

Introduction to Linear Cryptanalysis

Johan Wallén

Helsinki University of Technology
Laboratory for Theoretical Computer Science
johan@tcs.hut.fi

Overview

Linear cryptanalysis [Matsui, 1993] was the first known plaintext key recovery attack on DES that was faster than a brute force search (complexity: around 2^{43} known plaintexts and similar time complexity).

The attack on DES was verified experimentally already in [Matsui, 1994].

Linear cryptanalysis is still the best attack on DES and is one of the most efficient cryptanalytic methods for block and stream ciphers.

Since its introduction, resistance against the attack has been one of the standard design goals for new ciphers.

Overview (cont.)

Linear cryptanalysis studies the correlation between linear combinations of input and output bits of functions.

In this talk, we will give a brief introduction to the theory behind linear cryptanalysis including linear approximation, linear distinguishers, the basic key recovery attack, Fourier analysis, correlation matrices and linear trails.

Our focus is on block cipher cryptanalysis, but the theory is also relevant to stream cipher cryptanalysis.

Warning: the terminology and notation varies a lot from author to author. This talk is no exception. . .

Linear functionals

Our approach is maybe a bit more abstract than necessary. This allows us to ignore some details.

Let X be a finite dimensional inner-product space over $\mathbb{F}_2$. We denote the inner product of x and w by (x, w) .

We will often view w as the linear functional $x \mapsto (x, w)$.

In applications of linear cryptanalysis, one usually have $X = \mathbb{F}_2^n$ or $X = \mathbb{F}_{2^k}^n$ for some n and k .

When $X = \mathbb{F}_2^n$, it is natural to define $(x, w) = \sum_i x_i w_i$. That is, w **selects** the bit positions with $w_i = 1$ and (x, w) is the **parity** of the bits of x selected by w .

Linear functionals (cont.)

When $X = \mathbb{F}_{2^k}^n$, the set $\mathbb{F}_{2^k}^n$ is interpreted as a vector space of dimension kn over $\mathbb{F}_2$ in the natural way (same operations, but the scalar multiplication restricted to $\{0, 1\}$).

The inner product is defined by $(x, w) = \sum_i \text{Tr}(x_i w_i)$ for all $x, w \in \mathbb{F}_{2^k}^n$, where $x_i w_i$ denotes multiplication in $\mathbb{F}_{2^k}$ and $\text{Tr}: \mathbb{F}_{2^k} \rightarrow \mathbb{F}_2$ is the trace operator $\text{Tr}(y) = y + y^2 + \dots + y^{2^{k-1}}$.

This definition is convenient when the cipher uses field operations, since we do not have to specify how $\mathbb{F}_{2^k}^n$ is represented as nk -tuples of bits.

Boolean functions

Let X be a finite dimensional inner-product space over $\mathbf{F}_2$, and let $f, g: X \rightarrow \mathbf{F}_2$ be Boolean functions. Let $n = \dim X$.

The **correlation** between f and g is

$$\begin{aligned}\text{corr}(f, g) &= 2^{1-n} |\{x \in X \mid f(x) = g(x)\}| - 1 \\ &= 2^{-n} (|\{x \in X \mid f(x) = g(x)\}| - |\{x \in X \mid f(x) \neq g(x)\}|).\end{aligned}$$

This is simply the probability taken over x that $f(x) = g(x)$ scaled to a value in $[-1, 1]$.

As a notational convenience, we define $l_w: X \rightarrow \mathbf{F}_2$ by $l_w(x) = (x, w)$ for all $w \in X$.

Linear approximations

Let X and Y be finite dimensional inner-product spaces over $\mathbb{F}_2$. Let $f: X \rightarrow Y$ be a function.

A **linear approximation** of f is a pair of linear functionals $(u, w) \in Y \times X$. This is interpreted as the approximate relation $(f(x), u) = (x, w)$.

The strength of the approximation is measured by its **correlation**

$$\begin{aligned}\text{corr}_f(u, w) &= 2^{1-\dim X} |\{x \in X \mid (f(x), u) = (x, w)\}| - 1 \\ &= \text{corr}(l_u \circ f, l_w).\end{aligned}$$

This is simply the probability that the relation holds scaled to a value in $[-1, 1]$.

Linear distinguishers

We will describe a basic distinguishing attack on block ciphers using linear approximations. More precisely, we consider the following decision problem.

Let $f: X \rightarrow X$ be a function that is selected either (1) by choosing a key uniformly at random for a (fixed) block cipher, or (2) by choosing a permutation uniformly at random. We are given N known plaintext/ciphertext pairs $(x_i, f(x_i))$. Our goal is to decide if f is selected from (1) or (2).

We assume that the plaintexts x_i are disjoint and uniformly distributed (that is, we choose uniformly at random a subset $\{x_i\} \subset X$ of size N).

Linear distinguishers (cont.)

We furthermore assume that we have a linear approximation (u, w) with $u \neq 0$ such that the expectation of the square of the correlation $\mathbf{E}_f \text{corr}_f(u, w)^2$ is “fairly large” when f is sampled from the block cipher.

Note that the case $u = 0$ is of no interest (we would then ignore the ciphertext).

For simplicity, we assume that we are given an oracle $C(x)$ such that $C(x) = 1$ if $(f(x), u) = (x, w)$ and $C(x) = 0$ otherwise. (This can easily be implemented using an oracle for f .)

Linear distinguishers (still cont.)

The distinguisher uses an integer parameter $N/2 < \alpha < N$. The distinguisher operates as follows.

1. Choose uniformly at random a subset $\{x_i\} \subset X$ of size N .
2. Compute $c = \sum_i C(x_i)$.
3. Accept (that is, guess that f is the real cipher) if and only if $c \leq N - \alpha$ or $c \geq \alpha$.

Note that the computational cost is roughly N encryptions (including the computation of the oracle).

Note also that this can be seen as a known plaintext attack, if the plaintext distribution is “close” to uniform.

Linear distinguishers: success probability

The analysis of the success probability is based on the following observations. For simplicity, we assume that $w \neq 0$.

Suppose that f is selected from all permutations. Since the oracle queries are disjoint, we can replace the oracle C with an oracle that ignores the query and returns 1 with probability $1/2$.

Thus, the value $c = \sum_i C(x_i)$ is binomially distributed with parameters $1/2$ and N . The acceptance probability is

$$\sum_{k=0}^{N-\alpha} \binom{N}{k} 2^{-N} + \sum_{k=\alpha}^N \binom{N}{k} 2^{-N}.$$

Linear distinguishers: success probability (cont.)

Suppose that f is selected from the block cipher. For simplicity, we assume that the correlation for the selected key is $\epsilon = \text{corr}_f(u, w) > 0$. Let $p = 1/2 + \epsilon$.

Since the oracle queries are a uniformly selected subset of size N , it follows that the value $c = \sum_i C(x_i)$ is binomially distributed with parameters p and N . Thus, the acceptance probability is

$$\sum_{k=0}^{N-\alpha} \binom{N}{k} p^k (1-p)^{N-k} + \sum_{k=\alpha}^N \binom{N}{k} p^k (1-p)^{N-k}.$$

Linear distinguishers: success probability (still cont.)

The success probability is then computed by bounding the tails of the binomial distributions.

If the correlation of the linear approximation is $\epsilon = \text{corr}_f(u, w) > 0$, the optimal choice of the parameter α is the nearest integers to

$$N \cdot \frac{\log(1 - 2\epsilon)}{\log(1 - 2\epsilon) - \log(1 + 2\epsilon)}.$$

If ϵ is small, this is about $N(1/2 + \epsilon/2)$.

The expected success probability can then be expressed using the expectation $\mathbb{E}_f \text{corr}_f(u, w)^2$ when f is selected from the block cipher. (Is this actually a good idea?)

Linear distinguishers: some remarks

The analysis of the success probability is technical, and thus unsuitable for a short talk. We refer to [Vaudenay, 2003] and [Junod, 2003] for analyses.

The number of plaintexts needed for a high success probability is usually considered to be around $1 / \mathbb{E}_f \text{corr}_f(u, w)^2$, where the expectation is taken over the block cipher.

(If the expectation is taken over all permutations and $u, w \neq 0$, it is easy to show that $\mathbb{E}_f \text{corr}_f(u, w)^2 = \frac{1}{2^n - 1}$, where $n = \dim X$, by expressing the correlation using indicator variables or by direct counting.)

The basic key recovery attack

Using similar ideas, the distinguisher can be transformed into a key recovery attack.

We consider a generic r -round iterated cipher $f_r \cdots f_1$, where each $f_i: X \rightarrow X$ might depend on an unknown key k_i .

The basic attack uses an $r - 2$ -round linear approximation from round 2 to $r - 1$.

That is, we have an approximate relation of the form

$$(f_1(x, k_1), a) + (f_r^{-1}(y, k_r), b) = \sum_{i=2}^{r-1} (k_i, c_i), \quad (1)$$

where a, b, c_i are constants (the linear approximation), x is the plaintext and y is the ciphertext.

The basic key recovery attack (cont.)

Suppose that we are given N known plaintext-ciphertext pairs (x, y) . By guessing parts of the round keys k_1, k_r , the attacker can compute

$$(f_1(x, k_1), a) \oplus (f_r^{-1}(y, k_r), b)$$

for each plaintext-ciphertext pair, and count the number of times N_0 that this expression is zero.

If the guess is correct, one would expect that $2N_0/N - 1$ is plus or minus the correlation of the used linear approximation (1) (depending on the unknown constant $\sum_{i=2}^{r-1} (k_i, c_i)$).

If the guess is incorrect, one would expect that the values $(f_1(x, k_1), a)$ and $(f_r^{-1}(y, k_r), b)$ are uncorrelated and thus $2N_0/N - 1 \approx 0$.

The basic key recovery attack (still cont.)

Thus, we could sort the round subkeys candidates by the value $|2N_0/N - 1|$ and obtain a ranking of the most likely candidates.

We can then try the candidates in order and exhaustively search for the remaining part of the round keys, or use other linear approximations to recover more key material.

Under reasonable assumptions, the number of plaintexts needed for this attack is inversely proportional to the **square** of the correlation of the approximation just as for distinguishing attacks.

Note that the time complexity of this attack might be large.

Fourier analysis

There is a nice Fourier-theoretic framework for analysing the correlation of linear approximations.

Let X be an n -dimensional vector space over $\mathbb{F}_2$. For each Boolean function $f: X \rightarrow \mathbb{F}_2$, let $\hat{f}: X \rightarrow \mathbb{R}$ be the function $\hat{f}(x) = (-1)^{f(x)}$.

If $h(x) = f(x) + g(x)$, then $\hat{h}(x) = \hat{f}(x)\hat{g}(x)$.

Let $\mathcal{B}_n = \langle \hat{f} \mid f: X \rightarrow \mathbb{F}_2 \rangle$ be the real vector space generated by the Boolean functions (with pointwise addition and scalar multiplication). Define a multiplication by $(\xi\eta)(x) = \xi(x)\eta(x)$ for all $\xi, \eta \in \mathcal{B}_n$. Then $\mathcal{B}_n$ is a commutative real algebra with unit.

Fourier analysis (cont.)

Define an inner product in $\mathcal{B}_n$ by

$$(\xi, \eta) = 2^{-n} \sum_{x \in X} (\xi \eta)(x).$$

For Boolean functions $f, g: X \rightarrow \mathbf{F}_2$, we have $(\hat{f}, \hat{g}) = \text{corr}(f, g)$. Note that for all linear Boolean functions l_v, l_w ,

$$(\hat{l}_v, \hat{l}_w) = 2^{-n} \sum_x (-1)^{(x, v+w)} = \begin{cases} 1 & \text{if } v = w \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

For each Boolean function f , we have

$$\sum_w \left(2^{-n} \sum_x \hat{f}(x) \hat{l}_w(x) \right) \hat{l}_w(z) = \sum_x \hat{f}(x) 2^{-n} \sum_w (-1)^{(x, w) + (z, w)} = \hat{f}(z).$$

That is, $\{\hat{l}_w \mid w \in X\}$ is an orthonormal basis for $\mathcal{B}_n$.

Fourier analysis (still cont.)

It follows that each ξ in $\mathcal{B}_n$ has a unique representation as $\xi = \sum_w \alpha_w \hat{l}_w$, where $\alpha_w = (\xi, \hat{l}_w)$.

The corresponding Fourier transform $\mathcal{F}: \mathcal{B}_n \rightarrow \mathcal{B}_n$ is defined by $\mathcal{F}(\xi) = \Xi$, where Ξ is the mapping $w \mapsto (\xi, \hat{l}_w)$ (this is often called the Walsh-Hadamard transform). If $\hat{F} = \mathcal{F}(\hat{f})$, $\hat{F}(w) = \text{corr}(f, l_w)$.

The Fourier transform is clearly linear and bijective, and $\mathcal{F}(\xi\eta) = \mathcal{F}(\xi) * \mathcal{F}(\eta)$, where

$$(\nu * \mu) = \sum_{t \in X} \nu(x + t) \mu(t)$$

is the convolution operation in $\mathcal{B}_n$, as usual. It follows that $\mathcal{F}$ is an algebra isomorphism between the real algebras $\langle \mathcal{B}_n, +, \cdot \rangle$ and $\langle \mathcal{B}_n, +, * \rangle$.

Fourier analysis (still cont.)

Let $f: X \rightarrow Y$, where X and Y are binary vector spaces of dimensions n and m , respectively.

Let $\chi_f: X \times Y \rightarrow \mathbb{F}_2$ be the characteristic function of f , that is, $\chi_f(x, y) = 1$ if and only if $f(x) = y$. Then

$$\begin{aligned}\mathcal{F}(\chi_f)(w, u) &= 2^{-m-n} \sum_{x,y} \chi_f(x, y) (-1)^{(y,u)+(x,w)} \\ &= 2^{-m-n} \sum_x (-1)^{(f(x),u)+(x,w)} = 2^{-m} \text{corr}_f(u, w).\end{aligned}$$

Using the FFT, we can thus compute the correlation of all linear approximations of f in time $O((n + m)2^{n+m})$, or $O(n + m)$ per approximation.

Correlation matrices

Let $f: X \rightarrow Y$ and $g: Y \rightarrow Z$. Recall that

$$(-1)^{h(z)} = \sum_p \text{corr}(h, l_p)(-1)^{(z,p)} = \sum_p \text{corr}(h, l_p)(-1)^{l_p(z)}$$

for all Boolean valued functions h . Thus,

$$\begin{aligned} (-1)^{(l_u g f)(x)} &= \sum_v \text{corr}(l_u g, l_v)(-1)^{l_v(f(x))} \\ &= \sum_v \text{corr}(l_u g, l_v) \sum_w \text{corr}(l_v f, l_w)(-1)^{(x,w)} \\ &= \sum_w \left(\sum_v \text{corr}(l_u g, l_v) \text{corr}(l_v f, l_w) \right) (-1)^{(x,w)}. \end{aligned}$$

By the uniqueness of this expansion,

$$\text{corr}_{g \circ f}(u, w) = \sum_v \text{corr}_g(u, v) \text{corr}_f(v, w).$$

Correlation matrices (cont.)

If we let C^f be the $|Y| \times |X|$ matrix whose element in row u and column w is $\text{corr}_f(u, w)$, we thus have $C^{g \circ f} = C^g C^f$.

The matrix C^f is called the **correlation matrix** of f . They were introduced by [Daemen, 1995, Daemen et al., 1995].

If $f: X \rightarrow X$ is a bijection, we have

$$\text{corr}((f(x), u), (x, w)) = \text{corr}((x, u), (f^{-1}(x), w)).$$

It follows that $C^{(f^{-1})} = (C^f)^t$. From this, it is easy to see that f is invertible if and only if $(C^f)^t = (C^f)^{-1}$.

Correlation matrices (still cont.)

If f is the function $f(x) = x + k$, $(x + k, u) = (x, w)$ if and only if $(x, u + w) = (k, w)$. It follows that C^f is a diagonal matrix with $C_{uw}^f = (-1)^{(k,w)}$.

If L is a linear function, we have $(Lx, u) = (x, w)$ if and only if $(x, L^t u) = (x, w)$. It follows that $C_{u,w}^L = 1$ if $L^t u = w$ and $C_{u,w}^L = 0$ otherwise.

If $f(x) = (f_1(x_1), \dots, f_k(x_k))$, where $x = (x_1, \dots, x_k)$, and the inner products are defined componentwise $((x, w) = \sum_i (x_i, w_i))$, it is easy to see using the convolution property of the Fourier transform that $C_{u,w}^f = \prod_i C_{u_i, w_i}^{f_i}$, where $u = (u_1, \dots, u_k)$ and $w = (w_1, \dots, w_k)$.

We conclude that the correlation matrix of most ciphers (that do not use arithmetic operations) can be expressed as the product of a few simple matrices.

Linear trails

In practise, good linear approximations for ciphers are found by forming trails of linear approximations for the components. Suppose that $f = f_{r-1} \cdots f_0$, where $f_i: X_i \rightarrow X_{i+1}$.

A **linear trail** for f is a sequence $(w_r, \dots, w_0)$ with $w_i \in X_i$. The correlation of a trail is defined to be

$$\text{corr}_f(w_r, \dots, w_0) = \prod_i \text{corr}_{f_i}(w_{i+1}, w_i).$$

Then the correlation of approximations is given by the trail correlations,

$$\text{corr}_f(u, w) = \sum_{\substack{w_r, \dots, w_0 \\ w_r = u, w_0 = w}} \text{corr}_f(w_r, \dots, w_0).$$

Linear trails (cont.)

If the sum is dominated by a single trail, it is common to use it to approximate the correlation of the linear approximation.

Note that the correlation of a trail can usually be efficiently computed, while the actual correlation cannot be efficiently computed.

Furthermore, it is often possible to search the space of all trails for a cipher and find the ones with the highest correlation. These trails (and not the actual linear approximations) are used for linear attacks in practise.

Thus, some heuristics are involved.

Suggestion for future presentations

Some topics on linear cryptanalysis not covered in this talk are:

1. Linear cryptanalysis of stream ciphers.
2. Concrete linear attacks on ciphers.
3. Detailed analysis of the success probability of the distinguishers.

The first two topics would be very suitable for future presentations, but the third is probably too technical. Volunteers for these and other topics should contact Helger. Slots are free for the next seminars.

References

Joan Daemen. *Cipher and Hash Function Design: Methods Based on Linear and Differential Cryptanalysis*. PhD thesis, Katholieke Universiteit Leuven, March 1995.

Joan Daemen, René Govaerts, and Joos Vandewalle. Correlation matrices. In *Fast Software Encryption 1994*, volume 1008 of *Lecture Notes in Computer Science*, pages 275–285. Springer-Verlag, 1995.

Pascal Junod. On the optimality of linear, differential, and sequential distinguishers. In *Advances in Cryptology—Eurocrypt 2003*, volume 2656 of *Lecture Notes in Computer Science*, pages 17–32. Springer-Verlag, 2003.

Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology—Eurocrypt 1993*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397. Springer-Verlag, 1993.

Mitsuru Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology—Crypto 1994*, volume 839 of *Lecture Notes in Computer Science*, pages 1–11. Springer-Verlag, 1994.

Serge Vaudenay. Decorrelation: A theory for block cipher security. *Journal of Cryptology*, 16(4):249–286, 2003.