Aalto University
School of Science
and Technology

Linear Cryptanalysis

Kaisa Nyberg

Department of Computer Science
Aalto University School of Science
kaisa.nyberg@aalto.fi

S3, Sackville, August 11, 2015

Outline

- Linear characteristics and correlations
- Matsui's algorithms
- Traditional statistical models using normal distributions under key equivalence hypotheses
- Linear Hull theorem
- Key variance and more realistic key hypotheses

Expected outcome

- This lecture provides you with the basic concepts for understanding Matsui's algorithms and more general linear cryptanalysis
- This lecture is targeted to give you the necessary (and hopefully also sufficient) prerequisites for being able to read recent
  - CRYPTO 2015 paper by Jialin Huang, et al., (and possibly also the one by Bing Sun, et al.)
  - FSE 2013 paper by Andrey Bogdanov and Elmar Tischhauser
- and goes beyond by presenting a new comprehensive model of Matsui's Algorithm 2 that handles key variance for both wrong keys and right keys.

Section: Linear characteristics and correlations

Symmetric-key encryption

$k \in \mathcal{K}$	the key
$x \in \mathcal{P}$	the plaintext
$y \in \mathcal{C}$	the ciphertext

Encryption method is a family $\{E_k\}$ of transformations $E_k : \mathcal{P} \to \mathcal{C}$, parametrized using the key k such that for each encryption transformation E_k there is a decryption transformation $D_k : \mathcal{C} \to \mathcal{P}$, such that $D_k(E_k(x))) = x$, for all $x \in \mathcal{P}$.

The spaces are too large to allow deterministic analysis when observing data from a cipher. Therefore we use information theoretic and statistical analysis and consider the key, plaintext and ciphertext as random variables.

Assumption: Plaintext and key are independent random variables that follow uniform distribution.

Block cipher

The data to be encrypted is split into blocks x_i, $i = 1, \ldots, N$, of fixed length n. A typical value of n is 128. $\mathcal{P} = \mathcal{C} = \mathbb{Z}_2^n$, $\mathcal{K} = \mathbb{Z}_2^\ell$.

For the purposes of linear cryptanalysis a block cipher is considered as a vectorial Boolean function

$$f : \mathbb{Z}_2^n \times \mathbb{Z}_2^\ell \to \mathbb{Z}_2^n \times \mathbb{Z}_2^\ell \times \mathbb{Z}_2^n, \;\; f(x, k) = (x, k, E_k(x))$$

Inner product of a mask $u = (u_1, \ldots, u_n) \in \mathbb{Z}_2^n$ and a data vector $x = (x_1, \ldots, x_n) \in \mathbb{Z}_2^n$ is computed as

$$u \cdot x = u_1 x_1 + u_2 x_2 + \ldots u_n x_n.$$

Linear approximation with mask triple $(u, v, w) \in$ of a block cipher is a relation

$$u \cdot x + v \cdot k + w \cdot E_k(x).$$

Correlation

- Correlation between two Boolean functions $f : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ and $g : \mathbb{Z}_2^n \mapsto \mathbb{Z}_2$ is defined as

  $$c(f,g) = 2^{-n}\left(\#\{x \in \mathbb{Z}_2^n \mid f(x) = g(x)\} - \#\{x \in \mathbb{Z}_2^n \mid f(x) \neq g(x)\}\right)$$

- Correlation $c(f,0)$ is called the correlation of $f(x)$ over x, and also denoted as $c_x(f(x))$. Then

  $$c_x(f(x)) = 2\Pr(f(x) = 0) - 1.$$

- Linear cryptanalysis makes use of Boolean functions derived from ciphers with large correlations

  $$|c_x(u \cdot x + v \cdot k + w \cdot E_k(x))|.$$

- Useful expression:

  $$c_x(f(x)) = 2^{-n}\sum_x (-1)^{f(x)},$$

  where the sum is taken over integers.

Correlations in iterated block ciphers

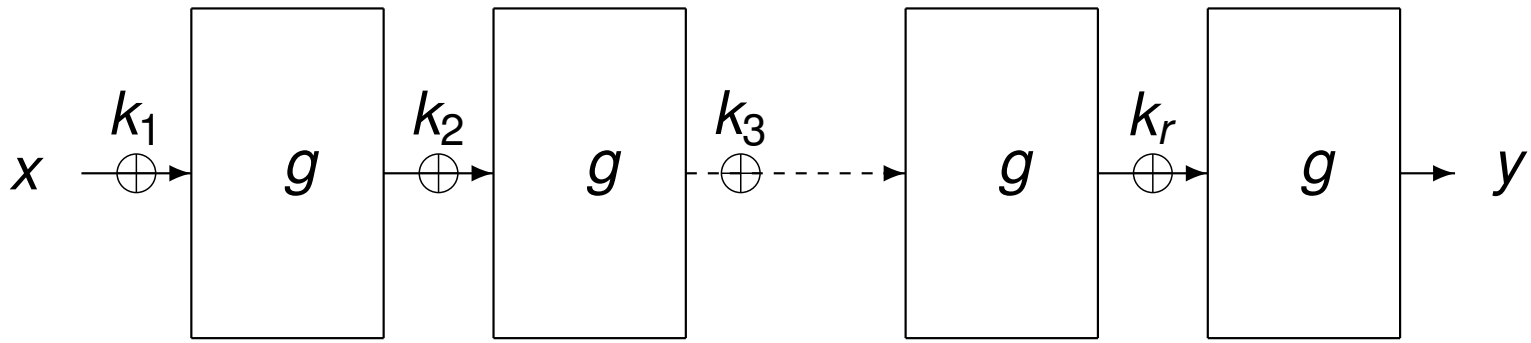
We focus on key alternating iterated block ciphers. Let $(k_1, k_2, \ldots, k_r)$ be the extended key with the round keys k_i derived from master key k. A key-alternating cipher E_k has following structure

$$E_k(x) = g(\ldots g(g(g(x + k_1) + k_2) \ldots) + k_r).$$

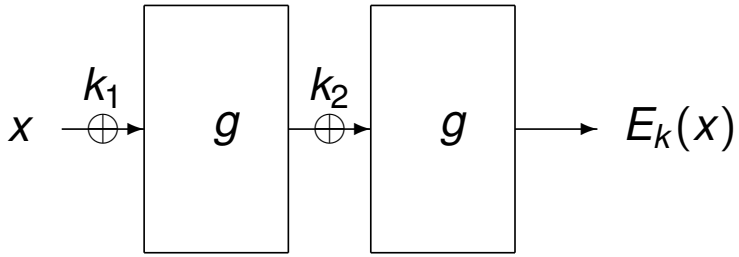
Then

$$c_x(u \cdot x + w \cdot E_k(x)) = \sum_{\tau} \prod_{i=1}^{r} (-1)^{\tau_i \cdot k_i} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)),$$

where $_\tau = (\tau_1 = u, \tau_2, \ldots, \tau_r, \tau_{r+1} = w)$. [JD94]



Proof in case $r = 2$



In this case $\tau = (u, \tau_2, w)$

$$c_x(u \cdot x + w \cdot E_k(x)) = 2^{-n} \sum_x (-1)^{u \cdot x + w \cdot E_k(x)}$$

$$= 2^{-n} \sum_x (-1)^{u \cdot x + w \cdot g(g(x+k_1)+k_2)}$$

$$= 2^{-2n} \sum_\tau \sum_x (-1)^{u \cdot x + \tau_2 \cdot g(x+k_1)} \sum_y (-1)^{\tau_2 \cdot y + w \cdot g(y+k_2)}$$

$$= 2^{-2n} \sum_\tau \sum_{z_1} (-1)^{u \cdot (z_1+k_1) + \tau_2 \cdot g(z_1)} \sum_{z_2} (-1)^{\tau_2 \cdot (z_2+k_2) + w \cdot g(z_2)}$$

$$= \sum_\tau (-1)^{u \cdot k_1 + \tau_2 \cdot k_2} c_{z_1}(u \cdot z_1 + \tau_2 \cdot g(z_1)) c_{z_2}(\tau_2 \cdot z_2 + w \cdot g(z_2)).$$

Linear characteristic

Similarly as in the previous proof, we set $z_1 = x + k_1$ and $z_i = g(z_{i-1}) + k_i$, $i = 2, \ldots, r$. Let $v = (v_1, \ldots, v_r, v_{r+1})$ be an $(r+1)$-tuple of masks such that $v_1 = u$ and $v_{r+1} = w$. Then

$$\bigoplus_{i=1}^{r} (v_i \cdot z_i + v_{i+1} \cdot g(z_i)) = u \cdot x + v_1 \cdot k_1 + \ldots + v_r \cdot k_r + w \cdot E_k(x).$$

The sequence $v = (v_1, \ldots, v_r, v_{r+1})$ is called a linear characteristic from u to w over the key-alternating cipher E_k.

We set $v \cdot k = v_1 \cdot k_1 + \ldots + v_r \cdot k_r$. Then the linear characteristic $v = (v_1, \ldots, v_r, v_{r+1})$ defines the linear approximation

$$u \cdot x + v \cdot k + w \cdot E_k(x).$$

Correlation of linear characteristic

Using

$$c_x(u \cdot x + w \cdot E_k(x)) = \sum_{\tau} (-1)^{\tau \cdot k} \prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)),$$

where $\tau = (u, \tau_2, \ldots, \tau_r, w)$, we obtain

$$\begin{aligned} c_x(u \cdot x + v \cdot k + w \cdot E_k(x)) &= (-1)^{v \cdot k} c_x(u \cdot x + w \cdot E_k(x)) \\ &= (-1)^{v \cdot k} \sum_{\tau} (-1)^{\tau \cdot k} \prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)) \\ &= \prod_{i=1}^{r} c_z(v_i \cdot z + v_{i+1} \cdot g(z)) + \sum_{\tau \neq v} (-1)^{\tau \cdot k} \prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)). \end{aligned}$$

Taking the average over $k = (k_1, \ldots, k_r)$, where all k_i take all possible values, will make the second term vanish if at least one $\tau_i \neq 0$. We can state the following theorem.

Average correlation of linear characteristic

Assumption. Round keys $k_1, \ldots, k_r$ take all possible values.

Theorem. Average correlation of a linear characteristic $v = (v_1, v_2, \ldots, v_r, v_{r+1})$ from u to w taken over extended keys $k = (k_1, \ldots, k_r)$ is

$$\begin{aligned} \tilde{c}(u, v, w) &= \mathrm{Avg}_k c_x(u \cdot x + v \cdot k + w \cdot E_k(x)) \\ &= \prod_{i=1}^{r} c_z(v_i \cdot z + v_{i+1} \cdot g(z)) \end{aligned}$$

- In the first practical cryptanalysis, the following estimate was used

$$c_x(u \cdot x + v \cdot k + w \cdot E_k(x)) = (-1)^{v \cdot k} c_x(u \cdot x + w \cdot E_k(x)) \approx \tilde{c}(u, v, w)$$

- $\tilde{c}(u, v, w)$ was computed using the Piling up lemma under the heuristic assumption of round-independence. We replaced this assumption by assuming independent round keys (i.e., long-key cipher [DR2007]).
- Is $\tilde{c}(u, v, w)$ a good estimate for any fixed key k?

Case of single dominant characteristic

Rewriting the previous equation gives

$$c_x(u \cdot x + w \cdot E_k(x)) \approx (-1)^{v \cdot k} \prod_{i=1}^{r} c_z(v_i \cdot z + v_{i+1} \cdot g(z)). \qquad (1)$$

The values $c_x(u \cdot x + w \cdot E_k(x))$ are estimated from cipher data. In reality,

$$c_x(u \cdot x + w \cdot E_k(x)) = \sum_{\tau} (-1)^{\tau \cdot k} \prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)).$$

A characteristic $v = (v_1, \ldots, v_{r+1})$ is called dominant characteristic from u to w, if

$$\tilde{c}(u, v, w) = \prod_{i=1}^{r} c_z(v_i \cdot z + v_{i+1} \cdot g(z)) \text{ is large, and}$$

$$\prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)) \approx 0, \text{ for } \tau \neq v.$$

Toy example

$E_k(x) = g(g(x) + k)$, where g is the AES S-box and k is eight bits. The maximum of $|c(u \cdot x + v \cdot g(x))|$ is 2^{-3}.Then for any 8-bit end masks u and w there exist many characteristics v with equally large correlations

$$|\tilde{c}(u, v, w)| = 2^{-6}$$

which is the maximum possible value 2^{-6}.
On the other hand, for any given (u, w) the true value $|c_x(u \cdot x + w \cdot E_k(x))|$ varies a lot with the key k.

Consider $(u, w) = (\mathtt{EA}, \mathtt{EA})$. Then we have $|c_x(u \cdot x + w \cdot E_k(x))| = 0$, for 21 keys k.
For these keys, linear cryptanalysis fails!

For the remaining 235 keys k we have $|c_x(u \cdot x + w \cdot E_k(x))| \geq 2^{-6}$.

There are no single dominant characteristics.

Correlations over SPN: S-box layer

$$\begin{aligned} x &= (x_1, x_2, \ldots, x_t) \\ g(x) &= (S_1(x_1), S_2(x_2), \ldots, S_t(x_t)) \end{aligned}$$

Assumption: Inputs x_j to different S-boxes are independent.

$$\begin{aligned} u &= (u_1, u_2, \ldots, u_t) \\ v &= (v_1, v_2, \ldots, v_t) \\ c_x(u \cdot x + v \cdot g(x)) &= \prod_{j=1}^{t} c_{x_j}(u_j \cdot x_j + v_j \cdot g(x_j)), \end{aligned}$$

by the Piling up lemma.
To maximize the correlation one usually takes almost all u_j and v_j equal to zero, since then $c_{x_j}(u_j \cdot x_j + v_j \cdot g(x_j)) = 1$.

Linear characteristics for SPN: linear layer

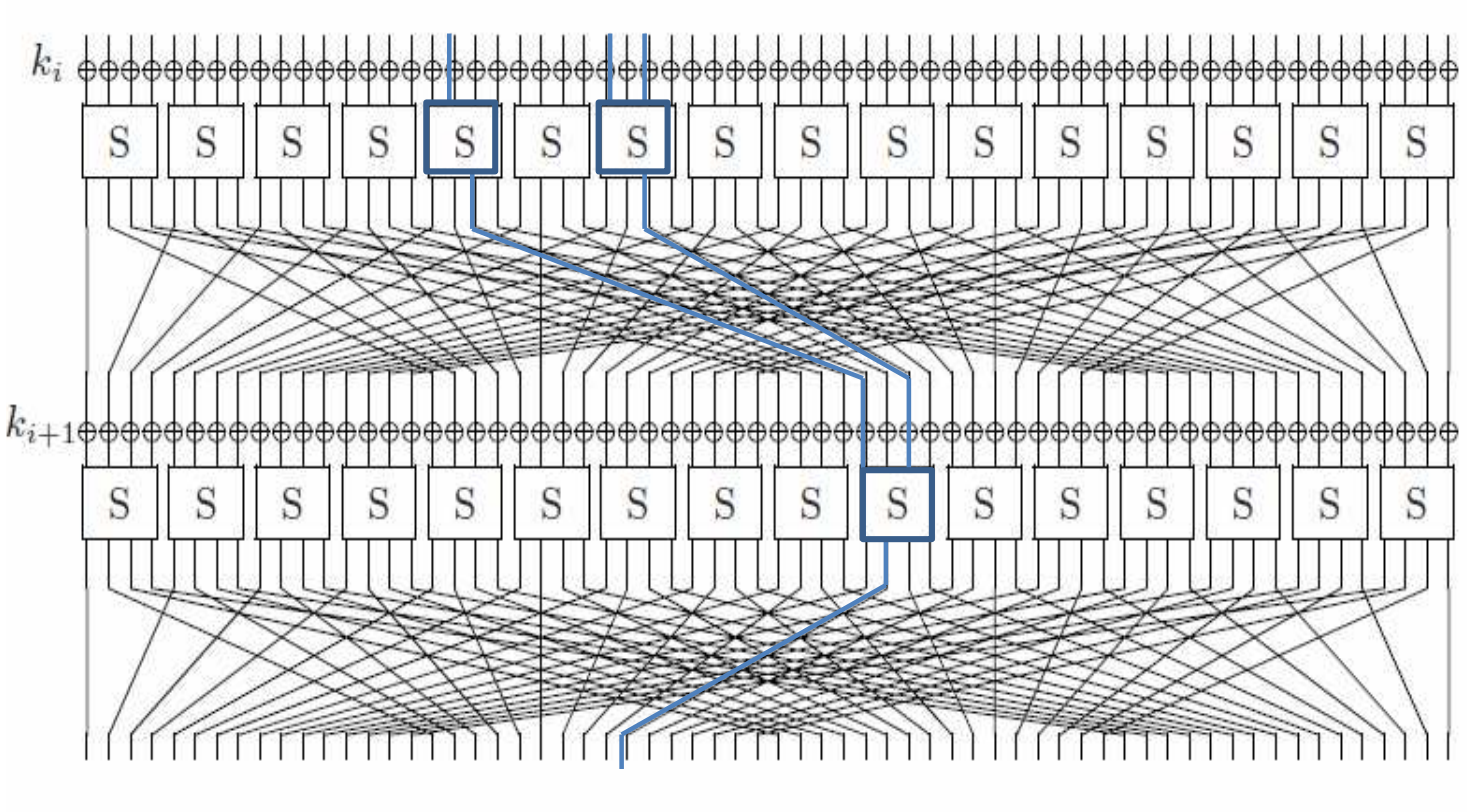
$$g(x) = Mx$$

$$\begin{aligned} c_x(u \cdot x + v \cdot Mx)) &= c_x(u \cdot x + M^t v \cdot x)) \\ &= \begin{cases} 1 & \text{if } u = M^t v \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

This uniquely determines the masks over the linear layer.

Text-books have nice concrete examples of how to construct linear characteristics over SPNs, see [Stinson] or [Knudsen-Robshaw].

SPN characteristic example



Section: Matsui's algorithms

Empirical correlation

Given a sample D of size N of plaintext-ciphertext pairs $(x, y = E_K(x))$ obtained from cipher with unknown key K the empirical correlation (or observed correlation) $\hat{c}(D, K)$ of linear approximation $u \cdot x + w \cdot y$) is computed as

$$\hat{c}(D, K) = \frac{1}{N} \left(\#\{(x, y) \mid u \cdot x + w \cdot y = 0\} - \#\{(x, y) \mid u \cdot x + w \cdot y \neq 0\} \right)$$

Algorithm 1

Matsui's Algorithm 1 is a statistical cryptanalysis method for finding one bit of the key with the following steps

1. Select a linear characteristic v such that the average correlation

$$\tilde{c} = \tilde{c}(u, v, w)$$

   deviates from 0 as much as possible.

2. Sample plaintext-ciphertext pairs $(x, E_K(x))$ for a fixed (unknown) key K and determine the empirical correlation $\hat{c}(D, K)$ of the linear approximation

$$u \cdot x + w \cdot E_K(x)$$

3. If $\tilde{c}$ and $\hat{c}$ are of the same sign, output $v \cdot K = 0$. Else output $v \cdot K = 1$.

Here estimate (1) is used: $c_x(u \cdot x + w \cdot E_K(x)) \approx (-1)^{v \cdot K} \tilde{c}(u, v, w)$.

Algorithm 2

Matsui's Algorithm 2 is a statistical cryptanalysis method for finding a part of the last round key for block ciphers where a relation (last round trick)

$$E_{k',k_r}(x) = G_{k_r}(E'_{k'}(x)),$$

where k_r is relatively short, can be constructed from the encryption transformation.

1. Select a linear characteristic v such that the average correlation $\tilde{c} = \tilde{c}(u, v, w)$ of the linear approximation

$$u \cdot x + v \cdot k' + w \cdot E'_{k'}(x)$$

deviates from 0 as much as possible.

Algorithm 2, cont'd

2. Take a sample D of plaintext-ciphertext pairs (x, E_{k',k_r}) of size N. For each last round key candidate K, compute pairs $(x, y = G_K^{-1}(E_{k',k_r}(x))$ and determine the empirical correlation $\hat{c}(D, K)$ of the linear approximation

$$u \cdot x + w \cdot y = 0$$

3. Output a set of values K that achieve top largest values $|\hat{c}(D, K)|$.
4. Additionally, one can determine the value $v \cdot k'$ as in Algorithm 1.

Key assumptions of Matsui's Algorithm 2

The statistical model of Algorithm 2 relies on two assumptions

- **Hypothesis of right-key equivalence:** The observed correlation $\hat{c}(D, K_R)$ computed from a data sample of size N obtained by partial decryption using the right key candidate $K = K_R$ follows, for all K_R, normal distribution with parameters

$$\begin{aligned} |\mathrm{Exp}_D(\hat{c}(D, K_R))| &= |\tilde{c}|, \text{ i.e. is the same for all } K_R \\ \mathrm{Var}_D(\hat{c}(D, K_R)) &= \mathrm{Exp}_D(\hat{c}(D, K_R) - \mathrm{Exp}_D(\hat{c}(D, K_R)))^2 = \frac{1}{N}. \end{aligned}$$

- **Wrong-key hypothesis:** The observed correlation $\hat{c}(D, K_W)$ computed from a data sample of size N obtained by partial decryption using a wrong key candidate $K = K_W$ follows, for all K_W, normal distribution with parameters

$$\begin{aligned} \mathrm{Exp}_D(\hat{c}(D, K_W)) &= 0 \\ \mathrm{Var}_D(\hat{c}(D, K_W)) &= \frac{1}{N}. \end{aligned}$$

Variance of observed correlation

We explain the wrong key variance. Let N be the size of the sample and N_0 be the number of observed pairs (x, y) computed with the wrong key that satisfy $u \cdot x + w \cdot y = 0$.

N_0 is binomially distributed with expected value $N/2$ and variance $N/4$. For large N we can approximate the binomial distribution with the normal one. From

$$\hat{c} = \frac{2N_0 - 1}{N}$$

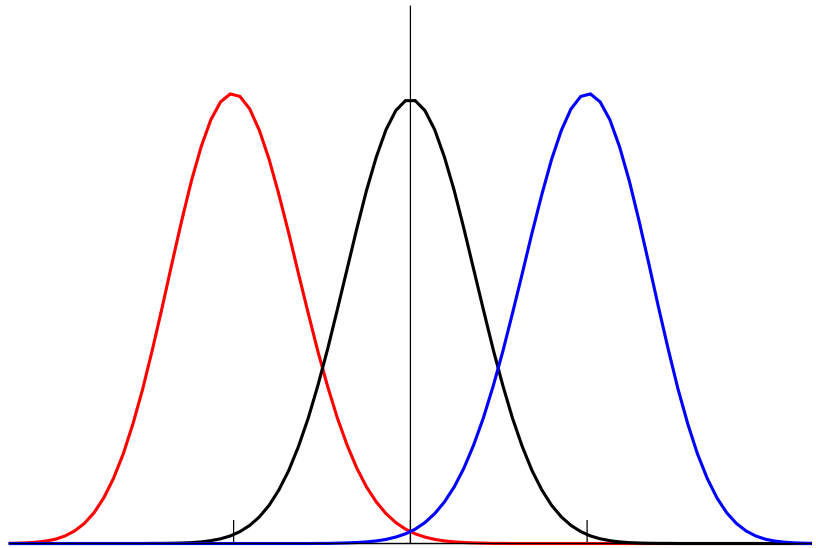
it follows that

$$\mathrm{Exp}_D\left(\hat{c}(D, K_W)\right) = 0 \text{ and } \mathrm{Var}_D\left(\hat{c}(D, K_W)\right) = \frac{1}{N}.$$

The data variance of the correlation computed for the correct key is approximately the same.

As N grows the variance decreases.

Normal distributions of observed correlations in Algorithm 2



Legend: black = wrong key, red = right key with $\mathrm{Exp}_D\left(\hat{c}(D, K_R)\right) < 0$, blue = right key with $\mathrm{Exp}_D\left(\hat{c}(D, K_R)\right) > 0$

Section: Success probability and data complexity

Statistical tests

- Linear cryptanalysis makes use of a statistical hypothesis test.
- Algorithm 1 makes a decision between

$$\begin{aligned} \mathrm{H}_0 &: \quad v \cdot k = 0 \\ \mathrm{H}_1 &: \quad v \cdot k = 1 \end{aligned}$$

- Algorithm 2 makes a decision between

$$\begin{aligned} \mathrm{H}_0 &: \quad K = K_R, \text{ that is, } G_K^{-1}(E_{k',K_R}(x)) = E'_{k'}(x), \text{ for all } x \\ \mathrm{H}_1 &: \quad K = K_W, \text{ that is, data pairs } (x, G_K^{-1}(E_{k',K_R}(x)) \\ &\quad \text{are not from the cipher} \end{aligned}$$

The setting

- Two normal deviates T_W and T_R such that

$$T_W \sim \mathcal{N}(\mu_W, \sigma_W^2) \text{ and } T_R \sim \mathcal{N}(\mu_R, \sigma_R^2).$$

  (w.l.g. assume $\mu_W < \mu_R$)

- Value of T computed from a sample drawn from either the distribution of T_W or the one of T_R. The task is to decide which one of the two.

Success probability

- Threshold value Θ:
  - $T \leq \Theta \Rightarrow T$ is drawn from the distribution of T_W,
  - $T \geq \Theta \Rightarrow T$ is drawn from the distribution of T_R.
- We set bounds to the error probabilities

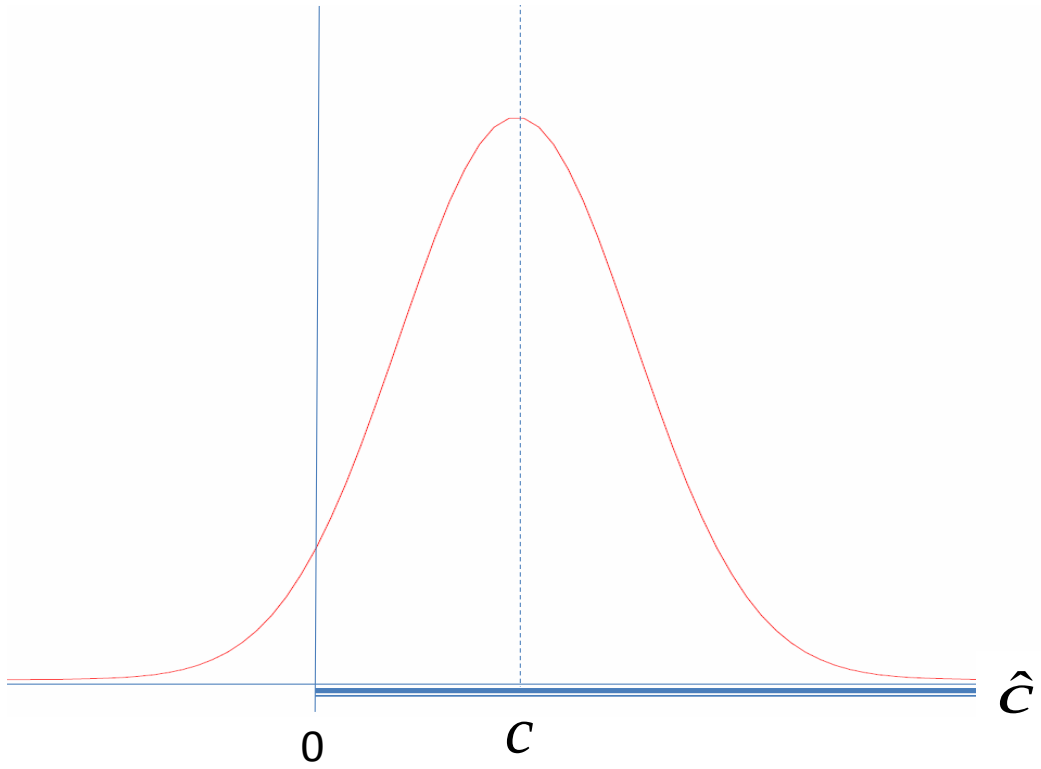
$$\Pr(T_W \mid T > \Theta) \leq \alpha_0 \text{ and } \Pr(T_R \mid T \leq \Theta) \leq \alpha_1,$$

which are satisfied if

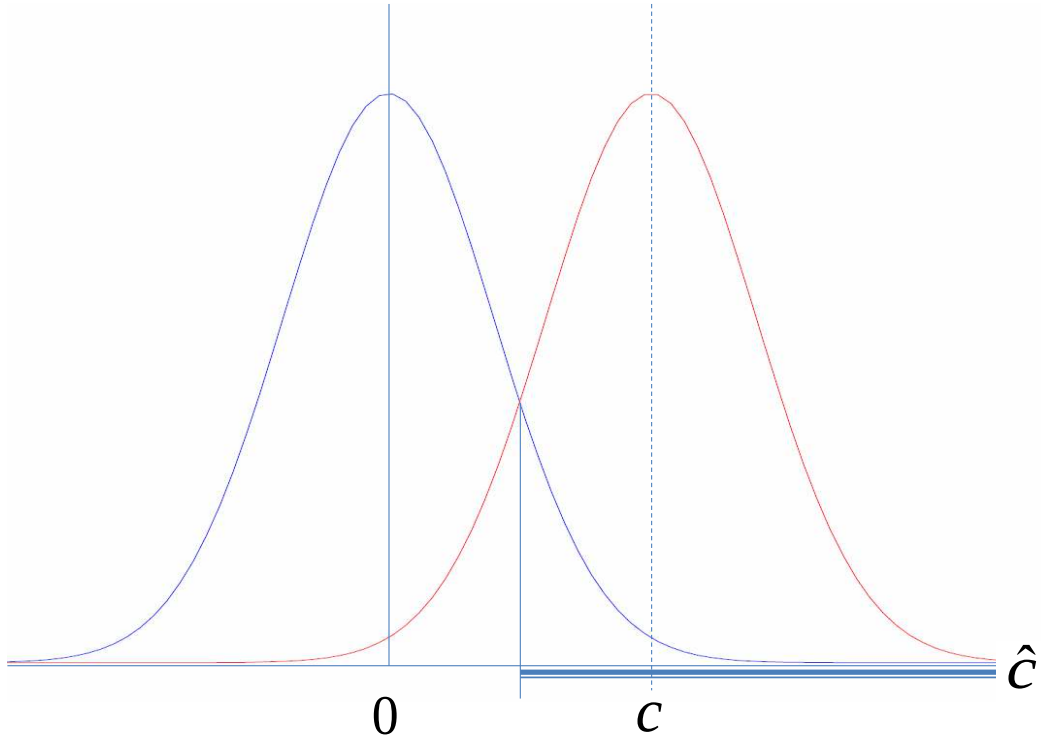
$$\mu_W + \sigma_W \zeta_0 \leq \Theta \leq \mu_W - \sigma_R \zeta_1,$$

where $\Phi(\zeta_i) = 1 - \alpha_i$ for $i = 1, 2$, and Φ is the cumulative distribution function of the standard normal distribution.

Example success area in Algorithm 1



Example success area in Algorithm 2



Success probability in Algorithm 2

- We denote

$$\alpha_0 = 2^{-a-1} \text{ and } \alpha_1 = 1 - P_S,$$

  where a is often called the advantage and P_S is the success probability that the right key is accepted. The value a indicates how many bits of the right key are correctly determined.
- Then a set of $2^{\ell-a}$ keys remains to be searched.
- Question: Why $\alpha_0 = 2^{-a-1}$ and not 2^{-a}?

Success probability in Algorithm 2

- Plugging in the parameters of distributions

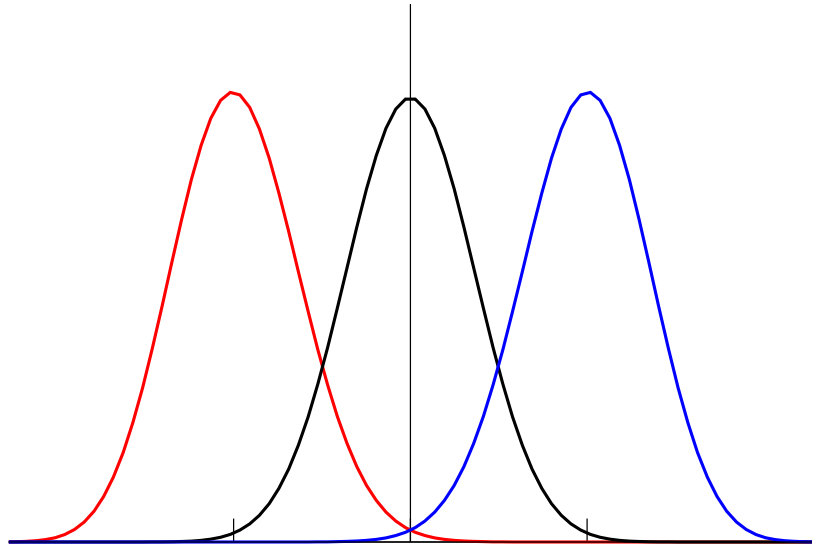
$$\frac{1}{\sqrt{N}}\Phi^{-1}(1-2^{-a-1}) \leq \Theta \leq |\tilde{c}| - \frac{1}{\sqrt{N}}\Phi^{-1}(P_S)$$

- Such a value Θ exists if

$$\begin{aligned} P_S &\leq \Phi(\sqrt{N}|\tilde{c}| - \Phi^{-1}(1-2^{-a-1}), \text{ or equivalently} \\ N &\geq \frac{\left(\Phi^{-1}(1-2^{-a-1}) + \Phi^{-1}(P_S)\right)^2}{|\tilde{c}|^2} \end{aligned}$$

This lower bound of N is an estimate of the data complexity required to get a bits of the right key with success probability P_S.

Why $\alpha_0 = 1 - 2^{-a-1}$?



Answer: Samples computed with wrong keys will be accepted on both sides. To have the total probability less than 2^{-a} both sides must have probability less than 2^{-a-1}.

Question: Why the success probability is not halved?

Answer: There is only one correct key. It can fail the test only on one side of the distribution.

Statistical model of Algorithm 2

- Φ the cumulative distribution function of the standard normal distribution
- P_S success probability, $\varphi_{P_S} = \Phi^{-1}(P_S)$
- 2^{-a} is the proportion of accepted wrong keys
- a is the advantage of the attack, $\varphi_a = \Phi^{-1}(1 - 2^{-a-1})$
- μ_R and σ_R^2 are the mean and variance of the normal deviate $\hat{c}(D, K_R)$ for the right key, and
- μ_W and σ_W^2 are the mean and variance of the normal deviate $\hat{c}(D, K_W)$ for the wrong key.

Then the success probability can be determined by

$$P_S \approx \Phi\left(\frac{\mu_R - \mu_W - \sigma_W \varphi_a}{\sigma_R}\right).$$

Summary of Algorithm 2

- Fix the advantage a, where $a \leq t$ and t is the length in bits of the last round key.
- Fix the success probability P_S
- Compute the sample size N from

$$N = \frac{\left(\Phi^{-1}(1 - 2^{-a-1}) + \Phi^{-1}(P_S)\right)^2}{|\tilde{c}|^2}$$

- Use a sample D of N known plaintext-ciphertext pairs
- Compute the observed correlations for all 2^t last round key candidates.
- Take those 2^{t-a} of the round key candidates that have the largest empirical correlation (in absolute value).
- Complement them to full ℓ-bit keys by appending the remaining $\ell - t$ bits to them (assuming the key-schedule allows this).
- Search the set of $2^{\ell-a}$ cipher key candidates exhaustively.

Section: Linear Hull theorem

Fixed-key correlation of a linear characteristic

Recall

$$c_x(u \cdot x + w \cdot E_k(x)) = \sum_{\tau} \prod_{i=1}^{r} (-1)^{\tau_i \cdot k_i} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)),$$

where $\tau =_(u, \tau_2, \ldots, \tau_r, w)$, and that for any v,

$$c_x(u \cdot x + v \cdot k + w \cdot E_k(x)) = (-1)^{v \cdot k} c_x(u \cdot x + w \cdot E_k(x))$$
$$= \prod_{i=1}^{r} c_z(v_i \cdot z + v_{i+1} \cdot g(z)) + \sum_{\tau \neq v} (-1)^{\tau \cdot k} \prod_{i=1}^{r} c_z(\tau_i \cdot z + \tau_{i+1} \cdot g(z)).$$

We also computed the average of this correlation over the keys. The Right Key Hypothesis states that all right keys behave as the average.

This is clearly not the case. Next we compute the variance of $c_x(u \cdot x + v \cdot k + w \cdot E_k(x))$ as the key k varies.

The Fundamental Theorem

By Jensen's inequality

$$\mathrm{Avg}_k c_x(u \cdot x + v \cdot k + w \cdot E_k(x))^2 \geq \tilde{c}(u, v, w)^2,$$

for all v, and in general the strict inequality holds. More accurately, the following theorem holds

The Linear Hull Theorem [KN94, KN01, DR2007] If the round keys of a block cipher E_k take on all values (aka E_k is a long-key cipher), then

$$\mathrm{Avg}_k c_x(u \cdot x + w \cdot E_k(x))^2 = \sum_{\tau} \tilde{c}(u, \tau, w)^2.$$

We denote

$$\mathrm{ELP}(u, w) = \mathrm{Avg}_k(c_x(u \cdot x + w \cdot E_k(x)))^2$$

and call it the expected linear potential of (u, w).

Toy example cont'd

Consider the previous example. We saw that in terms of single characteristics, all (u, w) are about equally good, but there are no dominant characteristics.

Also in terms of linear hulls, all (u, w) have about equally large ELP:

$$\mathrm{ELP}(\mathtt{33}, \mathtt{D5}) = 2^{-10.40} \leq \mathrm{ELP}(u, w) \leq 2^{-9.65} = \mathrm{ELP}(\mathtt{EA}, \mathtt{EA})$$

$|c(u \cdot x + w \cdot E_k(x))|^2 \geq \mathrm{ELP}(\mathtt{EA}, \mathtt{EA})$, for 76 keys k.

The weakest of (u, w) is $(\mathtt{33}, \mathtt{D5})$. For this mask pair $|c_x(u \cdot x + w \cdot E_k(x))| = 0$, for 33 keys k.

For the remaining 223 keys we have $|c_x(u \cdot x + w \cdot E_k(x))| \geq 2^{-6}$.

$|c(u \cdot x + w \cdot E_k(x))|^2 \geq \mathrm{ELP}(\mathtt{33}, \mathtt{D5})$, for 80 keys k.

Computing an estimate of $\mathrm{ELP}(u, w)$

$$
\begin{aligned}
\mathrm{ELP}(u,w) &= \mathrm{Avg}_k c_x(u\cdot x + w\cdot E_k(x))^2 = \sum_{\tau_2,\ldots,\tau_r}\prod_{i=1}^{r} c_z(\tau_i\cdot z + \tau_{i+1}\cdot g(z))^2 \\
&= \sum_{\tau_r} c_z(\tau_r\cdot z + w\cdot g(z))^2 \sum_{\tau_{r-1}} c_z(\tau_{r-1}\cdot z + \tau_r\cdot g(z))^2 \\
&\quad \cdots\cdots \quad \sum_{\tau_3} c_z(\tau_3\cdot z + \tau_4\cdot g(z))^2 \\
&\quad \sum_{\tau_2} c_z(\tau_2\cdot z + \tau_3\cdot g(z))^2 c_z(u\cdot z + \tau_2\cdot g(z))^2
\end{aligned}
$$

- This expression gives an iterative algorithm: start from the bottom line to compute for each τ_3 the value on the last line.
- Can be made feasible by restricting to τ with low Hamming weight and keeping only the largest values from each iteration.
- Restrictions on τ will lead to a lower bound of $\mathrm{ELP}(u, w)$, which is still much larger than any $\tilde{c}(u, v, w)^2$.

Section: Key variance and more realistic key hypotheses

Key-variance of observed correlation

Wrong key case
see also [BT2013]

- The wrong-key hypothesis states that $\mathrm{Exp}_D\left(\hat{c}(D, K_W)\right)$ is equal for all keys K_W. This is not true in practice.
- Denote $c(K_W) = \mathrm{Exp}_D\left(\hat{c}(D, K_W)\right)$.
- Then $\mathrm{Exp}_{K_W}\left(c(K_W)\right) = 0$
- To compute the variance, we use

$$\mathrm{Var}_{K_W}(c(K_W)) = \mathrm{Exp}_{K_W}\left(c(K_W)^2\right) - \left(\mathrm{Exp}_{K_W}\left(c(K_W)\right)\right)^2.$$

- By [DR2007], Corollary 7, $\mathrm{Exp}_{K_W}\left(c(K_W)^2\right) = 2^{-n}$.
- Then $\mathrm{Var}_{K_W}\left(c(K_W)\right) = 2^{-n}$.

Key-variance of observed correlation

Right key case (New!)

The Right-key equivalence hypothesis states that $|\mathrm{Exp}_D(\hat{c}(D, K_R)|)$ is equal for all keys K_R. This is not realistic.

- Denote $\mathrm{Exp}_D(\hat{c}(D, K_R)) = \tilde{c}(K_R)$. If $\tilde{c}(K_R) > 0$ set $K_R \in \mathcal{K}^+$. Else $K_R \in \mathcal{K}^-$
- Hypothesis of Algorithm 2 (single dominant characteristic):

$$\mathrm{Exp}_{K_R \in \mathcal{K}^-}(\tilde{c}(K_R)) \approx -\mathrm{Exp}_{K_R \in \mathcal{K}^+}(\tilde{c}(K_R))$$

- Denote $|\mathrm{Exp}_{K_R \in \mathcal{K}^+}(\tilde{c}(K_R))| = c$
- To compute the variance of $\tilde{c}(K_R)$, for $K_R \in \mathcal{K}^+$ (similarly for $K_R \in \mathcal{K}^-$) we apply again the rule

$$\mathrm{Var}_X(F(X)) = \mathrm{Exp}_X\left(F(X)^2\right) - \left(\mathrm{Exp}_X(F(X))\right)^2.$$

- Then we get $\mathrm{Var}_{K_R \in \mathcal{K}^+}(\tilde{c}(K_R)) = \mathrm{ELP} - c^2$.

Total variance of observed correlation

Wrong key case

- To compute the total variance over data and wrong keys, we recall the rule

$$\mathrm{Var}_{X,Y}\left(F(X,Y)\right) = \mathrm{Exp}_X\left(\mathrm{Var}_Y\left(F(X,Y)\right)\right) + \mathrm{Var}_X\left(\mathrm{Exp}_Y\left(F(X,Y)\right)\right)$$

- ... and get

$$\mathrm{Var}_{D,K_W}\left(\hat{c}(D,K_W)\right) = \frac{1}{N} + 2^{-n}.$$

Total variance of observed correlation

Right key case

- To compute the total variance of $|\hat{c}(D, K_R)|$ over data and right keys, we use the same rule as above
- ... and get

$$\begin{aligned} \mathrm{Exp}_{K_R}\left(\mathrm{Var}_D\left(\hat{c}(D, K_R)\right)\right) &= \frac{1}{N} \\ \mathrm{Var}_{K_R}\left(\mathrm{Exp}_D\left(\hat{c}(D, K_R)\right)\right) &= \mathrm{Var}_{K_R}\left(\tilde{c}(K_R)\right) = \mathrm{ELP} - c^2. \end{aligned}$$

- $\mathrm{Var}_{D,K_R\in\mathcal{K}^+}\left(\hat{c}(D, K_R)\right) = \mathrm{Var}_{D,K_R\in\mathcal{K}^-}\left(\hat{c}(D, K_R)\right) = \frac{1}{N} + \mathrm{ELP} - c^2$

How large is the right key variance

Matsui's Algorithm uses one dominant characteristic (u, v, w). By the Linear Hull theorem

$$\mathrm{ELP} - \tilde{c}(u, v, w)^2 = \sum_{\tau \neq v} \tilde{c}(u, \tau, w)^2,$$

and

$$|\mathrm{Exp}_{D,K_R} \hat{c}(D, K_R)| = c = |\tilde{c}(u, v, w)|.$$

In the ideal case, the sum on the right side of the first equation is the variance of pure noise, which has mean equal to zero and variance equal to 2^{-n}. We obtain

$$\mathrm{ELP} - c^2 \geq 2^{-n}.$$

Note that the classical hypothesis of right key equivalence assumes $\mathrm{ELP} - c^2 = 0$.

Recalling the statistical model of Algorithm 2

- Φ the cumulative distribution function of the standard normal distribution
- P_S success probability, $\varphi_{P_S} = \Phi^{-1}(P_S)$
- 2^{-a} is the proportion of accepted wrong keys
- a is the advantage of the attack, $\varphi_a = \Phi^{-1}(1 - 2^{-a-1})$
- μ_R and σ_R^2 are the mean and variance of the normal deviate $\hat{c}(D, K_R)$ for the right key, and
- μ_W and σ_W^2 are the mean and variance of the normal deviate $\hat{c}(D, K_W)$ for the wrong key.

Then the success probability can be determined by

$$P_S \approx \Phi\left(\frac{\mu_R - \mu_W - \sigma_W \varphi_a}{\sigma_R}\right).$$

Success probability and data complexity

Now we plug in the new parameters of Matsui's Algorithm 2 to get

$$P_S \approx \Phi\left(\frac{c\sqrt{N} - \sqrt{1 + N2^{-n}}\varphi_a}{\sqrt{N(\mathrm{ELP} - \mathrm{c}^2) + 1}}\right).$$

If $\mathrm{ELP} = c^2$ (key-equivalence) then this is identical to the result in [BT2013] Eq.(6).

In reality, we have $\mathrm{ELP} - c^2 > 2^{-n}$, and we derive the data complexity estimate as

$$N \geq \frac{(\varphi_a + \varphi_{P_S})^2}{c^2 - (\mathrm{ELP} - c^2)(\varphi_a + \varphi_{P_S})^2 + \varphi_a^2(\mathrm{ELP} - c^2 - 2^{-n})}$$

Extensions and variations

- Zero-correlation: $c = \tilde{c}(K_R) = 0$ for all keys
- Is it possible to handle a case with two dominant characteristics?

Example [AES book]
linear approximation is composed of more than one dominant characteristics

$$c(u \cdot x + w \cdot E_k(x)) = (-1)^{\gamma \cdot k} c_\gamma + (-1)^{\lambda \cdot k} c_\lambda$$

where c_γ and c_λ are the correlations of the characteristics
Then Avg $(c(u \cdot x + \gamma \cdot k + w \cdot \mathcal{E}_k(x)) = c_\gamma$.
But this gives a usable estimate only for a half of the keys.
Those are the keys for which $\lambda \cdot k = 0$ and then

$$c(u \cdot x + \gamma \cdot k + w \cdot E_k(x)) = c_\gamma + c_\lambda$$

For the remaining keys we have

$$c(u \cdot x + \gamma \cdot K + w \cdot \mathcal{E}_K(x)) = c_\gamma - c_\lambda \approx 0.$$

The case of about equally small characteristics

To resist linear cryptanalysis, [DR2007] advices designers to take care that no single dominant characteristic exists.
It will make linear cryptanalysis harder, but not impossible. We have for the right key

$$\mathrm{Exp}_D(\hat{c}(D, K_R)) = c_x(u \cdot x + w \cdot E_k(x)) \sim \mathcal{N}(0, \mathrm{ELP})$$

from where we get Theorem 22 of [DR2007]
Theorem. If the number of characteristics with non-zero correlation is large and all characteristic correlations are small compared to ELP, then

$$\frac{c_x(u \cdot x + w \cdot E_k(x))}{ELP} \sim \chi^2(1)$$

The case of about equally small characteristics, cont'd

For the wrong keys

$$\mathrm{Exp}_D(\hat{c}(D, K_W)) \sim \mathcal{N}(0, 2^{-n})$$

as before, and

$$2^n \mathrm{Exp}_D(\hat{c}(D, K_W))^2 \sim \chi^2(1)$$

If ELP $>> 2^{-n}$ the distributions of the observed correlations for the wrong keys and right keys can be distinguished.

Multiple/Multidimensional linear cryptanalysis

- So far we fixed (u, v).
- By collecting a number, say m, strong linear approximations with different mask pairs (u, w) the sum of their observed squared correlations is a multiple of a $\chi^2(m)$ distributed random variable
- ... provided that the correlations are independent random variables.
- To overcome this problem, it is possible to consider distributions of data extracted from plaintext-ciphertext pairs with an expected distribution which is far from uniform.
- Such non-uniform distributions can be found by with the help of linear approximations with high correlations.
- They may also be found with the help of truncated differentials with high probabilities.

Summary

- Linear cryptanalysis using one dominant characteristic
- First the statistical model of observed correlation was presented for a fixed key
- Traditional key-equivalence hypotheses were stated
- As they are known not to hold in practice for all keys,
- we studied the variance in the correlation due to the key for wrong keys and also for right keys.
- New estimates, with improved theoretical justification, of success probability and data complexity achieved.
- Among modern ciphers, cases with single dominant linear characteristics are very rare
- The statistical model for uniformly small characteristics correlations briefly described
- Enhancements by using more linear approximations was discussed.

Cited papers and books

[KN94] K. Nyberg: Linear approximation of block ciphers. In Advances in Cryptology - EUROCRYPT'94, volume 950 of Lecture Notes in Computer Science. Springer-Verlag, 1995.

[JD94] J. Daemen: Correlation Matrices. In Fast Software Encryption, FSE 2, volume 1008 of Lecture Notes in Computer Science. Springer-Verlag, 1995.

[KN01] K. Nyberg: Correlation theorems in cryptanalysis. Discrete Applied Mathematics, 111:177-188, 2001.

[AES book] J. Daemen and V. Rijmen: The Design of Rijndael: AES - The Advanced Encryption Standard. Springer-Verlag, 2002.

[Stinson] D. R. Stinson: Cryptography: Theory and Practice, 3rd ed.. CRC Press, 2005.

[Knudsen-Robshaw] L. R. Knudsen and M. Robshaw: The Block Cipher Companion. Springer, 2011.

[DR2007] Joan Daemen and Vincent Rijmen: Probability distributions of correlation and differentials in block ciphers. *J. Mathematical Cryptology*, 1(3):221–242, 2007.

[BT2013] Andrey Bogdanov and Elmar Tischhauser: On the wrong key randomisation and key equivalence hypotheses in Matsui's algorithm 2. FSE 2013, LNCS 8424, pages 19–38.