Some Open Issues in Side Channel Analysis

Christophe Giraud and **Emmanuel Prouff**
e.prouff,c.giraud@oberthur.com

Oberthur Technologies

Cryptography and Security for Embedded Systems – June. 2009

- Classical Point of View (Linear/Differential/Algebraic attacks):

$$C = \mathrm{E}_{\mathrm{Secret}}(M)$$

Goal: from pairs (m_i, c_i) and the math. description of E recover *Secret*.

- SCA Point of view: $\mathrm{E}_{\mathrm{Secret}}(M)$ is a sequence of operations

$$O_1(M, K_1), O_2(M, K_2), \cdots, O_j(M, K_j), \cdots ,$$

K_j is a function of *Secret* and $|K_j|$ is small (*e.g.* 8 bits). Associate $O_j(m_i, K_j)$ with an observation vector $\mathbf{l}_i(K_j)$.

Goal: from pairs $(m_i, \mathbf{l}_i)$ recover K_j for several j's and then *Secret*.

Side Channel Attacks of Block Ciphers

- SCA are divide-and-conquer attacks.
- Main Issue: recover a sub-key $k^\star$ from an observation $\mathbf{L}(k^\star)$ of the manipulation of $O(M, k^\star)$.
- (Almost) straightforward if $\mathbf{L}(k^\star) = O(M, k^\star)$ (perfect observation).
- What happens for unperfect observations?
- Core Idea: $\mathbf{L}(k^\star)$ and $O(M, k^\star)$ satisfy

$$\mathbf{L}(k^\star) = \varphi \circ O(M, k^\star) + B$$

where B is a noise and φ is a unknown leakage function.

- Assumption: $B \sim \mathcal{N}(0, \sigma)$.

Side Channel Attacks of Block Ciphers

- SCA are divide-and-conquer attacks.
- Main Issue: recover a sub-key $k^\star$ from an observation $\mathbf{L}(k^\star)$ of the manipulation of $O(M, k^\star)$.
- (Almost) straightforward if $\mathbf{L}(k^\star) = O(M, k^\star)$ (perfect observation).
- What happens for unperfect observations?
- Core Idea: $\mathbf{L}(k^\star)$ and $O(M, k^\star)$ satisfy

$$\mathbf{L}(k^\star) = \varphi \circ O(M, k^\star) + B$$

where B is a noise and φ is a unknown leakage function.

- Assumption: $B \sim \mathcal{N}(0, \sigma)$.

Side Channel Attacks of Block Ciphers

- SCA are divide-and-conquer attacks.
- Main Issue: recover a sub-key $k^\star$ from an observation $\mathbf{L}(k^\star)$ of the manipulation of $O(M, k^\star)$.
- (Almost) straightforward if $\mathbf{L}(k^\star) = O(M, k^\star)$ (perfect observation).
- What happens for unperfect observations?
- Core Idea: $\mathbf{L}(k^\star)$ and $O(M, k^\star)$ satisfy

$$\mathbf{L}(k^\star) = \varphi \circ O(k^\star) + B$$

where B is a noise and φ is a unknown leakage function.

- Assumption: $B \sim \mathcal{N}(0, \sigma)$.

Side Channel Attacks of Block Ciphers

Core assumptions:

1. $\mathbf{L}(K)$ depends on K.
2. $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.

Side Channel Attacks of Block Ciphers

Core assumptions:

1. $\mathbf{L}(K)$ depends on K.
2. $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.

What do they imply?

Side Channel Attacks of Block Ciphers

Core assumptions:

1. $\mathbf{L}(K)$ depends on K.
2. $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.

What do they imply?

- (1) $\Rightarrow \mathrm{I}(\mathbf{L}(K), K) \neq 0$.

Side Channel Attacks of Block Ciphers

Core assumptions:

1. $\mathbf{L}(K)$ depends on K.
2. $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.

What do they imply?

- (1) $\Rightarrow \mathrm{I}(\mathbf{L}(K), K) \neq 0$.
- (2) $\Rightarrow$There exist functions f s.t.

$$\mathrm{I}(\mathbf{L}(k^\star), f(O(k^\star))) \neq 0$$

and (possibly) pairs $(\mathcal{C}, f)$ s.t.

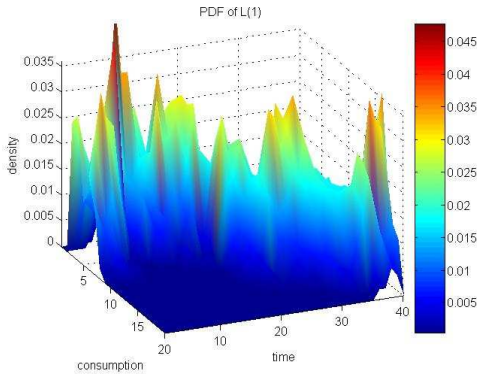
$$\rho(\mathcal{C} \circ \mathbf{L}(k^\star), f(O(k^\star)) \neq 0$$

Key dependency of Observations

Assumption: the pdf of $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.
Experimental Illustration: O is the first DES S-box computation.

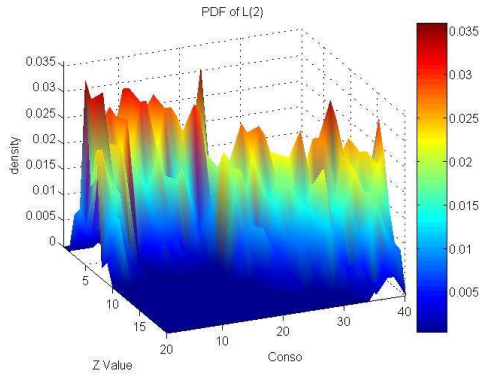
Key dependency of Observations

Assumption: the pdf of $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.
Experimental Illustration: O is the first DES S-box computation.



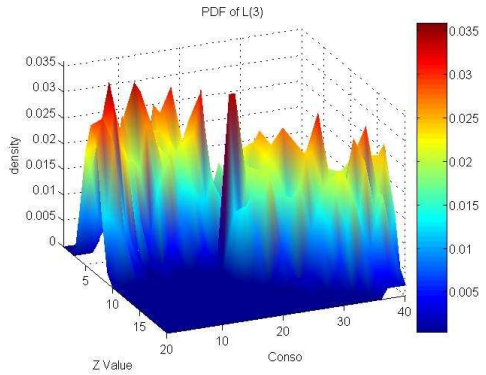
Key dependency of Observations

Assumption: the pdf of $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.
Experimental Illustration: O is the first DES S-box computation.



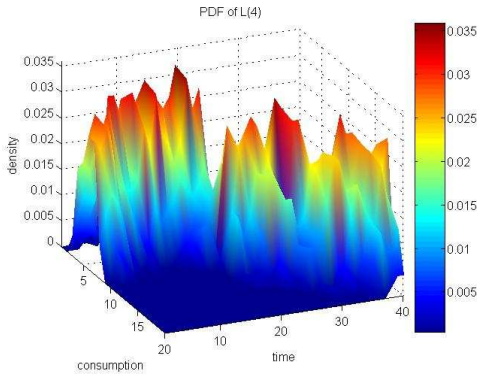
Key dependency of Observations

Assumption: the pdf of $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.
Experimental Illustration: O is the first DES S-box computation.



Key dependency of Observations

Assumption: the pdf of $\mathbf{L}(k^\star)$ depends on $O(k^\star)$.
Experimental Illustration: O is the first DES S-box computation.



Template Attacks [CRR03]

Goal: from a sample $(m_i, \mathbf{l}_i)_i \leftarrow (M, \mathbf{L}(k^\star))$ and O, retrieve $k^\star$.

Method

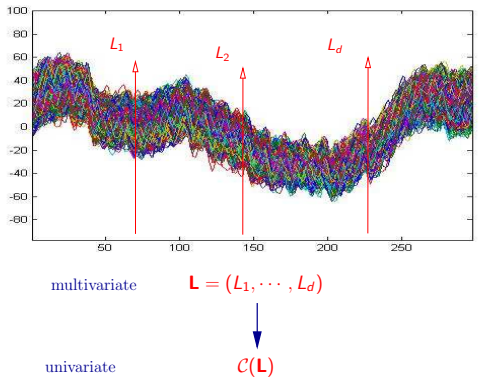
- For every k compute an estimation $\hat{\mathbf{L}}(k)$ of $\mathbf{L}(k)$
- Then for $(m_i, \mathbf{l}_i)_i \leftarrow (M, \mathbf{L})$ output

$$\hat{k} = \text{argmax}_k \prod_i \text{Prob}[\hat{\mathbf{L}}(k) = \mathbf{l}_i \mid M = m_i] \ .$$

- If the template estimations $\hat{\mathbf{L}}(k)$ are good enough, then $\hat{k} = k^\star$ [Maximum Likelihood Approach].
- Issues: estim. of the $\mathbf{L}(k)$'s, high dim. of vector $\mathbf{L}$.
- Open Question: robustness of the templates.

Linear Correlation based attacks (LCA) [KJJ99,BCO04]

Idea: the joint distribution of some coordinates of $\mathbf{L}(K)$ depends on K.



Linear Correlation based attacks (LCA) [KJJ99,BCO04]

Goal: from a sample $(m_i, \mathbf{l}_i)_i \leftarrow (M, \mathbf{L}(k^\star))$ and O, retrieve $k^\star$.

Method

- Convert $\mathbf{L}(k^\star)$ into an univ. data $\mathcal{C}(\mathbf{L}(k^\star))$
- Then select an appropriate prediction function f and output

$$\hat{k} = \text{argmax}_k \, |\hat{\rho}\,(\mathcal{C}(\mathbf{L}(k^\star), f(O(k)))| \ ,$$

  where $\hat{\rho}$ is an estimation of the linear correlation coefficient based on $(m_i, \mathbf{l}_i)_i \leftarrow (M, \mathbf{L}(k^\star))$.
- If there is enough linear dependency, then $\hat{k} = k^\star$.
- (Open) Issues: design of the functions $\mathcal{C}$ and f such that $\forall k \neq k^\star$:

$$|\rho\,(\mathcal{C}(\mathbf{L}), f(O(k^\star)))| >> |\rho\,(\mathcal{C}(\mathbf{L}), f(O(k)))| \ .$$

Mutual Information based Attacks (MIA) [GBP08,RP09]

Goal: from a sample $(m_i, \mathbf{l}_i)_i \leftarrow (M, \mathbf{L}(k^\star))$ and O, retrieve $k^\star$.

Method

- Select an appropriate prediction function f and output

$$\hat{k} = \text{argmax}_k \hat{\mathrm{I}}(\mathbf{L}(k^\star), f(O(k))) \quad ,$$

  where $\hat{\mathrm{I}}$ is an estimation of the mutual information based on $(m_i, \mathbf{l}_i)_i \leftarrow (M, \mathbf{L}(k^\star))$.
- If there is any dependency and if the estimations are good enough, then $\hat{k} = k^\star$.
- (Open) Issues: design f and estimate the mutual information.

Study in Two Contexts

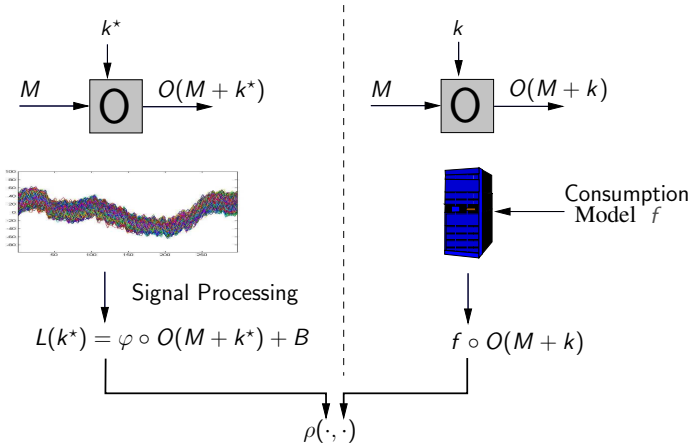
Univariate Study: there exists instantaneous leakage [observations $L(k^\star)$ of dim. 1 are sufficient!].

Unprotected Implementation.

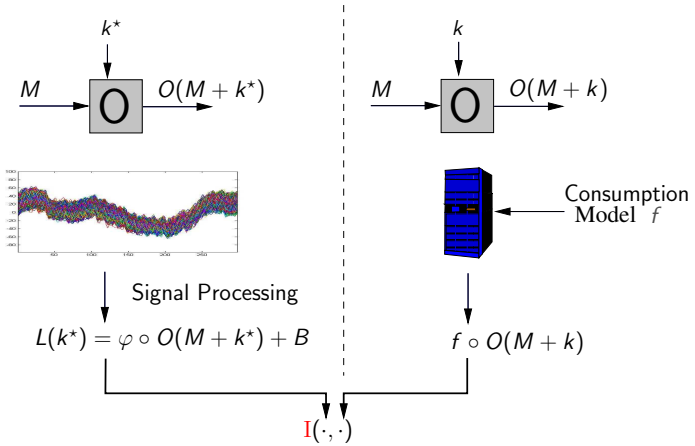
Multivariate Study: there does not exist instantaneous leakage [observations $\mathbf{L}(k^\star)$ of dim. greater than 2 are required!].

Protected Implementation.

Linear Correlation Based SCA in the Univariate Case



Mutual Information Based SCA in the Univariate Case



Mutual Information Based SCA in the Univariate Case

Mutual Information Based SCA in the Univariate Case

- How to model the consumption? How to choose f?

Mutual Information Based SCA in the Univariate Case

- How to model the consumption? How to choose f?
  Answer: select $f = HW$ (practical validation in [Mes00a,LP07]).

Mutual Information Based SCA in the Univariate Case

- How to model the consumption? How to choose f?
  Answer: select $f = HW$ (practical validation in [Mes00a,LP07]).
- Why does the MIA succeed in recovering the key $k^\star$?

Mutual Information Based SCA in the Univariate Case

- How to model the consumption? How to choose f?
  Answer: select $f = HW$ (practical validation in [Mes00a,LP07]).
- Why does the MIA succeed in recovering the key $k^\star$?
  Partial answer in [PR09].

Mutual Information Based SCA in the Univariate Case

- How to model the consumption? How to choose f?
  Answer: select $f = HW$ (practical validation in [Mes00a,LP07]).
- Why does the MIA succeed in recovering the key $k^\star$?
  Partial answer in [PR09].
- How to estimate the mutual information?

Mutual Information Based SCA in the Univariate Case

- How to model the consumption? How to choose f?
  Answer: select $f = HW$ (practical validation in [Mes00a,LP07]).
- Why does the MIA succeed in recovering the key $k^\star$?
  Partial answer in [PR09].
- How to estimate the mutual information?
  Partial answer in [GBTP08,PR09,...].

How to estimate the mutual information?

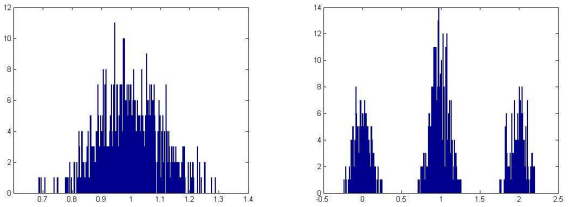
- Assumption: $(L(k^\star) \mid Z(k) = z)$ is Gaussian iff $k = k^\star$ (and is a Gaussian Mixture otherwise).
- Issue: estimate $\mathrm{H}\,(L(k^\star) \mid Z(k) = z)$ for every z and every k.
- Idea: first estimate the pdf $(L(k^\star) \mid Z(k) = z)$ and then compute/estimate the entropy.
- Various estimation methods: Histogram Method, Kernel Method, Parametric Method.

How to estimate the mutual information?

- Assumption: $(L(k^\star) \mid Z(k) = z)$ is Gaussian iff $k = k^\star$ (and is a Gaussian Mixture otherwise).
- Issue: estimate $\mathrm{H}\,(L(k^\star) \mid Z(k) = z)$ for every z and every k.
- Idea: first estimate the pdf $(L(k^\star) \mid Z(k) = z)$ and then compute/estimate the entropy.
- Various estimation methods: Histogram Method, Kernel Method, Parametric Method.
- Open Question: are there other estimation methods more efficient in our context?

Hist. Method Based MIA: Attack Simulations

- 10000 observations, noise std equal to 0.1, O equal to the first DES S-box, $f = Id$, #bins $= 285$.



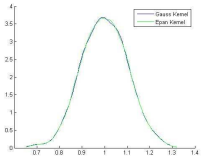
- Left: $\hat{\mathrm{H}}(L(11)|Z(11) = 1) = -1.31$ (theoretical: -1.27).
- Right: $\hat{\mathrm{H}}(L(11)|Z(5) = 1) = -0.0345$.
- Drawbacks: rough estimation, choice of #bins [Sil86].

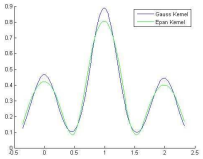
Kernel and Parameterized Methods Based MIA

- Kernel Methods [Par62]: same idea as Hist. Method except that the pdf is locally approximated by a continuous function called Kernel function.
- Assets: quality of the estimation is improved, require less observations than Hist. Meth.
- Drawbacks: selection of the Kernel and of the bandwith [Sil 86].

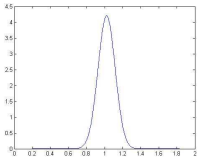
- Param. Methods: approximate the pdf by a Gaussian distribution.
- Assets: require less observations than Kernel Meth.
- Drawbacks: rough estimation (Gaussian Assumption).

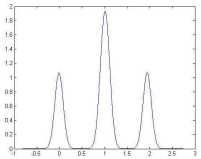
Kern. and Param. Meth. Based MIA: Attack Simulations





(a) 1st order for $k = 11$. (b) 1st order for $k = 5$.





SCA Comparisons in the Gaussian Model –Simulations–

- Number of measurements required to achieve a success rate of 90% according to the noise std σ.

Attack \ σ	0.5	1	2	5	10	15	20	50	100
LCA, $f = \mathsf{Id}$	30	30	100	1000	3000	7000	15000	70000	260000
MIA_H (Hist), $f = \mathsf{Id}$	80	160	600	4000	20000	50000	95000	850000	10^6+
MIA_K (Kernel), $f = \mathsf{Id}$	70	140	500	3000	15000	35000	60000	500000	10^6+
MIA_P (Param.), $f = \mathsf{Id}$	60	100	300	2000	5000	15000	20000	150000	500000
LCA, $f = \mathsf{HW}$	30	30	70	400	2000	4000	7000	45000	170000
MIA_H (Hist), $f = \mathsf{HW}$	40	70	300	1500	7000	20000	40000	320000	10^6+
MIA_K (Kernel), $f = \mathsf{HW}$	30	60	190	1500	5500	15000	25000	190000	900000
MIA_P (Param.), $f = \mathsf{HW}$	70	70	150	1000	3000	7000	15000	65000	300000

- LCA is always is better!
- Explanation: hypotheses and observations have strong linear dependency.

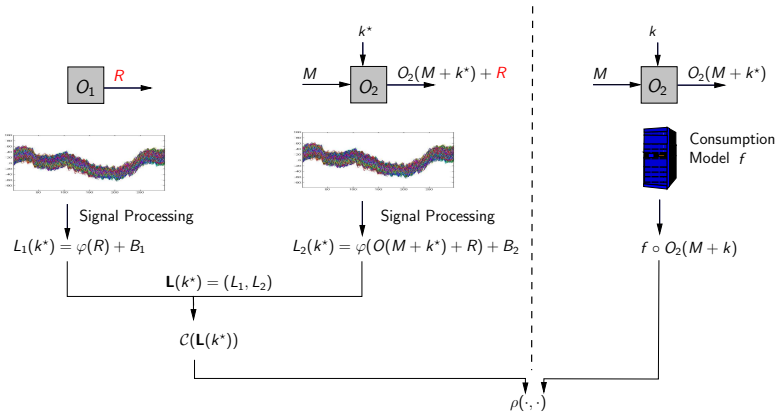
SCA Comparisons in the Gaussian Model –Simulations–

- Number of measurements required to achieve a success rate of 90% according to the noise std σ.

Attack \ σ	0.5	1	2	5	10	15	20	50	100
LCA, $f = \mathrm{Id}$	30	30	100	1000	3000	7000	15000	70000	260000
$\mathrm{MIA}_{\mathcal{H}}$ (Hist), $f = \mathrm{Id}$	80	160	600	4000	20000	50000	95000	850000	10^6+
MIA_{K} (Kernel), $f = \mathrm{Id}$	70	140	500	3000	15000	35000	60000	500000	10^6+
MIA_{P} (Param.), $f = \mathrm{Id}$	60	100	300	2000	5000	15000	20000	150000	500000
LCA, $f = \mathrm{HW}$	30	30	70	400	2000	4000	7000	45000	170000
$\mathrm{MIA}_{\mathcal{H}}$ (Hist), $f = \mathrm{HW}$	40	70	300	1500	7000	20000	40000	320000	10^6+
MIA_{K} (Kernel), $f = \mathrm{HW}$	30	60	190	1500	5500	15000	25000	190000	900000
MIA_{P} (Param.), $f = \mathrm{HW}$	70	70	150	1000	3000	7000	15000	65000	300000

- LCA is always is better!
- Explanation: hypotheses and observations have strong linear dependency.
- Open Question: does it exist practical contexts for which MIA is better than LCA?

Recall: Linear Correlation Based SCA for $d = 2$



Multivariate LCA [Mes00,JPS05,PSD05,OMHT05,PR09]

- Open Issue: Determine the pair $(\mathcal{C}, f)$ that maximizes $\rho_{k^\star}$ and minimizes ρ_k for $k \neq k^\star$.

Multivariate LCA [Mes00,JPS05,PSD05,OMHT05,PR09]

- Open Issue: Determine the pair $(\mathcal{C}, f)$ that maximizes $\rho_{k^\star}$ and minimizes ρ_k for $k \neq k^\star$.
- Partial Answer [PR09]: we can determine the optimal prediction function f that maximizes $\rho_{k^\star}$ knowing $\mathcal{C}$.

Multivariate LCA [Mes00,JPS05,PSD05,OMHT05,PR09]

- Open Issue: Determine the pair $(\mathcal{C}, f)$ that maximizes $\rho_{k^\star}$ and minimizes ρ_k for $k \neq k^\star$.
- Partial Answer [PR09]: we can determine the optimal prediction function f that maximizes $\rho_{k^\star}$ knowing $\mathcal{C}$.
- Partial Answer [PR09]: for $d = 2$ the product function is the better proposed combining function.

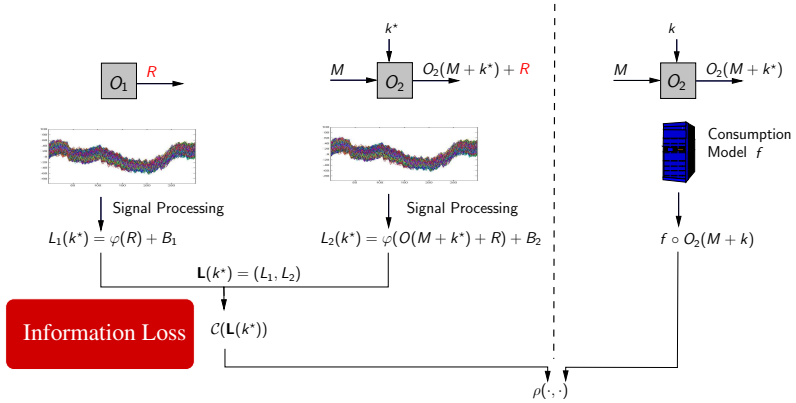
Multivariate LCA [Mes00,JPS05,PSD05,OMHT05,PR09]

- Open Issue: Determine the pair $(\mathcal{C}, f)$ that maximizes $\rho_{k^\star}$ and minimizes ρ_k for $k \neq k^\star$.
- Partial Answer [PR09]: we can determine the optimal prediction function f that maximizes $\rho_{k^\star}$ knowing $\mathcal{C}$.
- Partial Answer [PR09]: for $d = 2$ the product function is the better proposed combining function.
- Open Issue: is it the optimal one?

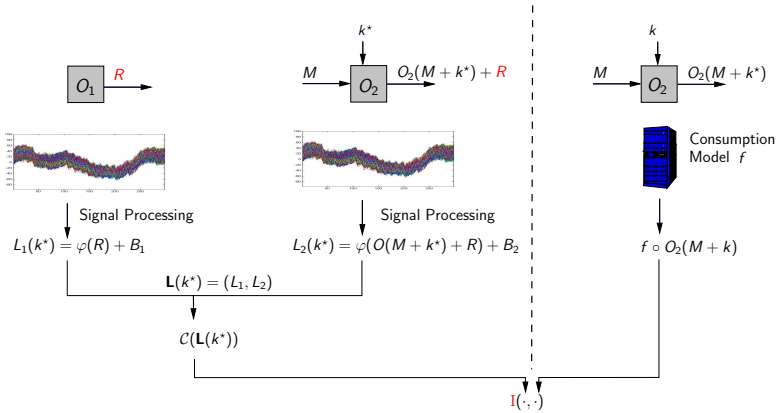
Multivariate LCA [Mes00,JPS05,PSD05,OMHT05,PR09]

- Open Issue: Determine the pair $(\mathcal{C}, f)$ that maximizes $\rho_{k^\star}$ and minimizes ρ_k for $k \neq k^\star$.
- Partial Answer [PR09]: we can determine the optimal prediction function f that maximizes $\rho_{k^\star}$ knowing $\mathcal{C}$.
- Partial Answer [PR09]: for $d = 2$ the product function is the better proposed combining function.
- Open Issue: is it the optimal one?
- Open Issue: what for $d > 3$?

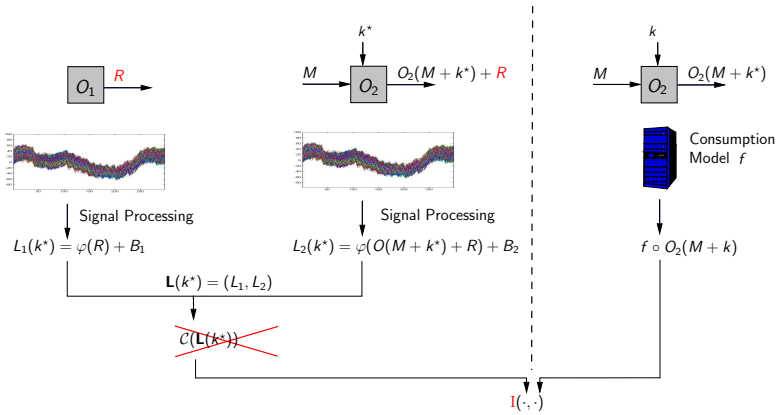
Recall: Linear Correlation Based SCA for $d = 2$



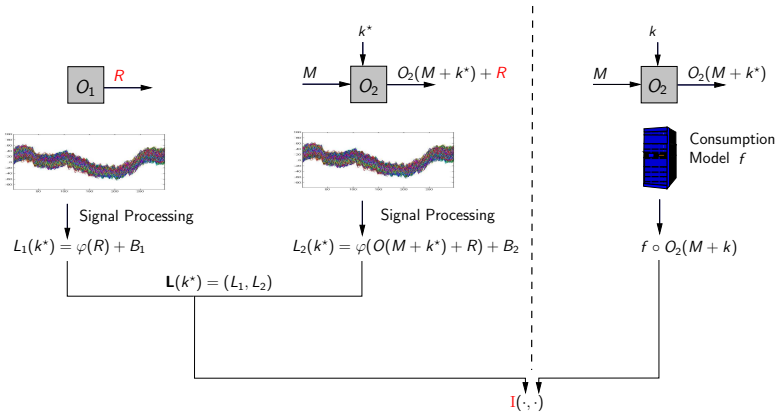
Recall: Mutual Information Based SCA for $d = 2$



A New Multivariate Distinguisher: The Mutual Information



A New Multivariate Distinguisher: The Mutual Information



Multivariate MIA [PR09-ACNS]

We have

$$\mathrm{H}(\mathbf{L}(k^\star)|Z(k)) = -\sum_{z \in \mathrm{Im}(f)} \mathrm{p}_Z(z) \times \mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k) = z\right) \ .$$

When noise is Gaussian, then we showed that:

Multivariate MIA [PR09-ACNS]

We have

$$\mathrm{H}(\mathbf{L}(k^\star)|Z(k)) = -\sum_{z \in \mathrm{Im}(f)} \mathrm{p}_Z(z) \times \mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k) = z\right) \ .$$

When noise is Gaussian, then we showed that:

- The pdf of $(\mathbf{L}(k^\star) \mid Z(k) = z)$ is a Gaussian Mixture g_θ

Multivariate MIA [PR09-ACNS]

We have

$$\mathrm{H}(\mathbf{L}(k^\star)|Z(k)) = -\sum_{z \in \mathrm{Im}(f)} \mathrm{p}_Z(z) \times \mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k) = z\right) \ .$$

When noise is Gaussian, then we showed that:

- The pdf of $(\mathbf{L}(k^\star) \mid Z(k) = z)$ is a Gaussian Mixture g_θ
- The key hypothesis k only plays a part in the definition of the weights of g_θ.

Multivariate MIA [PR09-ACNS]

We have

$$\mathrm{H}(\mathbf{L}(k^\star)|Z(k)) = -\sum_{z \in \mathrm{Im}(f)} \mathrm{p}_Z(z) \times \mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k) = z\right) \ .$$

When noise is Gaussian, then we showed that:

- The pdf of $(\mathbf{L}(k^\star) \mid Z(k) = z)$ is a Gaussian Mixture g_θ
- The key hypothesis k only plays a part in the definition of the weights of g_θ.
- The number of components in the mixture is minimum when $k = k^\star$.

Multivariate MIA [PR09-ACNS]

We have

$$\mathrm{H}(\mathbf{L}(k^\star)|Z(k)) = -\sum_{z\in\mathrm{Im}(f)} \mathrm{p}_Z(z) \times \mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k)=z\right) \ .$$

When noise is Gaussian, then we showed that:

- The pdf of $(\mathbf{L}(k^\star) \mid Z(k)=z)$ is a Gaussian Mixture g_θ
- The key hypothesis k only plays a part in the definition of the weights of g_θ.
- The number of components in the mixture is minimum when $k = k^\star$.
- Theoretical Condition: choose f s.t. $f \circ O$ is not injective.

Multivariate MIA [PR09-ACNS]

We have

$$\mathrm{H}(\mathbf{L}(k^\star)|Z(k)) = -\sum_{z \in \mathrm{Im}(f)} \mathrm{p}_Z(z) \times \mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k) = z\right) \ .$$

When noise is Gaussian, then we showed that:

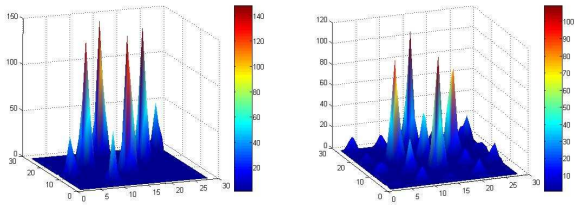
- The pdf of $(\mathbf{L}(k^\star) \mid Z(k) = z)$ is a Gaussian Mixture g_θ
- The key hypothesis k only plays a part in the definition of the weights of g_θ.
- Open Issue: prove that the entropy is minimum when $k = k^\star$? We have only experimental evidences!
- Theoretical Condition: choose f s.t. $f \circ O$ is not injective.

Estimation of the Multivariate MIA [PR09-ACNS]

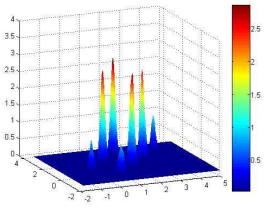
- Issue: estimate $\mathrm{H}\left(\mathbf{L}(k^\star) \mid Z(k) = z\right)$ for every z and every k.
- Idea: first estimate the pdf $\left(\mathbf{L}(k^\star) \mid Z(k) = z\right)$ and then compute/estimate the entropy.
- Various estimation methods: Histogram Method, Kernel Method, Parametric Method.
- Open Issue: improve the efficiency of the parametric method for multivariate observations.

Hist. Method Based MIA: Attack Simulations

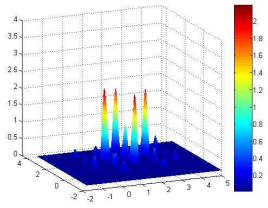
- 10000 observations, noise std equal to 0.1, O equal to the first DES S-box, $f = Id$, #bins $= 285$.



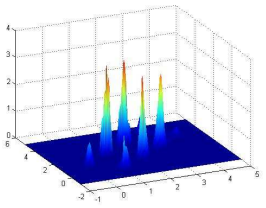
- Left: $\hat{\mathrm{H}}(\mathbf{L}(11)|Z(11) = 1) = 0.22$.
- Right: $\hat{\mathrm{H}}(\mathbf{L}(11)|Z(5) = 1) = 1.12$.
- Drawbacks: rough estimation, choice of #bins [Sil86].

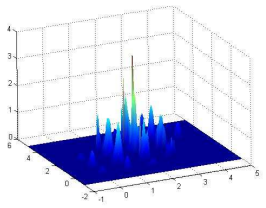


(a) 2nd order for $k = 11$.



(b) 2nd order for $k = 5$.





Comparisons

Table: Second Order Attack on DES S-box – Number of measurements required to achieve a success rate of 90% according to the noise standard deviation σ.

Attack σ	0.5	1	2	5	7	10
2O-LCA ($f = \mathrm{HW}$, $\mathcal{C}^{\star}_{diff}$)	300	600	4000	90000	10^6	10^6+
2O-LCA ($f = \mathrm{HW}$, $\mathcal{C}^{\star}_{prod}$)	300	600	3500	85000	400000	10^6+
2O-MIA$_H$ ($f = \mathrm{Id}$)	7000	7000	8000	15000	30000	55000

- For Univariate SCA: Correlation Based SCA are better than MI Based SCA: this means that our consumption models are quite good
- For Multivariate SCA: multivariate dependency quantifiers (as the Mutual Information) offer better results
- New Issues: pdf estimation for multivariate observations, improvement of the models, prove that combining by product is optimal (or not), extend analyses for orders greater than 2, is there any alternative to Correlation and MI based attacks(?), etc.

Thank you!
Questions and/or Comments?