Linear Cryptanalysis

See:

Motivation
1. All the operations except for the S boxes are linear.
2. Mixing the key in all the rounds is done by a linear operation (XOR).
3. Approximate linear relations between input bits and output bits of the S boxes can be derived.
4. Approximate linear equations of the plaintext/plaintext and key bits by
   \[ P_{i+2r} \oplus C_{i+2r} \oplus K_{i+2r} = 0 \]
   with some probability (\( \neq 1/2 \)).

Linear Cryptanalysis
The second and most successful method to attack DES. Reduces the complexity of attacking DES to \( 2^{43} \) known plaintexts.

Note: As in differential cryptanalysis we ignore the existence of the initial and the final permutations, since they do not affect the analysis.

Linear Approximations of the S Boxes
Assume \( X \) is an input of an S box, and \( Y \) is the output of the S box: \( Y = S(X) \).

Notation: Denote the bits of \( X \) and \( Y \) by \( X = X_1X_2X_3X_4X_5 \) and \( Y = Y_1Y_2Y_3Y_4 \).

Notation: \( X' = X_1'X_2'X_3'X_4' \) and \( Y' = Y_1'Y_2'Y_3'Y_4' \) represent subsets of the bits of \( X \) and \( Y \):
\[ \{X'\} = \{i|X' = 1, i \in \{1, 2, 3, 4, 5\}\} \]
\[ \{Y'\} = \{i|Y' = 1, i \in \{1, 2, 3, 4\}\} \]

Notation: Let \( XX' \) be the (boolean) scalar product of the vectors \( X \) and \( X' \) (i.e., the parity of the bits of \( X \) chosen by \( X' \)) (and similarly for \( YY' \)).

Linear Approximation Tables — Probability

**Definition:** The probability of an entry \([X'; Y']\) in a linear approximation table is the fraction of inputs \( X \) whose parity \( XX' \oplus YY' = 0 \).

**Result:** The probability of an entry \([X'; Y']\) is \( 1/2 \) plus the value in the entry divided by the number of possible inputs of the S box (64 in DES).

**Example:** The probability of an entry with value 8 is \( 1/2 + 8/64 = 1/2 + 1/8 \), and the probability of the entry \([0, 1]\) in S5 is \( 1/2 - 20/64 = 12/64 \).
Linear Characteristics

In linear cryptanalysis we study the behavior of parities of subsets of bits during encryption, and we wish to find statistical information on the parity of a subset of ciphertext bits and key bits given the parity of a subset of plaintext bits.

As in differential cryptanalysis we define characteristics, however, linear characteristics do not represent differences: they represent the subsets of bits participating in the parity approximation.

Note that linear characteristics represent Matsui’s linear relations, in a different notation.

**Linear Characteristics (cont.)**

**Definition:** An $n$-round characteristic:

\[ \Omega^1 = (\Omega^1_1, \Omega^1_2, \Omega^1_N, 1/2 + p_1) \]

can be concatenated with an $m$-round characteristic:

\[ \Omega^2 = (\Omega^2_1, \Omega^2_2, \Omega^2_N, 1/2 + p_2) \]

if $\Omega^1_N$ equals the swapped value of the two halves of $\Omega^2_N$. The concatenation of the characteristics $\Omega^1$ and $\Omega^2$ (if they can be concatenated) is the $(n+m)$-round characteristic:

\[ \Omega = (\Omega^1_1, \Omega^1_2, \Omega^1_N \oplus \Omega^2_N, 1/2 + 2 - p_1 - p_2). \]

---

**Linear Characteristics - Examples**

The best one-round characteristic does not have any active S box. This characteristic has probability 1:

\[ \begin{array}{c|c}
    0 & 0 \\
    \hline
    0 & 1 \\
    \hline
    1 & 1 \\
    \hline
    1 & 0 \\
\end{array} \]

with probability 1, no affected key bits

\[ \begin{array}{c|c}
    0 & 0 \\
    \hline
    0 & 1 \\
    \hline
    1 & 1 \\
    \hline
    1 & 0 \\
\end{array} \]

---

**Linear Characteristics - Examples (cont.)**

The best three round characteristic is the concatenation of the above two one-round characteristics:

\[ \begin{array}{c|c|c|c|c|c}
    0 & 1 & 0 & 0 & 0 & 0 \\
    \hline
    0 & 1 & 0 & 1 & 0 & 0 \\
    \hline
    1 & 0 & 1 & 0 & 1 & 0 \\
    \hline
    1 & 0 & 0 & 1 & 0 & 1 \\
\end{array} \]

with probability 1/2 - 20/36 = 1/2 - 5/9, no affected key bits

\[ \begin{array}{c|c|c|c|c|c}
    0 & 1 & 0 & 0 & 0 & 0 \\
    \hline
    0 & 1 & 0 & 1 & 0 & 0 \\
    \hline
    1 & 0 & 1 & 0 & 1 & 0 \\
    \hline
    1 & 0 & 0 & 1 & 0 & 1 \\
\end{array} \]

with probability 1/2 - 20/36 = 1/2 - 5/9, no affected key bits

\[ \begin{array}{c|c|c|c|c|c}
    0 & 1 & 0 & 0 & 0 & 0 \\
    \hline
    0 & 1 & 0 & 1 & 0 & 0 \\
    \hline
    1 & 0 & 1 & 0 & 1 & 0 \\
    \hline
    1 & 0 & 0 & 1 & 0 & 1 \\
\end{array} \]

with probability 1/2 - 20/36 = 1/2 - 5/9, no affected key bits

\[ \begin{array}{c|c|c|c|c|c}
    0 & 1 & 0 & 0 & 0 & 0 \\
    \hline
    0 & 1 & 0 & 1 & 0 & 0 \\
    \hline
    1 & 0 & 1 & 0 & 1 & 0 \\
    \hline
    1 & 0 & 0 & 1 & 0 & 1 \\
\end{array} \]

it has probability 1/2 + 2/(\Omega^1_2) = 1/2 + 25/128.

---

**Linear Characteristics (cont.)**

**Definition:** A one-round characteristic is a tuple

\[ (\Omega_1, \Omega_2, \Omega_K, 1/2 + p), \]

in which $(\Omega_1)_L = (\Omega_2)_L = \Lambda$, $(\Omega_1)_R \oplus (\Omega_2)_R = \alpha$, and in which $1/2 + p$ is the probability that a random input block $P$ and its one-round encryption $C$ under the key $K$ satisfies $P \cdot \Omega_1 \oplus K \cdot \Omega_2 = 0$. $\Omega_2$ is the subset of bits from the data before the round, $\Omega_1$ is the subset of bits from the data after the round, and $\Omega_K$ is the subset of bits from the key whose parity is approximated.

---

**Linear Characteristics (cont.)**

The probability $1/2 + 2 - p_1 - p_2$ holds since the expected zero parity is achieved either if both original parities are zero, or if both original parities are one, and

\[ (1/2 + p_1)(1/2 + p_2) + (1/4 - p_1)(1/4 - p_2) = \frac{1}{2} + \frac{1}{2}p_1 + \frac{1}{2}p_2 + \frac{1}{2}p_1p_2 \]

\[ = \frac{1}{2} + 2p_1p_2. \]

When we concatenate $I$ characteristics (that can be concatenated) the probability of the resultant characteristic is

\[ 1/2 + p = 1/2 + 2^{I-1} \sum_{i=1}^{I} p_i. \]
Matsui’s Best Characteristic (cont.)

This eight-round characteristic has probability about $1/2 + 2^{-17}$. Related to a 16-round characteristic, it has probability about $1/2 + 2^{-34}$. Matsui got a 16-round characteristic with probability about $2^{-21}$, by replacing the first and last rounds by locally better ones.

Properties of Linear Characteristics

In differential cryptanalysis, when the data is duplicated during encryption, the number of times the characteristic predicts each possible value of the parity of the XORed key bits is counted in the reverse order to the DES standard.

Another Notation

Note that the bit indices here are counted in the reverse order to the DES standard.

Properties of Linear Characteristics (cont.)

In duplication, we should keep the parity the same as of the original, thus when we have $Z = Y = X$, and $X$ participate in the parity, then we can replace $X$ by $Y := Z$, if $X$ does not participate in the parity, then $X := Y := Z = 0$.

Moreover, we can also use bits which do not appear in the original subset. When we have $Z = Y = X$, and $X$ does not participate in the parity, then we can add both $Y$ and $Z$ since we actually add $Y \oplus Z = 0$.

Properties of Linear Characteristics (cont.)

This characteristic uses the expansion of two bits into both $S$ boxes $S_7$ and $S_9$. The input subset is $a' = 0$ — the empty subset. However, the characteristic assumes that some bits are duplicated to set both common bits in the input subsets of the $S$ boxes. Since they are added twice (once in each $S$ box) they cancel each other in $a' = 0$, and we get an empty input difference with a non-empty output difference. This characteristic is very similar to the iterative characteristic of differential cryptanalysis, except that the zero and the non-zero values exchange their sides.

Properties of Linear Characteristics (cont.)

A two round iterative characteristic:

$$S_0 = 00 00 00 \oplus Z,$$

$$S_1 = 00 00 00 \oplus Y.$$
Linear Attacks (cont.)

Let the linear expression predicted by the characteristic be

$$Pp_T \oplus C p_T = K p_R$$

with probability $1/2 + p$ for some $p \neq 0$.

**Algorithm 1**:

1. Given $N$ plaintext/ciphertext pairs, for a sufficiently large $N$, let $M$ be the number of plaintexts whose left hand side of the above equation equals to zero:

$$Pp_T \oplus C p_T = 0.$$  

2. The algorithm guesses that the parity of the key bits $K p_R$ is

$$(M > N/2) \oplus (1/2 + p > 1/2),$$

i.e.,

| $M > N/2$ | 1 |
| $M < N/2$ | 0 |

Clearly, as the number of plaintexts $N$ is larger, and as the probability’s distance from half $|p|$ is larger, the success rate of this algorithm is larger as well. This algorithm finds only one parity bit of the key for each characteristic.

Linear Attacks (cont.)

The following algorithm have two advantages over the previous one:

1. It uses shorter characteristics, and thus they have higher $|p|$.
2. It finds several key bits, in addition to the parity of $K p_R$.

Linear Attacks (cont.)

We are interested in the number $N$ of known plaintexts required for the attacks.

**Lemma**. Let $N$ be the number of random plaintexts and $1/2 + p$ the probability that the equation

$$Pp_T \oplus C p_T = K p_R$$

holds, and assume that $|p|$ is sufficiently small. Then, the success rate of Algorithm 1 is

$$N - \frac{2p}{|p|} - \frac{2p}{|p|^2} - \frac{2p}{|p|^3} \mathrm{d}x.$$

The success rates are

| $N$ | Success rate | $84.1\%$ | $92.1\%$ | $97.7\%$ | $99.8\%$ |

In general, linear cryptanalysis attacks require $O(|p|^2)$ known plaintexts. Algorithm 2 also requires $O(|p|^2)$ known plaintexts with

Matsui’s Attack on DES

Matsui uses the best 14-round characteristic (derived from the above eight-round iterative characteristic) with probability about $1/2 \pm 2^{-30}$, and tries all the possible values of 12 key bits: six in the first round and six in the last round, by algorithm 2.

This attack requires only $2^{15}$ known plaintexts. Matsui has implemented this attack on a network of workstations, and found the key within a few months.

Linear Attacks (cont.)

Dependence of the Probabilities on the Key

The probability of the characteristics (and thus of the differentials) is defined as the probability that a random pair is a right pair with respect to a random key. However, for fixed keys the probability that a random pair is a right pair might differ from this probability, and in actual attacks we are interested in the probability that a random pair is a right pair with respect to the used secret key.
Dependence of the Probabilities on the Key (cont.)

Example: The approximations

\[ S_1: X[4] \oplus Y[9, 17, 31] = K[5], \quad p = 1/2 - 2/64 \]
\[ S_2: X[4] \oplus Y[2, 18, 28] = K[7], \quad p = 1/2 \]

are combined to the characteristic with probability 1/2:

\[ \Omega P = 40 \quad 80 \]
\[ C_0 \quad 12 \quad 00 \quad 00 \quad 00 \quad 00 \]
\[ A' = 40 \quad 80 \]
\[ C_0 \quad 12 \]
\[ a' = 0 \]
\[ p = 1/2 + 2^{-2/64} = 1/2 \]

Since it has probability exactly 1/2, it seems that it cannot be used for linear attacks!

Dependence of the Probabilities on the Key (cont.)

There are also examples in which

1. The defined probability is smaller than 1/2, but for most fixed keys it is higher than 1/2, and
2. The defined probability is smaller than half, but for many keys it is half.

For these keys linear cryptanalysis might not be used.
3. And others.

Linear Differentials

In differential cryptanalysis, a characteristic’s probability is a lower bound for the probability of the differential.

In linear cryptanalysis it is not!

When there are several characteristics with the same \( \Omega P \) and same \( \Omega T \) – i.e., a differential – the probabilities might add, but also might cancel each other.

For example such cancellation occurs when

1. the probabilities of the two characteristics are similar but the parities of the key subsets are different for the particular secret key.
2. when the probabilities of the two characteristics are complements (1/2 + p and 1/2 − p) and the parities of the key subsets are the same.

Ciphertext Only Linear Cryptanalysis

Linear cryptanalysis can in certain cases be applied to ciphertext only attacks, if the attacker knows the parity of \( P \Omega P \) by the assumptions on the plaintext language, or even if the attacker only knows the distribution of the parity \( P \Omega P \) in the unknown plaintexts.

In particular, if ASCII text is used, the attacker knows that the most significant bit of every byte is zero. If he finds a characteristic whose plaintext subset \( \Omega P \) contains only these bits, he can apply the corresponding attack without knowing the plaintext at all.

Extensions of Linear Cryptanalysis

1. Differential-Linear cryptanalysis (Hellman, Langford)
2. Linear cryptanalysis with multiple approximations (Robshaw, Kaliski)
3. Statistical cryptanalysis (Vaudenay)
4. Partitioning cryptanalysis (Harpes, Massey)
5. Interpolations attacks (Knudsen)
6. Matsui’s algorithm for finding best characteristics (for both differential and linear cryptanalysis)
7. Constraints on the size of S boxes and sums of characteristics
8. Non-Linear approximations in linear cryptanalysis (Knudsen, Robshaw)
9. Reduction to \( 25/34 \approx 74\% \) of the number of chosen plaintexts required by Matsui (March 1998).