On Linear Cryptanalysis Using Multiple Linear Approximations

Echternach, January 7, 2008

Kaisa Nyberg

kaisa.nyberg@tkk.fi

Department of Information and Computer Science
Helsinki University of Technology
and Nokia Research Center, Finland

Outline

1. Linear Cryptanalysis; Matsui's Algorithm 1 and 2 on DES
2. Distinguisher and Statistical Distinguishing Attack
3. Hypothesis Testing and Data Complexity
4. Survey on Previous Work
5. Multidimensional Linear Cryptanalysis
6. Linear Approximation of Boolean Functions
7. Linear Resistance
8. Conclusions

Surveyd papers

1. Biryukov, et al. On Multiple Linear Approximations. Crypto 2004
2. Baignères, et al. How Far Can We Go Beyond Linear Cryptanalysis? Asiacrypt 2004
3. Maximov, et al.
   (a) Maximov. Some Word on Cryptanalysis of Stream Ciphers. Thesis 2006.
   (b) Englund and Maximov. Attack the Dragon. INDOCRYPT 2005

Linear Cryptanalysis

- One of the most powerful cryptanalytic attacks on DES
- Matsui 1993
- Notation:

  Let m and n be integers, $m \leq n$.

  Data $X = (X_1, X_2, \ldots, X_n)$

  Mask $\xi = (\xi_1, \xi_2, \ldots, \xi_m)$,

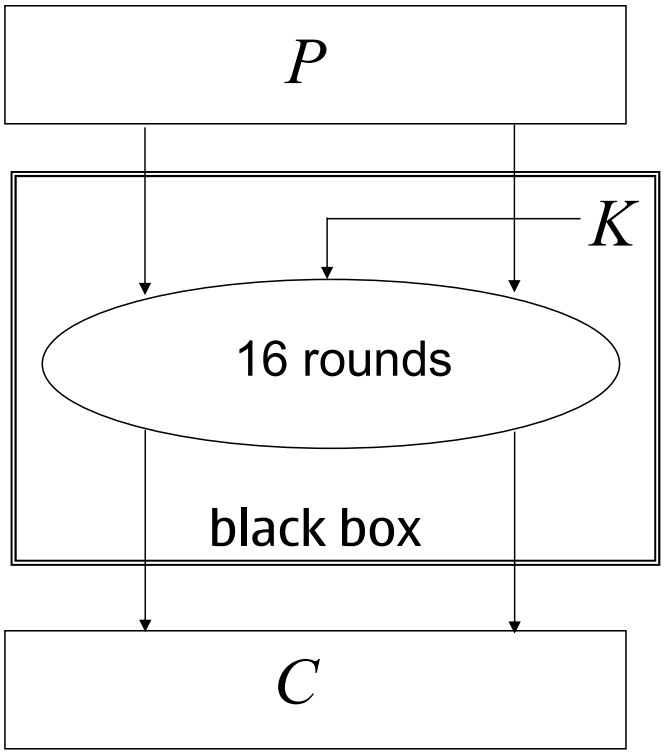
  where $1 \leq \xi_1 < \xi_2 < \ldots < \xi_m \leq n$

  Masked data:

  $X(\xi) = X_{\xi_1} \oplus X_{\xi_2} \oplus \ldots \oplus X_{\xi_m}$

  addition modulo 2, substring parity

Algorithm 1



Relation

$$P(\pi) \oplus C(\gamma) = K(\kappa)$$

holds with bias

$$\varepsilon = p - \tfrac{1}{2} \neq 0$$

If $\varepsilon > 0$ and known P, C pairs support

$$P(\pi) \oplus C(\gamma) = 1$$

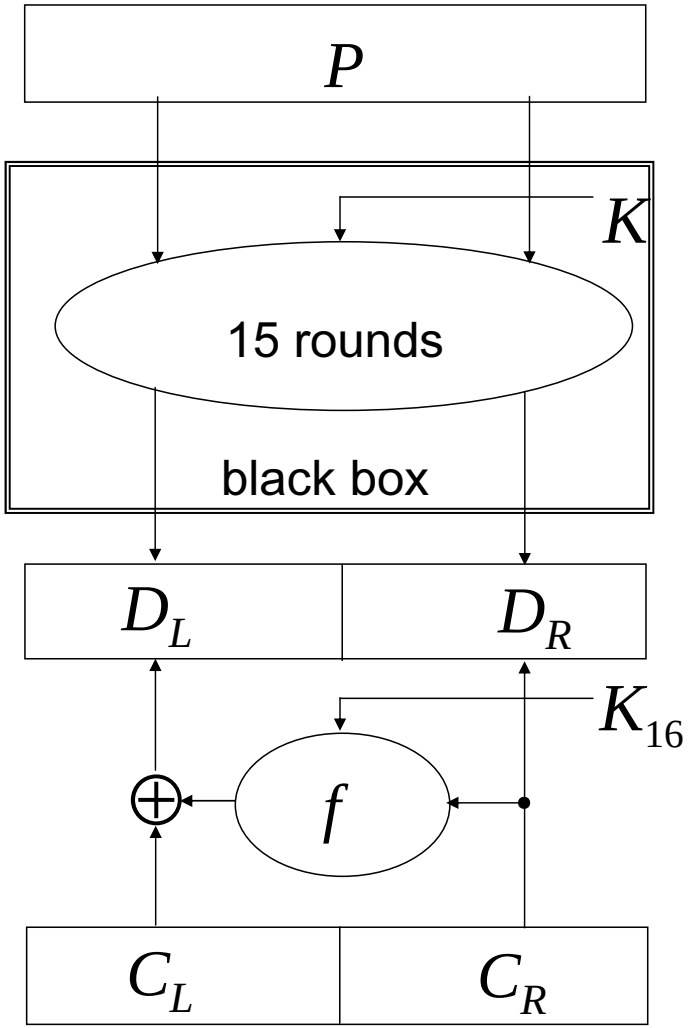
select $K(\kappa)$=1, else $K(\kappa)$= 0.

If $\varepsilon < 0$ and known P, C pairs support

$$P(\pi) \oplus C(\gamma) = 1$$

select $K(\kappa)$=0, else $K(\kappa)$= 1.

Algorithm 2



Decrypt over the last round with K_{guess}

$$D_L = C_L \oplus f(K_{guess}, C_R)$$

Relation

$$P(\pi) \oplus D(\delta) = K(\kappa)$$

holds with absolute bias

$$\left| p - \tfrac{1}{2} \right| \neq 0$$

if $K_{guess} = K_{16}$, but with much less bias, i.e., the behavior is more random, if $K_{\text{guess}} \neq K_{16}$.

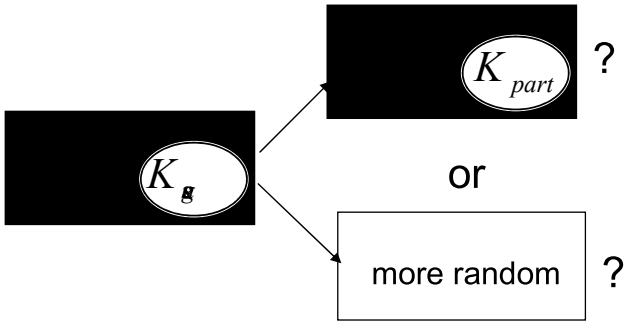
Distinguishers

Matsui's Algorithm 1 is of Type 1, and Algorithm 2 is of Type 2.

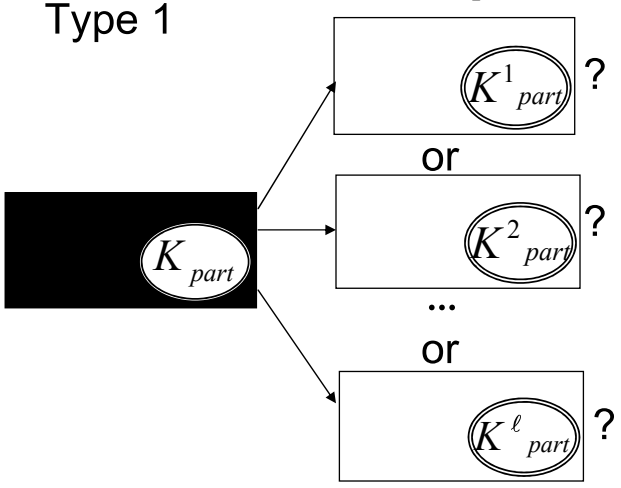
Derived reduced functionality as black box

Distinguisher is a reduced functionality possibly dependent of a part of the key. Two types of distinguishers for information deduction:

Distinguishing from more random
Type 2



Selecting the correct K_{part}
Type 1



Statistical Distinguishing Attack

Distinguishing attack operates in two phases

- **Design phase**

  Design of Distinguisher

  Distinguisher is a method to transform empirical data from the cipher to test data with known statistics

- **Statistical Inference phase**

  Statistical hypothesis testing is performed on the test data.

Challenge: Find a transformation which gives distinguishable statistics with as little data as possible.

Hypothesis Testing

- Probability distributions

  $\mathcal{D}_1$ estimated theoretical distribution from the cipher

  $\mathcal{D}_0$ some other known distribution, e.g., uniform distribution

- Hypotheses

  H_0: test data is drawn from $\mathcal{D}_0$

  H_1: test data is drawn from $\mathcal{D}_1$

- Statistics computed from the data

- Estimation of the data complexity N for sufficient confidence level.

Biryukov, et al.: Multiple Approximations (1)

- Multiple linear approximations:

$$
\begin{aligned}
X(\xi^1) \oplus Y(\eta^1) &= K(\kappa^1) \text{ with bias } \varepsilon_1 \\
X(\xi^2) \oplus Y(\eta^2) &= K(\kappa^2) \text{ with bias } \varepsilon_2 \\
&\cdots \\
X(\xi^m) \oplus Y(\eta^m) &= K(\kappa^m) \text{ with bias } \varepsilon_m
\end{aligned}
$$

- Biryukov, et al.: Generalisations of Matsui's Algorithms 1 and 2
- Assumption: Linear approximations statistically independent.
- Simulations on DES.
- Performance for ciphers with uniformly low biases (such as AES)?

Biryukov, et al.: Multiple Approximations (2)

Sample data $Z_1, Z_2, Z_3, \ldots, Z_N$, $Z_i = (P_i, D_i)$

m separate two-valued distributions considered

$p_j = 1 - q_j = (1 + c_j)/2 = \frac{1}{2} + \varepsilon_j$, $j = 1, \ldots, m$

Assumption: Distributions statistically independent.

Empirical distribution $\hat{p}_j = \frac{N_j}{N}$, $j = 1, \ldots, m$,

where $N_j = \#\{i \,|\, P_i(\xi^j) \oplus D_i(\eta^j) = 0\}$. Set $\hat{c}_j = 2\hat{p}_j - 1$.

The relative likelihood of a key class $z = (z_1, z_2, \ldots, z_m)$ is determined by the Euclidean distance (ℓ_2-distance) between the vectors $\hat{c}$ and $c_z = ((-1)^{z_1} c_1, \ldots, ((-1)^{z_m} c_m)$. That is, the least squares regression is used to select the key z, for which the the data fits the model best.

Required data N is inversely proportional to the "capacity" = $\sum_{j=1}^{m} \varepsilon_j^2$.

Baignères, et al.: Multidimensional Approximations in Markov Model (1)

Presented the optimal statistics for distinguishing between two distributions:

$\mathcal{D}_1$ theoretical distribution from the cipher with variance Var

$\mathcal{D}_0$ a different distribution, e.g., the uniform distribution

H_0: test data is drawn from $\mathcal{D}_0$

H_1: test data is drawn from $\mathcal{D}_1$

Sample data $Z_1, Z_2, Z_3, \ldots, Z_N$,

Optimal statistics: Log-likelihood ratio

$$LLR = \sum_{i=1}^{N} \log_2 \frac{\mathcal{D}_1(Z_i)}{\mathcal{D}_0(Z_i)} = \sum_{z \in Z} N_z \log_2 \frac{\mathcal{D}_1(z)}{\mathcal{D}_0(z)}$$

where $N_z = \#\{i \mid Z_i = z\}$.

Baignères, et al.: Multidimensional Approximations in Markov Model (2)

Required data N is inversely proportional to

$$\sum_{z} \frac{(\mathcal{D}_0(z) - \mathcal{D}_1(z))^2}{\mathcal{D}_0(z)}$$

If $\mathcal{D}_0$ uniform, then $N \approx 1/2^m$Var.

Baignères, et al., investigated probability distributions related to the transition matrices of approximations in the Markov cipher model, that is, the conditional probabilities

$$Pr[Y(\eta)|X(\xi)]$$

averaged over the key.

Maximov, et al.: Multidimensional Linear Cryptanalysis of Stream Ciphers

Distinguishing attacks on stream ciphers.

Multidimensional linear approximations over Boolean vector functions $f : GF(2)^n \to GF(2)^m$, $f(X) = Y$ in stream cipher constructions. Let $M : GF(2)^n \to GF(2)^n$ be linear. The m-bit noice is defined as

$$Z = MX \oplus Y,$$

and it takes 2^m values.

$\mathcal{D}_1$ is the distribution of the noice Z with variance Var.

$\mathcal{D}_0$ is the uniform distribution.

Data complexity $N \approx 1/2^m$Var.

(Originally, Maximov, et al., used the statistical distance, that is, ℓ_1-norm to give an estimate of the data complexity.)

Multidimensional Linear Approximation of Block Ciphers – Fixed Key Approach

$\mathcal{D}_1$ is the probability distribution of $Z = (Z_1, \ldots, Z_m)$, where:

$$Z_j = X(\xi_j) \oplus Y(\eta_j) \oplus K(\kappa_j), \text{ with bias } \varepsilon_j.$$

Consider the linear span of m linear approximations:

$$Z(\mu) = X(\xi) \oplus Y(\eta) \oplus K(\kappa), \text{ with bias } \varepsilon_\mu,$$

$\mu = (\xi, \eta, \kappa) \in \mathcal{M} =$ linear span of $\{(\xi^i, \eta^i, \kappa^i) | i = 1, 2, \ldots, m\} \setminus \{0\}$.

Baignères, et al.: Distinguishing from the uniform distribution $\mathcal{D}_0$ has data complexity inversely proportional to

$$2^m \text{Var} = \sum_{\mu \in \mathcal{M}} \varepsilon_\mu^2 > \sum_{j=1}^{m} \varepsilon_j^2.$$

Linear Approximation

$f : GF(2)^n \rightarrow GF(2)^m$, $m < n$, vector-valued Boolean function.
$Y = f(X)$, linear approximation $Y(\eta) = X(\xi)$, with correlation
$c(Y(\eta), X(\xi)) = 1 - 2\Pr[Y(\eta) = X(\xi)]$.

- Parseval's Theorem: Linear approximations with non-zero correlation exist.
- Linearity of f

$$\mathcal{L}(f) = \max_{\xi,\eta \neq \emptyset} |c(Y(\eta), X(\xi))|.$$

- The goal is to find ξ and η which maximise linearity, first for relevant parts of the cipher, and then for the entire cipher.

Linear Resistance

N. & Knudsen 1992, N. 1994: Provable security against linear cryptanalysis

- Use functions f with $\mathcal{L}(f)$ as small as possible.

Example (N. 1993): $x \in GF(2^n)$,

$$f(x) = \begin{cases} x^{-1}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Standard (one-dimensional) linear attack data complexity $\approx 2^n$.

With $n = 8$ used as the only source of nonlinearity in the Advanced Encryption Standard (2002), $\mathcal{L}(f) = 2^{-3}$.

Multidimensional Linear Resistance

Resistance against Matsui's one-dimensional linear cryptanalysis may not guarantee resistance against multidimensional linear cryptanalysis.

Example. The vector-valued Boolean function $f^{(n)}$ in $GF(2)^n$ defined by the function $x \mapsto x^{-1}$ in $GF(2^n)$ has variance

$$\mathrm{Var}(f^{(n)}) = \begin{cases} 2^{-n}(1-2^{1-n}) \approx 2^{-n}, & \text{for } n \text{ even} \\ 2^{-n}, & \text{for } n \text{ odd .} \end{cases}$$

It follows that the data complexity of n-dimensional linear distinguishing attack applied to $f^{(n)}$ is constant (independent of n).

N.& Hermelin 2007: Vectorial bent functions are optimal against multidimensional linear approximation.

Conclusions

- Previous approaches to multiple and multidimensional linear approximations surveyed.
- Fixed key multidimensional linear approximation of block ciphers discussed.
- Resistance against multidimensional linear cryptanalysis is an open question.