1}{8},$$

$$T_{13} = T_1 \oplus T_3 \text{ with bias } \varepsilon_{13} = 2\varepsilon_1\varepsilon_3 = \frac{1}{8}.$$

Then $T_{12}$ and $T_{23}$ cannot be independent. If they were independent, then by the Piling-Up Lemma the bias of $T_{13} = T_{12} \oplus T_{23}$ 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 $T_1$ and $T_2$ are discrete random variables with biases $\varepsilon_1$ and $\varepsilon_2$. If the bias $\varepsilon$ of $T = T_1 \oplus T_2$ satisfies

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

  then $T_1$ and $T_2$ are independent.

- To give the proof we introduce first the Walsh-Hadamard transform.
Walsh-Hadamard Transform

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

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

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):

\[
f(x) = 2^{-n} \sum_{w \in \{0,1\}^n} F(w)(-1)^{w \cdot x}, \quad \text{for all } x \in \{0, 1\}^n.
\]
Probability Distribution and Bias of \((T_1, T_2)\)

- Suppose \(Z = (T_1, T_2)\) is a pair of binary random variables, 
  \(a = (a_1, a_2)\) be a pair of bits and \(\varepsilon_a\) be the bias of 
  \(a \cdot Z = a_1 T_1 \oplus a_2 T_2\).

- **Lemma**

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

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

  \[
  2\varepsilon_a = 2\Pr[a \cdot Z = 0] - 1 = \Pr[a \cdot Z = 0] - \Pr[a \cdot Z = 1]
  \]

  \[
  = \sum_{t, a \cdot t = 0} \Pr[Z = t] - \sum_{t, a \cdot t = 1} \Pr[Z = t] = \sum_{t} \Pr[Z = t] (-1)^{a \cdot t}.
  \]
Probability Distribution and Bias of \((T_1, T_2)\)

- Indeed, \(\varepsilon_a = F(a)\) is the Walsh-Hadamard transform of \(f(t) = \Pr[Z = t]\).

- Using the inverse Walsh-Hadamard transform we get the following

\[
\Pr[Z = t] = \frac{1}{2} \sum_{(a_1, a_2)} \varepsilon_a (-1)^{a_1 t_1 \oplus a_2 t_2}.
\]
Proof of the Converse of the Piling-Up Lemma, \( k = 2 \)

- **Claim.** If the bias of \( T_1 \oplus T_2 \) is equal to \( 2\epsilon_1\epsilon_2 \) then \( T_1 \) and \( T_2 \) are independent.

- **Proof.** For \( a = (a_1, a_2) \in \{0, 1\}^2 \), we use \( \epsilon_a \) to denote the bias of \( a \cdot Z = a_1 T_1 \oplus a_2 T_2 \). Then

\[
\Pr[ T_1 = t_1, T_2 = t_2 ] = \sum_a \epsilon_a (-1)^{a_1 t_1 \oplus a_2 t_2}
\]

\[
= \epsilon_{(0,0)} + \epsilon_{(1,0)} (-1)^{t_1} + \epsilon_{(0,1)} (-1)^{t_2} + \epsilon_{(1,1)} (-1)^{t_1 \oplus t_2}
\]

\[
= \frac{1}{2} + \epsilon_1 (-1)^{t_1} + \epsilon_2 (-1)^{t_2} + 2\epsilon_1\epsilon_2 (-1)^{t_1} (-1)^{t_2}
\]

\[
= (\epsilon_1 (-1)^{t_1} + \frac{1}{2})(\epsilon_2 (-1)^{t_2} + \frac{1}{2})
\]

\[
= \Pr[ T_1 = t_1] \Pr[ T_2 = t_2]
\]
Linear Attack on the SPN
The four random variables have biases that are high in absolute value. By the Piling-Up Lemma we get the linear approximation

\[ T = X_5 \oplus X_7 \oplus X_8 \oplus U_6^4 \oplus U_8^4 \oplus U_{14}^4 \oplus U_{16}^4 \quad (3.3) \]

with bias \( |2^3 \left( \frac{1}{4} \right)^4| = \frac{1}{32} \) 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 $U_6^4$, $U_8^4$, $U_{14}^4$, and $U_{16}^4$, then the distribution of $T$ 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.