FSE 2020

Multiple Linear Cryptanalysis Using Linear Statistics

Jung-Keun Lee and Woo-Hwan Kim

ETRI

Our contribution

- improved and extended approach of multiple linear cryptanalysis[BCQ04]
  (exploit dominant statistically independent linear trails)
  - Algorithm 1 and Algorithm 2 style attacks
    - threshold based, rank based, combined
    - provide formulas for success probability and advantage
      in terms of data size, correlations of the trails, and threshold parameter
    - under some hypotheses on statistical independence of wrong key & right key statistics
- application to full DES, exploiting 4 linear trails
  - get attacks with complexity better than or comparable with existing linear attacks on DES
  - provide strong experimental verification

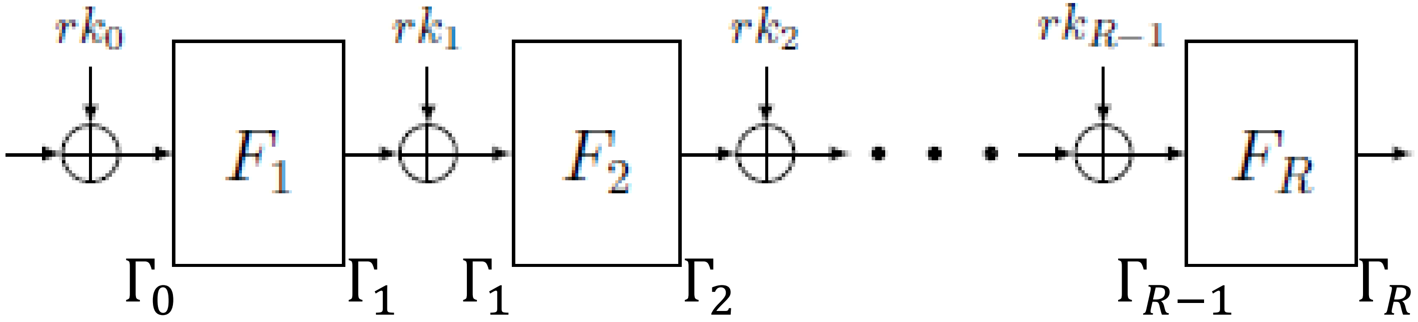
Organization

- Introduction and Preliminaries
- Our multiple linear attacks
- Application to DES
- Generalization
- Conclusion

Linear Trails and Linear Hulls

- key-alternating iterative block cipher

long key cipher $\tilde{E}$



- linear trail $\Gamma = [\Gamma_0, \ldots, \Gamma_R]$: sequence of linear masks
- linear hull $\mathcal{H}(\gamma, \gamma')$: the set of linear trails with the initial mask γ and final mask γ'

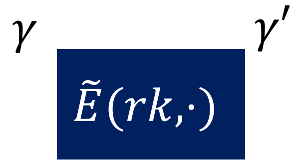
Linear Correlations

- $\varepsilon(\gamma,\gamma';F) \coloneqq \frac{1}{2^l}\sum_x(-1)^{\langle\gamma,x\rangle\oplus\langle\gamma',F(x)\rangle}$

  linear correlation of $F\colon \mathbb{F}_2^l \to \mathbb{F}_2^m$ w.r.t. pair of masks (γ,γ')

- $\varepsilon\left(\gamma,\gamma';\tilde{E},rk\right) \coloneqq \varepsilon\left(\gamma,\gamma';\tilde{E}(rk,\cdot)\right)$

  linear correlation of a linear hull for a given long key rk



- $C\left(\Gamma;\tilde{E}\right) = \prod_{i=0}^{R-1}\varepsilon(\Gamma_i,\Gamma_{i+1};F_{i+1})$

  (key-independent) linear correlation of a trail

- $\hat{\varepsilon}\left(\gamma,\gamma';\tilde{E},rk,D\right) \coloneqq \frac{1}{|D|}\sum_{(P,C)\in D}(-1)^{\langle\gamma,P\rangle\oplus\langle\gamma',C\rangle}$

  undersampled correlation

  D: data (consisting of plaintext-ciphertext pairs)

Linear Correlations

parity bit determined by Λ and rk

$$\varepsilon(\gamma,\gamma';\tilde{E},rk) = \sum_{\Lambda\in\mathcal{H}(\gamma,\gamma')}(-1)^{\boxed{\bigoplus_{i=0}^{R-1}\langle\Lambda_i,rk_i\rangle}}C(\Lambda;\tilde{E})$$

- Γ: a dominant trail

$\Rightarrow$
- $\varepsilon(\gamma,\gamma';\ rk) \approx (-1)^{\bigoplus_{i=0}^{R-1}\langle\Gamma_i,rk_i\rangle}C(\Gamma)$, or
- $(-1)^{\bigoplus_{i=0}^{R-1}\langle\Gamma_i,rk_i\rangle}\varepsilon(\gamma,\gamma';rk) \approx C(\Gamma)$

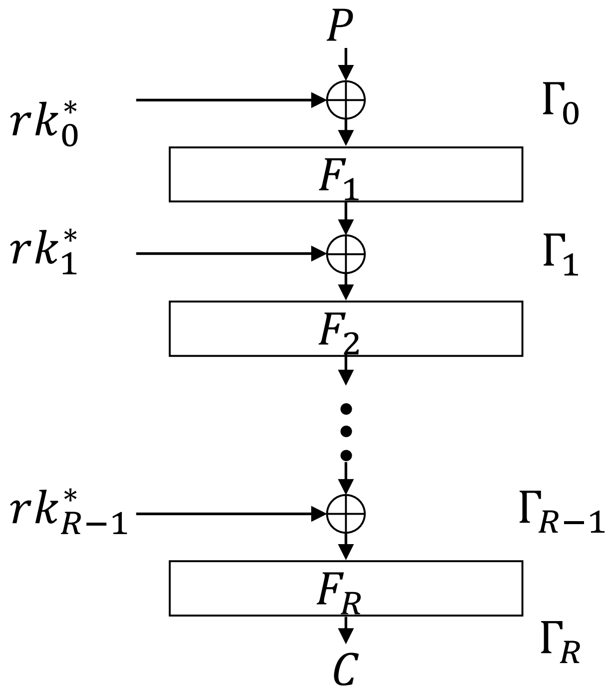
regardless of rk

Unless mentioned otherwise, we assume:-
- Γ, Γ^j: dominant, fixed
- $N = |D| \ll 2^n$, n: block size
- $|C(\Gamma)|, |C(\Gamma^j)| \gg 2^{-n/2}$
- K^* and rk^*(correct key, long key): fixed

Algorithm 1

- Use a single dominant trail $\Gamma = [\Gamma_0, \dots, \Gamma_R]$
  - try to recover the parity bit $\beta^* = \bigoplus_{i=0}^{R-1} \langle \Gamma_i, rk_i^* \rangle$
- Given a sample or data D, compute the undersampled correlation $\hat{\varepsilon}(\Gamma_0, \Gamma_{R-1}; rk^*, D)$
  - determine β^* to be 0 iff $\hat{\varepsilon}(\Gamma_0, \Gamma_{R-1}; rk^*, D)C(\Gamma)$>0

$$\hat{\varepsilon}(\gamma, \gamma'; rk, D) \coloneqq \frac{1}{|D|} \sum_{(P,C)\in D} (-1)^{\langle \gamma, P \rangle \oplus \langle \gamma', C \rangle}$$



Algorithm 1

- Right Key Hypothesis
  - Γ: dominant trail

$\Rightarrow$ $X = (-1)^{\beta^*} \hat{\varepsilon}(\gamma, \gamma'; rk^*, D)$: random variable letting D vary with $|D| = N$

$X \sim \mathcal{N}(\epsilon, 1/N)$ $\epsilon = C(\Gamma)$ $\beta^* = \bigoplus_{i=0}^{R-1} \langle \Gamma_i, rk_i^* \rangle$

- Success Probability
  - $P_{\mathrm{S}} = \Pr_{X \sim \mathcal{N}(\epsilon, 1/N)} (\epsilon X > 0) = \Phi\left(\sqrt{N}|\epsilon|\right)$

Algorithm 2

- Add outer rounds to a trail $\Gamma = [\Gamma_s, \dots, \Gamma_{s+r}]$ for the inner cipher $E|_s^{s+r}$
  - recover a parity bit and some outer round key bits
- Given D,
  - Use the statistic $(-1)^{\beta}\hat{\varepsilon}(\Gamma, rk^*, \kappa, D)$ to pick out candidates for (β^*, κ^*)
  - Proceed with trial encryption

$\beta^* = \bigoplus_{i=s}^{s+r-1} \langle \Gamma_i, rk_i^* \rangle$

β: indeterminate, binary

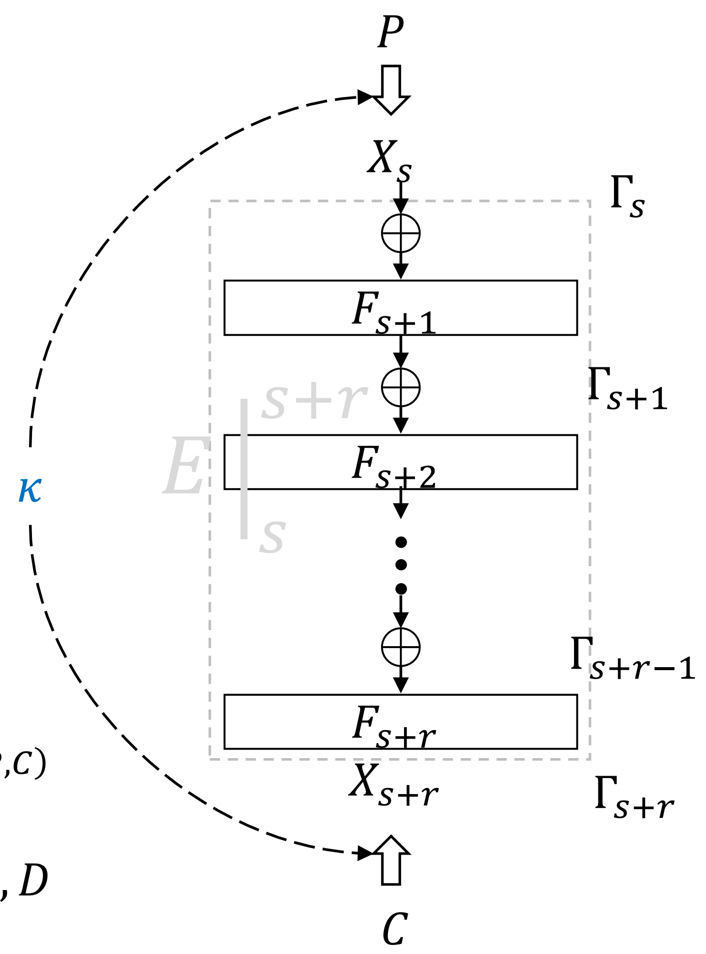
threshold based or rank based

κ: bit string obtained by concatenating outer round key bits involved in the outer round computation of $\langle \Gamma_s, X_s \rangle \oplus \langle \Gamma_{s+r}, X_{s+r} \rangle$

$$\hat{\varepsilon}(\Gamma, rk^*, \kappa, D) \coloneqq \frac{1}{|D|} \sum_{(P,C)\in D} (-1)^{g(\kappa,P,C)}$$

undersampled correlation gotten from κ, D

$\langle \Gamma_s, X_s \rangle \oplus \langle \Gamma_{s+r+1}, X_{s+r+1} \rangle = g(\kappa, P, C)$



Algorithm 2

- Right Key Hypothesis (on the distribution of right key statistic)
  - $(-1)^{\beta^*}\hat{\varepsilon}(\Gamma,\kappa^*,D)\sim\mathcal{N}(\epsilon,\frac{1}{N})$ as D varies with $|D|=N$
- Wrong Key Hypothesis (on the distribution of wrong key statistic)
  - $\hat{\varepsilon}(\Gamma,\kappa,D)\sim\mathcal{N}(0,\frac{1}{N})$ as (κ,D) varies with $\kappa\neq\kappa^*$
- Hypothesis on independence [Sel08]
  - the order statistics for the wrong key statistics & the right key statistic are independent

success probability, advantage can be estimated
for threshold/rank based methods

Algorithm 2 style attacks (multiple appr.)

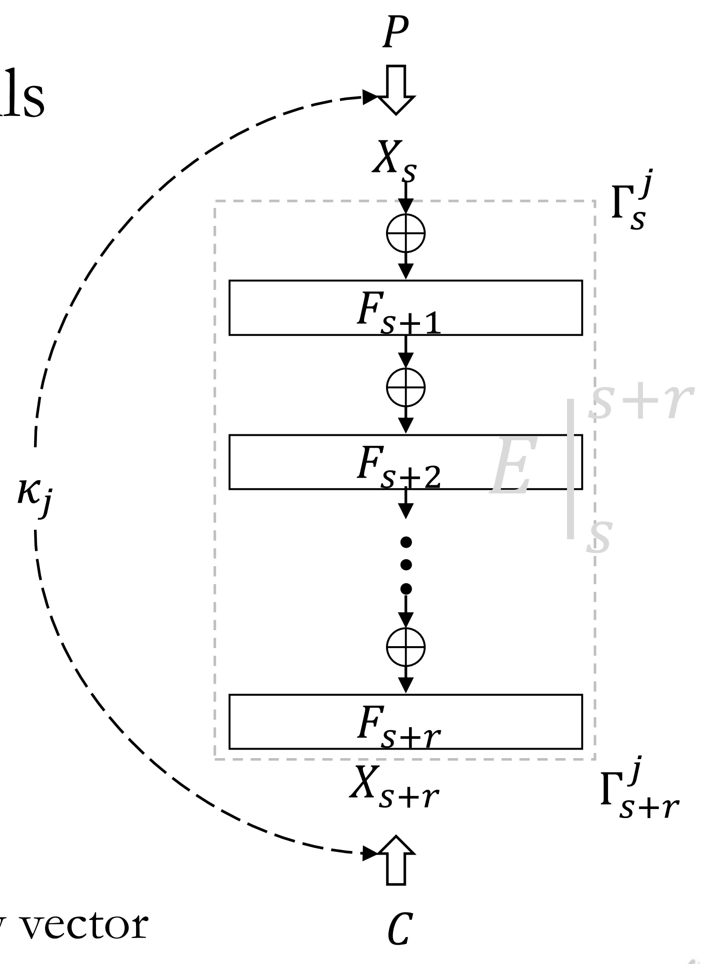
- $\Gamma^1, \Gamma^2, \dots, \Gamma^m$: dominant, statistically independent trails
  - $\epsilon_j = C(\Gamma^j)$ $(j = 1, \dots, m)$, $\epsilon = \sqrt{\sum_j \epsilon_j^2}$
- Given data D, recover $(\boldsymbol{\kappa}^*, \boldsymbol{\beta}^*)$,
  - $\boldsymbol{\kappa}^*$: correct value of the outer key $\boldsymbol{\kappa}$

    $\boldsymbol{\kappa}$: bit string obtained by combining of κ_j's (removing redundancy)
  - $\boldsymbol{\beta}^* = (\beta_1^*, \dots, \beta_m^*), \beta_j^* = \bigoplus_{i=s}^{s+r-1} \langle \Gamma_i^j, rk^* \rangle$
- Use the statistic $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) := \sum_j (-1)^{\beta_j} \epsilon_j \tau_j(\kappa_j, D)$

κ_j: bit string obtained by concatenating outer round key bits involved in the outer round computation of $\langle \Gamma_s^j, X_s \rangle \oplus \langle \Gamma_{s+r}^j, X_{s+r} \rangle$

$\tau_j(\kappa_j, D) := N\hat{\varepsilon}(\Gamma^j, \kappa_j, D)$

$\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$: any binary vector



assume for simplicity that bits of κ_j's are either identical or independent

Algorithm 2 style attacks (multiple appr.)

$$T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) \coloneqq \sum_j (-1)^{\beta_j} \epsilon_j \tau_j(\kappa_j, D)$$

- Algorithm 2MT (Threshold based):
  Pick out $(\boldsymbol{\kappa}, \boldsymbol{\beta})$'s with $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) \geq \theta = tN^2$
- Algorithm 2MR (Rank based):
  Rank $(\boldsymbol{\kappa}, \boldsymbol{\beta})$'s according to $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D)$
- Algorithm 2MC (Combined):
  Pick out candidates $(\boldsymbol{\kappa}, \boldsymbol{\beta})$'s with $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) \geq \theta$ and then rank them
  - yields better advantage than Algorithm 2MT for $P_S \approx 1$

Algorithm 2 style attacks (multiple appr.)

- Wrong key types
  - For $J_O \subsetneq \{1, \dots, m\}$,
    $\boldsymbol{\kappa}$ is said to have the wrong key type J_O if $\{j: \kappa_j = \kappa_j^*\} = J_O$

    $W(J_O)$: the set of $\boldsymbol{\kappa}$'s having the wrong key type J_O

  - For $J_O, J_I \subset \{1, \dots, m\}$ s.t. $J_O \neq \{1, \dots, m\}$ or $J_I \neq \{1, \dots, m\}$,
    $(\boldsymbol{\kappa}, \boldsymbol{\beta})$ is said to have the wrong key type (J_O, J_I) if
    - $\boldsymbol{\kappa}$ has the wrong key type J_O and $\boldsymbol{\beta}$ has the type J_I

For $J \subset \{1, \dots, m\}$,
$\boldsymbol{\beta}$ is said to have the type J if $\{j: \beta_j = \beta_j^*\} = J$
If $\boldsymbol{\beta}$ has the type J , denote it by $\boldsymbol{\beta}^J$

$W(J_O, J_I)$: the set of $(\boldsymbol{\kappa}, \boldsymbol{\beta})$'s having the wrong key type(J_O, J_I)

Multivariate Normal Distributions

$\boldsymbol{\mu} \in \mathbb{R}^m$, $\boldsymbol{\Sigma}$: positive definite $m \times m$ matrix over $\mathbb{R}$

- An m-variate random variable $\boldsymbol{X}$ is said to have the normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if it has the p.d.f.

$$\boldsymbol{x} \mapsto \frac{1}{(2\pi)^{m/2}|\det(\boldsymbol{\Sigma})|^{1/2}} e^{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}}$$

$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

- Probability that an m-variate normal random variable satisfies a linear inequality
  - $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{a} \in \mathbb{R}^m$, $\boldsymbol{a} \neq 0$, $b \in \mathbb{R}$
    - $\Pr_{\boldsymbol{X}}(\langle \boldsymbol{a}, \boldsymbol{X} \rangle + b \geq 0) = \Phi(\frac{\langle \boldsymbol{a}, \boldsymbol{\mu} \rangle + b}{|\boldsymbol{\sigma}^T \boldsymbol{a}|})$

$\boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$

Φ: c.d.f. of the std normal distribution

Algorithm 2 style attacks (multiple appr.)

For each $J_O \subset \{1, \ldots, m\}$

- $\boldsymbol{X}_{J_O}$: vector-valued random variable having the distribution determined by the values $((-1)^{\beta_1^*}\epsilon_1\tau_1(\kappa_1, D), \ldots, (-1)^{\beta_m^*}\epsilon_m\tau_m(\kappa_m, D))$

$|D| = N, \boldsymbol{\kappa} \in W(J_O)$

- Hypothesis: $\boldsymbol{X}_{J_O} \sim \mathcal{N}(\boldsymbol{\mu}_{J_O}, \boldsymbol{\Sigma}_{J_O})$
  - $\boldsymbol{\mu}_{J_O} = (\mu_1, \ldots, \mu_m); \mu_j = N\epsilon_j^2$ for $j \in J_O, \mu_j = 0$ for $j \notin J_O$
  - $\boldsymbol{\Sigma}_{J_O} = \mathrm{diag}(N\epsilon_1^2, \ldots, N\epsilon_m^2)$

$\Rightarrow$ distribution $\mathcal{D}_{J_O}$

Algorithm 2 style attacks (multiple appr.)

For each J_O

Let $\{1, \dots, m\} \setminus J_O = \{j_1, \dots, j_u\}$

- $\widehat{\boldsymbol{X}}_{J_O}$: vector-valued random variable having the distribution determined by

$$\big(\underbrace{(-1)^{\beta_1^*}\epsilon_1\tau_1(\kappa_1^*, D), \dots, (-1)^{\beta_m^*}\epsilon_m\tau_m(\kappa_m^*, D)}_{\text{right key statistic}}, \underbrace{\epsilon_{j_1}\tau_{j_1}\left(\kappa_{j_1}, D\right), \dots, \epsilon_{j_u}\tau_{j_u}\left(\kappa_{j_u}, D\right)}_{\text{wrong key statistic}}\big)$$

$|D| = N, \boldsymbol{\kappa} \in W(J_O)$

- Hypothesis (Stronger): $\widehat{\boldsymbol{X}}_{J_O} \sim \mathcal{N}(\widehat{\boldsymbol{\mu}}_{J_O}, \widehat{\boldsymbol{\Sigma}}_{J_O})$
  - $\widehat{\boldsymbol{\mu}}_{J_O} = (\mu_1, \dots, \mu_{m+u}), \widehat{\boldsymbol{\Sigma}}_{J_O} = \mathrm{diag}(\sigma_1^2, \dots, \sigma_{m+u}^2);$

$$(\mu_j, \sigma_j^2) = (N\epsilon_j^2, N\epsilon_j^2) \text{ for } j \in \{1, \dots, m\}, \ (\mu_{m+l}, \sigma_{m+l}^2) = (0, N\epsilon_{j_l}^2) \text{ for } l \in \{1, \dots, u\}$$

➡ distribution $\hat{\mathcal{D}}_{J_O}$

Algorithm 2MT

$$T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) := \sum_j (-1)^{\beta_j} \epsilon_j \tau_j(\kappa_j, D)$$

- Determine $(\boldsymbol{\kappa}, \boldsymbol{\beta})$ to be correct if
  - $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) \geq tN\epsilon^2$
- Success Probability $p_S(t)$:
  - $\Pr_D(T(\boldsymbol{\kappa}^*, \boldsymbol{\beta}^*, D) \geq tN\epsilon^2)$
    $= \Pr_{\boldsymbol{X} \sim \mathcal{D}_{\{1,\dots,m\}}} (X_1 + \cdots + X_m \geq tN\epsilon^2) = \Phi((1-t)\sqrt{N}\epsilon)$

    linear inequality
- False alarm probability: $\frac{1}{2^{k_O+m}} \times \sum_{(J_O,J_I):wrong} |W(J_O)| p_{\mathrm{fa}}^{2T,(J_O,J_I)}(t)$
  - $p_{\mathrm{fa}}^{2T,(J_O,J_I)}(t)$: probability that $(\boldsymbol{\kappa}, \boldsymbol{\beta})$ of type (J_O, J_I) satisfies the threshold condition

k_O: number of bits in $\boldsymbol{\kappa}$

Algorithm 2MT

- False alarm probability $p_{\text{fa}}^{2T,(J_O,J_I)}$ for type (J_O, J_I)

  $\Pr_{D,\boldsymbol{\kappa}\in W(J_O)}(T(\boldsymbol{\kappa},\boldsymbol{\beta}^{J_I},D) \geq tN\epsilon^2) = \Pr_{\boldsymbol{X}\sim\mathcal{D}_{J_O}}(\sum_{j\in J_O\cap J_I} X_j - \sum_{j\in J_O\setminus J_I} X_j + \sum_{l=1}^{u}(-1)^{\beta_{j_l}} X_{m+l}) \geq tN\epsilon^2)$

  $= \Phi(\sqrt{N}(\sum_{j\in J_O\cap J_I}\epsilon_j^2 - \sum_{j\in J_O\setminus J_I}\epsilon_j^2 - t\epsilon^2)/\epsilon)$

  linear inequality

- The false alarm probability $p_{\text{fa}}^{2\text{T}}(t)$
  - $\frac{1}{2^{k_O+m}}\sum_{(J_O,J_I):\text{wrong}} |W(J_O)| p_{\text{fa}}^{2T,(J_O,J_I)}(t)$

  $\approx \Phi(-t\sqrt{N}\epsilon)$ (in many cases)

- Advantage: $-\log_2 p_{\text{fa}}^{2\text{T}}(t)$

Algorithm 2MR

- Rank $(\boldsymbol{\kappa}, \boldsymbol{\beta})$ according to the statistic $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D)$
- Success Probability = 1
- False alarm probability: $\frac{1}{2^{k_O+m}} \times \sum_{(J_O,J_I):wrong} |W(J_O)| p_{\text{fa}}^{2R,(J_O,J_I)}$
  - $p_{\text{fa}}^{2R,(J_O,J_I)}$: probability that $(\boldsymbol{\kappa}, \boldsymbol{\beta})$ of type (J_O, J_I) is ranked higher than $(\boldsymbol{\kappa}^*, \boldsymbol{\beta}^*)$

Algorithm 2MR

- False alarm probability $p_{\mathrm{fa}}^{2R,(J_O,J_I)}$ for type (J_O, J_I):
$$\Pr_{D,\boldsymbol{\kappa}\in W(J_O)}(T(\boldsymbol{\kappa},\boldsymbol{\beta}^{J_I},D) \geq T(\boldsymbol{\kappa}^*,\boldsymbol{\beta}^*,D))$$
$$= \Pr_{\boldsymbol{X}\sim\mathcal{D}_{J_O}}\Big(-\sum_{j:j\leq m, j\notin J_O} X_j - 2\sum_{j\in J_O\setminus J_I}(-1)^{\beta_j^*}X_j + \sum_{l=1}^{u}(-1)^{\beta_{j_l}^*}X_{m+l}\Big) \geq tN\epsilon^2\Big)$$
linear inequality
$$= \Phi\left(-\left(N\Big(\sum_{j\in J_O\setminus J_I}\epsilon_j^2 + \frac{1}{2}\sum_{j\in\{1,\ldots,m\}\setminus J_O}\epsilon_j^2\right)^{1/2}\right)$$
- The false alarm probability $p_{\mathrm{fa}}^{2\mathrm{R}}$
  - $\frac{1}{2^{k_O+m}}\sum_{(J_O,J_I):\mathrm{wrong}}|W(J_O)|p_{\mathrm{fa}}^{2R,(J_O,J_I)} \approx \Phi(-\sqrt{N/2}\epsilon)$ (in many cases)
- Advantage: $-\log_2 p_{\mathrm{fa}}^{2\mathrm{R}} - 1$

Algorithm 2MC

- Pick out $\boldsymbol{\beta}$'s with $T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) \geq tN\epsilon^2$ and then rank them according to the statistic
- Success Probability: the same as in Algorithm 2MT
  - $\Phi((1-t)\sqrt{N}\epsilon)$
- False alarm probability: $\frac{1}{2^{k_O+m}} \times \sum_{(J_O,J_I):wrong} |W(J_O)| p_{\mathrm{fa}}^{2C,(J_O,J_I)}(t)$
  - $p_{\mathrm{fa}}^{2C,(J_O,J_I)}(t)$: probability that $(\boldsymbol{\kappa}, \boldsymbol{\beta})$ of type (J_O, J_I) is ranked higher than $(\boldsymbol{\kappa}^*, \boldsymbol{\beta}^*)$ *and satisfies the threshold condition*

Algorithm 2MC

- False alarm probability $p_{\mathrm{fa}}^{2C,(J_O,J_I)}(t)$ for type (J_O,J_I):

$$\Pr_{D,\boldsymbol{\kappa}\in W(J_O)}\left(T(\boldsymbol{\kappa},\boldsymbol{\beta}^{J_I},D)\geq T(\boldsymbol{\kappa}^*,\boldsymbol{\beta}^*,D),T(\boldsymbol{\kappa},\boldsymbol{\beta}^{J_I},D)\geq tN\epsilon^2\right)$$

Two linear inequalities

can be estimated numerically or by simulation

- The false alarm probability $p_{\mathrm{fa}}^{2\mathrm{C}}(t)$
  - $\frac{1}{2^{k_O+m}}\sum_{(J_O,J_I):\mathrm{wrong}}|W(J_O)|p_{\mathrm{fa}}^{2\mathrm{C},(J_O,J_I)}(\mathrm{t}) \approx p_{\mathrm{fa}}^{2\mathrm{C},(\emptyset,\emptyset)}(\mathrm{t})$ (in many cases)

- Advantage: $-\log_2 p_{\mathrm{fa}}^{2\mathrm{C}}(t)$

Application to DES

- Exploit 4 linear trails [BV17]
  - Γ^1: $\epsilon_1 = C(\Gamma_1) = -2^{-19.75}$, $k_O^1 = 12$
  - Γ^2: $\epsilon_2 = C(\Gamma_2) = -2^{-20.07}$, $k_O^2 = 18$
  - Γ^3: $\epsilon_3 = C(\Gamma_3) = -2^{-19.75}$, $k_O^3 = 12$
  - Γ^4: $\epsilon_4 = C(\Gamma_4) = -2^{-20.07}$, $k_O^4 = 18$

κ_1, κ_2 share 6 bits

κ_3, κ_4 share 6 bits

$\kappa_1 || \kappa_2$ and $\kappa_3 || \kappa_4$ does not have any bits in common

$\boldsymbol{\kappa}$ has 48 bits: $k_O = 48$

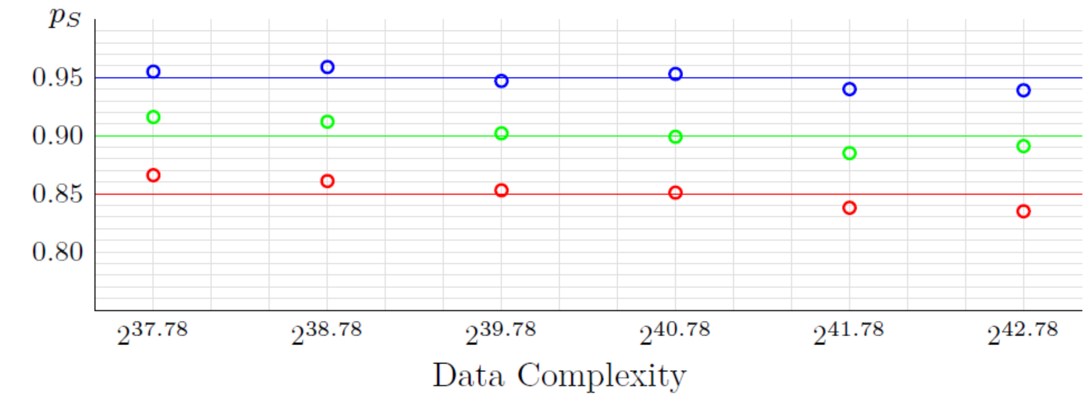
$\epsilon = 2^{-18.89}$

- Perform Algorithm 2MC, given data D of size N:
  - compress data and get 4 lists L_j's applying FWHT.
  - combine lists L_1 and L_2 to get a list $L_{1,2}$; combine lists L_3 and L_4 to get a list $L_{3,4}$
  - Sort $L_{1,2}$ and $L_{3,4}$ and get the list $L_{1,2,3,4}$ considering the threshold condition
  - Try the candidates in $L_{1,2,3,4}$ one by one

$T(\boldsymbol{\kappa}, \boldsymbol{\beta}, D) \geq tN\epsilon^2$

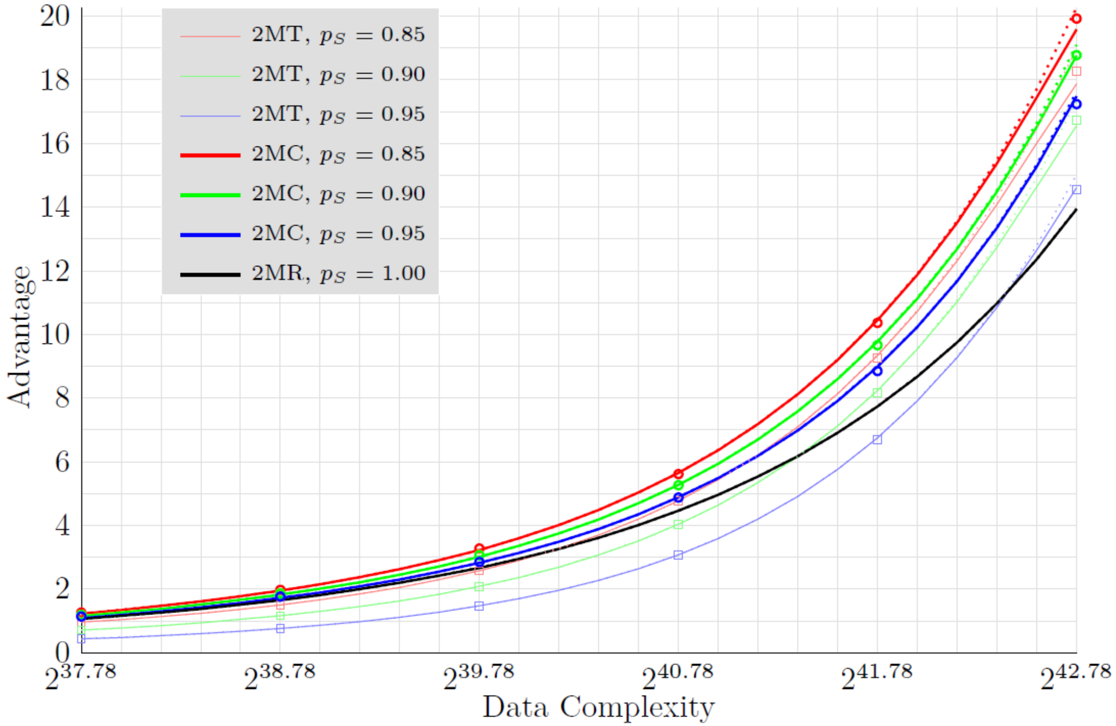
Application to DES

theoretical/experimental P_S



1,000 experiments
N up to $2^{42.78}$

theoretical/experimental advantage



Multiple linear cryptanalysis [BCQ04]

- Algorithm 1 and Algorithm 2 style attacks
  - formulas for advantage estimated in terms of trail correlations and data complexity
    - rank based, P_S fixed to 1
- limitations
  - advantage not analyzed theoretically for $P_S < 1$
  - experimental advantage not satisfactory
    - e.g. when applied to DES [BV17]

Multidimensional linear cryptanalysis [HCN09]

- Algorithm 1 and Algorithm 2 style attacks
  - threshold based or rank based
  - use LLR statistic or χ^2 statistic
  - approximate, asymptotic advantages theoretically provided
  - under certain independence assumptions
  - does not require trails to be dominant
- does not yield attack better than [Mat94] on DES
  - advantage not satisfactory when using a small number of trails
  - LLR method more effective, but not separable in general:
    adding outer rounds requires much overhead

Recent linear attacks on DES

- multiple linear cryptanalysis using 8 dependent trails [BV17]
- conditional linear cryptanalysis [BP19]
- analysis using a separable statistic [FS19]

Our attacks have comparable complexities;
advantageous with smaller data size.

cf. 2^{43} data/ 2^{43} time / 0.85 [Mat94]

Attack	Data	Time	p_S	Reference
Multiple LC	$2^{42.78}$	$2^{38.86}$	0.85	[BV17]
	$2^{41.00}$	$2^{49.76}$	0.80	
MultiDim. LC	$2^{41.81}$	$2^{41.81} + O(2^{41.81})$	0.83	[FS18]
	$2^{41.85}$	$2^{41.85} + O(2^{41.85})$	0.85	
Conditional LC	$2^{42.00}$	$2^{41.00}$	0.82	[BP18]
	$2^{41.90}$	$2^{41.90}$	0.85	
	$2^{41.00}$	$2^{50.00}$	0.92	
	$2^{40.00}$	$2^{52.00}$	0.82	
Multiple LC	$2^{42.75}$	$2^{38.87}$	0.85	**This Work**
	$2^{42.00}$	$2^{42.35}$	0.80	
	$2^{41.90}$	$2^{43.77}$	0.85	
	$\mathbf{2^{41.00}}$	$\mathbf{2^{48.17}}$	**0.80**	
	$\mathbf{2^{41.00}}$	$\mathbf{2^{49.23}}$	**0.95**	
	$\mathbf{2^{40.00}}$	$\mathbf{2^{51.14}}$	**0.80**	
	$\mathbf{2^{40.00}}$	$\mathbf{2^{51.89}}$	**0.95**	

Merits of the attack

- Why efficient?
  - the linear statistic
    - separable: overhead in adding outer rounds minimized
    - almost the same as the optimal LLR statistic up to a constant
  - parity bits recovered at the same time ⇒ advantage increased
    - χ^2 method does not consider recovering parity bits
    - existing LLR methods usually assume parity bits are known
  - multivariate normal distribution
    - allows to get estimates of attack complexity better than using order statistics

Generalization

- Exploit close-to-dominant, dependent trails
- Use modified hypotheses on the distributions of multivariate random variables
  - presume multivariate normal distributions but with different mean vectors and covariance matrices – need to be precomputed in advance
- Perform the same procedure with similar statistics
  - Use linear statistics with varying coefficients
- P_{S}, P_{fa} can be computed in the same way for each attack
  - probability of regions represented by linear inequalities for an multivariate normal random variable

Conclusion

- Multiple linear attacks using multiple dominant linear trails
  - statistical models regarding the distribution of vector valued random variables consisting of component statistics
  - closed formulas for success probability and advantage of various Algorithm 1 and Algorithm 2 style attacks in terms of data size, correlations of the trails, and threshold parameter incorporating the decomposition of outer key bits
  - best advantage among existing linear attacks when exploiting multiple dominant statistical independent trails
- Application to DES
  - exhibit the validity of the statistical models
  - show the effectiveness of the attack