Institut
Mines-Télécom

Less is More
Dimensionality Reduction from a Theoretical Perspective

CHES 2015 – Saint-Malo, France – Sept 13 - 16

Nicolas Bruneau, Sylvain Guilley, Annelie Heuser,
Damien Marion, and Olivier Rioul

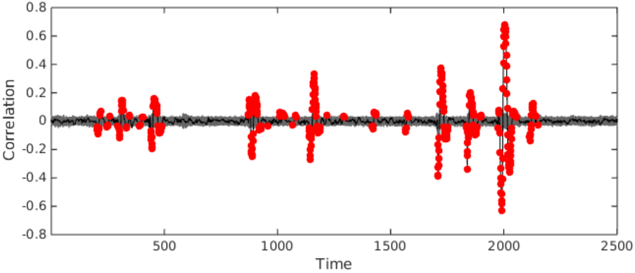
About us...

Nicolas BRUNEAU	Sylvain GUILLEY	Annelie HEUSER	Damien MARION	Olivier RIOUL
is also with	*is also with*	*is PhD fellow at*	*is also with*	*is also Prof at*
ST	SECURE IC THE TRUSTED COMPUTING COMPANY	Google	SECURE IC THE TRUSTED COMPUTING COMPANY	l'X ÉCOLE POLYTECHNIQUE UNIVERSITÉ PARIS-SACLAY

Overview

Overview

Motivation



large number of samples/ points of interest

Motivation

Problem (*profiled* and *non-profiled* side-channel distinguisher)

How to reduce dimensionality of multi-dimensional measurements?

Motivation

Problem (*profiled* and *non-profiled* side-channel distinguisher)

How to reduce dimensionality of multi-dimensional measurements?

Wish list

- simplification of the problem
- concentration of the information (to distinguish using fewer traces)
- improvement of the computational speed

State-of-the-Art I

Selection of points of interest

- manual selection of educated guesses [Oswald et al., 2006]
- automated techniques: sum-of-square differences (SOSD) and t-test (SOST) [Gierlichs et al., 2006]
- wavelet transforms [Debande et al., 2012]

State-of-the-Art I

Selection of points of interest

- manual selection of educated guesses [Oswald et al., 2006]
- automated techniques: sum-of-square differences (SOSD) and t-test (SOST) [Gierlichs et al., 2006]
- wavelet transforms [Debande et al., 2012]

Leakage detection metrics

- ANOVA (e.g. [Choudary and Kuhn, 2013, Danger et al., 2014]) or [Bhasin et al., 2014] (*Normalized Inter-Class Variance* (NICV))

State-of-the-Art II

Principal Component Analysis

- compact templates in [Archambeau et al., 2006]
- reduce traces in [Batina et al., 2012]
- eigenvalues as a security metric [Guilley et al., 2008]
- eigenvalues as a distinguisher [Souissi et al., 2010]

TELECOM ParisTech

State-of-the-Art II

Principal Component Analysis

- compact templates in [Archambeau et al., 2006]
- reduce traces in [Batina et al., 2012]
- eigenvalues as a security metric [Guilley et al., 2008]
- eigenvalues as a distinguisher [Souissi et al., 2010]

easily and accurately computed with no divisions involved

maximizing inter-class variance, but not intra-class variance

State-of-the-Art II

Linear Discriminant Analysis

- improved alternative
- takes inter-class variance and intra-class variance into account
- empirical comparisons [Standaert and Archambeau, 2008, Renauld et al., 2011, Strobel et al., 2014]

not easily and accurately computed with no divisions involved

maximizing inter-class variance and intra-class variance

State-of-the-Art II

Linear Discriminant Analysis

- improved alternative
- takes inter-class variance and intra-class variance into account
- empirical comparisons [Standaert and Archambeau, 2008, Renauld et al., 2011, Strobel et al., 2014]

But..

- advantages due to the statistical tools, their implementation, data set ...
- no clear rationale to prefer one method!

Contribution

- dimensional reduction in SCA from a theoretical viewpoint
- assuming attacker has full knowledge of the leakage
- derivation of the optimal dimensionality reduction

“Less is more”

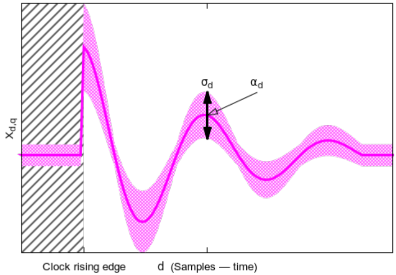
Advantages of dimensionality reduction can come with no impact on the attack success probability!

- comparison to PCA and LDA: theoretically and practically

Notations

- unknown secret key k^*, key byte hypothesis k
- D different samples, $d = 1, \ldots, D$
- Q different traces/ queries, $q = 1, \ldots, Q$
- matrix notation $M^{D,Q}$ (D rows, Q columns)
- leakage function φ
- sensitive variable: $\boxed{Y_q(k) = \varphi(T_q \oplus k)}$ (normalized variance $\forall q$)

Model



- trace $\boxed{X_{d,q} = \alpha_d Y_q(k^*) + N_{d,q}}$
- traces $X^{D,Q} = \alpha^D Y^Q(k^*) + N^{D,Q}$
- noise: zero-mean Gaussian distribution, covariance Σ
- independent of q but can be correlated among d

Overview

TELECOM ParisTech

Optimal distinguisher

Data processing theorem [Cover and Thomas, 2006]

Any preprocessing like dimensionality reduction can only decrease information.

- optimal means optimizing the success rate
- known leakage model: optimal attack $\Rightarrow$ template attack
- maximum likelihood principle

Optimal distinguisher

Data processing theorem [Cover and Thomas, 2006]

Any preprocessing like dimensionality reduction can only decrease information.

- optimal means optimizing the success rate
- known leakage model: optimal attack $\Rightarrow$ template attack
- maximum likelihood principle

- Given:
  - Q traces of dimensionality D in a matrix $x^{D,Q}$
  - for each trace x_q^D: a plaintext/ciphertext t_q

Optimal distinguisher

$$\begin{aligned}\mathcal{D}(x^{D,Q}, t^Q) &= \arg\max_k \; p(x^{D,Q}|t^Q, k^* = k) \\ &= \arg\max_k \; p_{N^{D,Q}}(x^{D,Q} - \alpha^D y^Q(k)) \\ &= \arg\max_k \; \prod_{q=1}^{Q} p_{N_q^D}(x_q^D - \alpha^D y_q(k))\end{aligned}$$

where

$$p_{N_q^D}(z^D) = \frac{1}{\sqrt{(2\pi)^D |\det \Sigma|}} \exp\Big(-\frac{1}{2}(z^D)^{\mathsf{T}} \Sigma^{-1} z^D\Big).$$

Optimal dimension reduction

Theorem

The optimal attack on the multivariate traces $x^{D,Q}$ is equivalent to the optimal attack on the monovariate traces $\tilde{x}^{Q}$, obtained from $x^{D,Q}$ by the formula:

$$\tilde{x}_q = \left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} x_q^D \qquad (q = 1, \ldots, Q).$$

Optimal dimension reduction

Theorem

The optimal attack on the multivariate traces $x^{D,Q}$ is equivalent to the optimal attack on the monovariate traces $\tilde{x}^{Q}$, obtained from $x^{D,Q}$ by the formula:

$$\tilde{x}_q = \left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} x_q^D \qquad (q = 1, \ldots, Q).$$

scalar $=$ column $D \cdot D \times D \cdot$ row D

Proof I

- taking the logarithm, the optimal distinguisher $\mathcal{D}(x^{D,Q}, t^Q)$ rewrites

$$\mathcal{D}(x^{D,Q}, t^Q) = \arg\min_k \sum_{q=1}^{Q} \left(x_q^D - \alpha^D y_q(k)\right)^{\mathsf{T}} \Sigma^{-1} \left(x_q^D - \alpha^D y_q(k)\right).$$

Proof I

- taking the logarithm, the optimal distinguisher $\mathcal{D}(x^{D,Q}, t^Q)$ rewrites

$$\mathcal{D}(x^{D,Q}, t^Q) = \arg\min_k \sum_{q=1}^{Q} \left(x_q^D - \alpha^D y_q(k)\right)^{\mathsf{T}} \Sigma^{-1} \left(x_q^D - \alpha^D y_q(k)\right).$$

- expansion gives

$$\underbrace{(x_q^D)^{\mathsf{T}} \Sigma^{-1} x_q^D}_{\text{cst. } C \text{ independent of } k} - 2(\alpha^D)^{\mathsf{T}} y_q(k) \Sigma^{-1} x_q^D + (y_q(k))^2 (\alpha^D)^{\mathsf{T}} \Sigma^{-1} \alpha^D$$

$$= C - 2y_q(k)\left[(\alpha^D)^{\mathsf{T}} \Sigma^{-1} x_q^D\right] + (y_q(k))^2 \left[(\alpha^D)^{\mathsf{T}} \Sigma^{-1} \alpha^D\right]$$

$$= \left[(\alpha^D)^{\mathsf{T}} \Sigma^{-1} \alpha^D\right] \left(y_q(k) - \frac{(\alpha^D)^{\mathsf{T}} \Sigma^{-1} x_q^D}{(\alpha^D)^{\mathsf{T}} \Sigma^{-1} \alpha^D}\right)^2 + C'.$$

Proof II

- so, for $\mathcal{D}(x^{D,Q}, t^Q)$ we obtain

$$\mathcal{D}(x^{D,Q}, t^Q) = \arg\min_k \sum_{q=1}^{Q} \left(y_q(k) - \frac{\left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} x_q^D}{\left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} \alpha^D} \right)^2 \left[\left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} \alpha^D \right]$$

$$= \arg\min_k \sum_{q=1}^{Q} \frac{\left(\tilde{x}_q - y_q(k)\right)^2}{\tilde{\sigma}^2},$$

where

$$\begin{cases} \tilde{x}_q & = \tilde{\sigma}^2 \cdot \left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} x_q^D, \\ \tilde{\sigma} & = \left(\left(\alpha^D\right)^{\mathsf{T}} \Sigma^{-1} \alpha^D \right)^{-1/2}. \end{cases}$$

Discussion

Optimal dimension reduction

Optimal distinguisher can be computed either:

- on multivariate traces x_q^D, with a noise covariance matrix Σ
- on monovariate traces $\tilde{x}_q$, with scalar noise of variance $\tilde{\sigma}^2$.

Discussion

Optimal dimension reduction

Optimal distinguisher can be computed either:

- on multivariate traces x_q^D, with a noise covariance matrix Σ
- on monovariate traces $\tilde{x}_q$, with scalar noise of variance $\tilde{\sigma}^2$.

- optimal dimensionality reduction does not depend on the distribution of $Y^D(k)$
- also not on the *confusion coefficient* [Fei et al., 2012]
- only on the signal weights α^D and on the noise covariance Σ

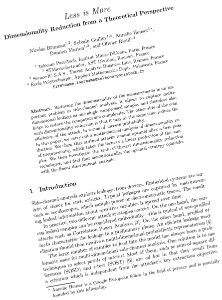
SNR

Corollary

After optimal dimensionality reduction, the signal-noise-ratio is given by

$$\frac{1}{\tilde{\sigma}^2} = (\alpha^D)^\mathsf{T} \Sigma^{-1} \alpha^D.$$

In the paper...



Less is More

Dimensionality Reduction from a Theoretical Perspective

Nicolas Bruneau[1,2], Sylvain Guilley[1,3], Annelie Heuser[1*], Damien Marion[1,3], and Olivier Rioul[1,4]

[1] Telecom ParisTech, Institut Mines-Télécom, Paris, France
[2] STMicroelectronics, AST Division, Rousset, France
[3] Secure-IC S.A.S., Threat Analysis Business Line, Rennes, France
[4] École Polytechnique, Applied Mathematics Dept., Palaiseau, France
firstname.lastname@telecom-paristech.fr

Abstract. Reducing the dimensionality of the measurements is an important problem in side-channel analysis. It allows to capture multi-dimensional leakage as one single compressed sample, and therefore also helps to reduce the computational complexity. The other side of the coin with dimensionality reduction is that it may at the same time reduce the efficiency of the attack, in terms of success probability. In this paper, we carry out a mathematical analysis of dimensionality reduction. We show that optimal attacks remain optimal after a first pass of preprocessing, which takes the form of a linear projection of the samples. We then investigate the state-of-the-art dimensionality reduction techniques, and find that asymptotically, the optimal strategy coincides with the linear discriminant analysis.

1 Introduction

Side-channel analysis exploits leakages from devices. Embedded systems are targets of choice for such attacks. Typical leakages are captured by instruments such as oscilloscopes, which sample power or electromagnetic traces. The resulting leaked information about sensitive variables is spread over time.

In practice, two different attack strategies coexist. On the one hand, the various leaked samples can be considered individually—this is typical of *non-profiled attacks* such as Correlation Power Analysis [2]. On the other hand, *profiled attacks* characterize the leakage in a preliminary phase. An efficient leakage modelization should then involve a multi-dimensional probabilistic representation [4].

The large number of samples to feed into the model has always been a problematic issue for multi-dimensional side-channel analysis. One solution is to use techniques to select *points of interest*. Most of them, such as sum-of-square differences (SOSD) and t-test (SOST) [9], are *ad hoc* in that they result from a criterion which is independent from the attacker's key extraction objective.

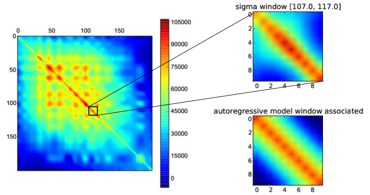
* Annelie Heuser is a Google European fellow in the field of privacy and is partially founded by this fellowship.

Examples

- white noise:

$$\boxed{\widetilde{\mathrm{SNR}} = \sum_{d=1}^{D} \mathrm{SNR}_d}$$

- autoregressive noise (confirmed on dpacontest v2)



TELECOM ParisTech

Overview

Comparison to PCA

Classical PCA

- centered data $M_{d,q} = X_{d,q} - \frac{1}{Q}\sum_{q'=1}^{Q} X_{d,q'}$ $(1 \leq q \leq Q, 1 \leq d \leq D)$
- directions of PCA: eigenvectors of $M^{D,Q}(M^{D,Q})^\mathsf{T}$
- drawback: depends both on data and noise

Comparison to PCA

Classical PCA

- centered data $M_{d,q} = X_{d,q} - \frac{1}{Q}\sum_{q'=1}^{Q} X_{d,q'}$ $(1 \leq q \leq Q, 1 \leq d \leq D)$
- directions of PCA: eigenvectors of $M^{D,Q}(M^{D,Q})^{\mathsf{T}}$
- drawback: depends both on data and noise

Inter-class PCA [Archambeau et al., 2006]

- centered column $\frac{1}{\sum_{\substack{1 \leq q \leq Q \\ Y_q = y}} 1} \sum_{\substack{1 \leq q \leq Q \\ Y_q = y}} X_q^D$
- takes into account the sensitive variable Y
- noise is averaged away

Comparison to PCA

For classical PCA

Asymptotically as $Q \longrightarrow +\infty$,

$$\frac{1}{Q} M^{D,Q} (M^{D,Q})^{\mathsf{T}} \longrightarrow \alpha^{D} (\alpha^{D})^{\mathsf{T}} + \Sigma.$$

Eigenvectors?

Comparison to PCA

For classical PCA

Asymptotically as $Q \longrightarrow +\infty$,

$$\frac{1}{Q} M^{D,Q} (M^{D,Q})^{\mathsf{T}} \longrightarrow \alpha^D (\alpha^D)^{\mathsf{T}} + \Sigma.$$

Eigenvectors?

Proposition

Asymptotically, Inter-class PCA has only one principal direction, namely the vector α^D.

Comparison to PCA

Proposition

The asymptotic SNR after projection using Inter-class PCA is equal to $\frac{\left\|\alpha^D\right\|_2^4}{(\alpha^D)^\mathsf{T}\Sigma\alpha^D}$.

Comparison to PCA

Proposition

The asymptotic SNR after projection using Inter-class PCA is equal to $\frac{\left\|\alpha^D\right\|_2^4}{(\alpha^D)^\mathsf{T}\Sigma\alpha^D}$.

Theorem

The SNR of the asymptotic Inter-class PCA is smaller than the SNR of the optimal dimensionality reduction.

Comparison to PCA

Proposition

The asymptotic SNR after projection using Inter-class PCA is equal to $\frac{\left\|\alpha^D\right\|_2^4}{(\alpha^D)^\mathsf{T}\Sigma\alpha^D}$.

Theorem

The SNR of the asymptotic Inter-class PCA is smaller than the SNR of the optimal dimensionality reduction.

Corollary

The asymptotic Inter-class PCA has the same SNR as the optimal dimensionality reduction if and only if α^D *is an eigenvector of* Σ. *In this case, both dimensionality reductions are equivalent.*

Comparison to LDA

- computes the eigenvectors of $S_w^{-1} S_b$
- S_w is the *intra-class scatter matrix*, asymptotically equal to Σ
- S_b is the *inter-class scatter matrix*, equal to $\alpha^D (\alpha^D)^\mathsf{T}$.

Proposition

Asymptotically, LDA has only one principal direction, namely the vector $\Sigma^{-1} \alpha^D$.

Comparison to LDA

- computes the eigenvectors of $S_w^{-1} S_b$
- S_w is the *intra-class scatter matrix*, asymptotically equal to Σ
- S_b is the *inter-class scatter matrix*, equal to $\alpha^D (\alpha^D)^\mathsf{T}$.

Proposition

Asymptotically, LDA has only one principal direction, namely the vector $\Sigma^{-1}\alpha^D$.

Theorem

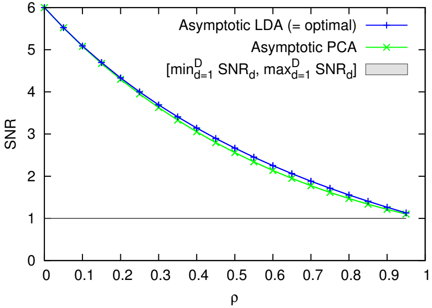
The asymptotic LDA computes exactly the optimal dimensionality reduction.

Asymptotic PCA and LDA

- $D = 6$ for autoregressive noise with $\sigma = 1$ and different ρ

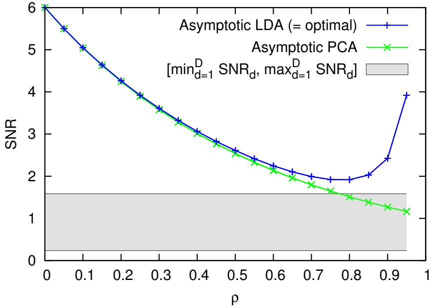
(a) Equal SNR$_d = 1$, $1 \leq d \leq D$

$$\alpha^D = (1, 1, 1, 1, 1, 1)^\mathsf{T}$$



(b) Varying SNR$_d$, $1 \leq d \leq D$

$$\alpha^D = \sqrt{6.0/6.4} \cdot (1.0, 1.1, 1.2, 1.3, 0.9, 0.5)^\mathsf{T}$$



Overview

Practical Validation

- DPA CONTEST V2, one clock cycle $D = 200$
- normalized Hamming weight
- precharacterization of the model parameter α^D and Σ (details in the paper)

- $\max_{d=1}^{D} \hat{\alpha}_d^2 / \hat{\Sigma}_{d,d} = 1.69 \cdot 10^{-3}$ (no dimensionality reduction)
- $\mathrm{SNR}_{\mathrm{PCA}} = \frac{((\hat{\alpha}^D)^{\mathsf{T}} \hat{\alpha}^D)^2}{(\hat{\alpha}^D)^{\mathsf{T}} \hat{\Sigma} \hat{\alpha}^D} = 1.36 \cdot 10^{-3}$ (PCA)
- $\mathrm{SNR}_{\mathrm{LDA}} = (\hat{\alpha}^D)^{\mathsf{T}} \hat{\Sigma} \hat{\alpha}^D = 12.78 \cdot 10^{-3}$ (LDA)

Overview

Conclusion and Perspectives

Optimal dimension reduction...

- is part of the optimal attack
- can be achieved *without* losing success probability

Conclusion and Perspectives

Optimal dimension reduction...

- is part of the optimal attack
- can be achieved *without* losing success probability

- LDA asymptotically achieves the same projection as optimal
- when weakly correlated (Σ is identity matrix) PCA is nearly equivalent to optimal/ LDA

Conclusion and Perspectives

Optimal dimension reduction...

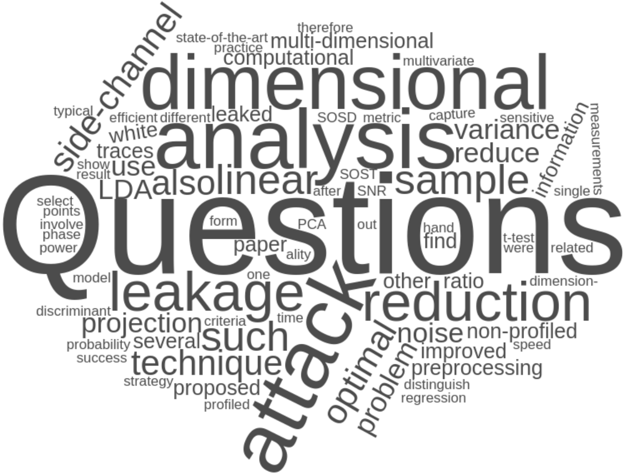
- is part of the optimal attack
- can be achieved *without* losing success probability

- LDA asymptotically achieves the same projection as optimal
- when weakly correlated (Σ is identity matrix) PCA is nearly equivalent to optimal/ LDA

⋆ extend to non-Gaussian noise

⋆ comparison to machine-learning techniques

Thank you!



TELECOM ParisTech

References I

[Archambeau et al., 2006] Archambeau, C., Peeters, É., Standaert, F.-X., and Quisquater, J.-J. (2006).
Template Attacks in Principal Subspaces.
In *CHES*, volume 4249 of *LNCS*, pages 1–14. Springer.
Yokohama, Japan.

[Batina et al., 2012] Batina, L., Hogenboom, J., and van Woudenberg, J. G. J. (2012).
Getting more from PCA: first results of using principal component analysis for extensive power analysis.
In Dunkelman, O., editor, *Topics in Cryptology - CT-RSA 2012 - The Cryptographers' Track at the RSA Conference 2012, San Francisco, CA, USA, February 27 - March 2, 2012. Proceedings*, volume 7178 of *Lecture Notes in Computer Science*, pages 383–397. Springer.

References II

[Bhasin et al., 2014] Bhasin, S., Danger, J.-L., Guilley, S., and Najm, Z. (2014).
Side-channel Leakage and Trace Compression Using Normalized Inter-class Variance.
In *Proceedings of the Third Workshop on Hardware and Architectural Support for Security and Privacy*, HASP '14, pages 7:1–7:9, New York, NY, USA. ACM.

[Choudary and Kuhn, 2013] Choudary, O. and Kuhn, M. G. (2013).
Efficient template attacks.
In Francillon, A. and Rohatgi, P., editors, *Smart Card Research and Advanced Applications - 12th International Conference, CARDIS 2013, Berlin, Germany, November 27-29, 2013. Revised Selected Papers*, volume 8419 of *LNCS*, pages 253–270. Springer.

References III

[Cover and Thomas, 2006] Cover, T. M. and Thomas, J. A. (2006).
Elements of Information Theory.
Wiley-Interscience.
ISBN-10: ISBN-10: 0471241954, ISBN-13: 978-0471241959, 2nd edition.

[Danger et al., 2014] Danger, J.-L., Debande, N., Guilley, S., and Souissi, Y. (2014).
High-order timing attacks.
In *Proceedings of the First Workshop on Cryptography and Security in Computing Systems*, CS2 '14, pages 7–12, New York, NY, USA. ACM.

[Debande et al., 2012] Debande, N., Souissi, Y., Elaabid, M. A., Guilley, S., and Danger, J. (2012).
Wavelet transform based pre-processing for side channel analysis.
In *45th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO 2012, Workshops Proceedings, Vancouver, BC, Canada, December 1-5, 2012*, pages 32–38. IEEE Computer Society.

TELECOM ParisTech

References IV

[Fei et al., 2012] Fei, Y., Luo, Q., and Ding, A. A. (2012).
A Statistical Model for DPA with Novel Algorithmic Confusion Analysis.
In Prouff, E. and Schaumont, P., editors, *CHES*, volume 7428 of *LNCS*, pages 233–250. Springer.

[Gierlichs et al., 2006] Gierlichs, B., Lemke-Rust, K., and Paar, C. (2006).
Templates vs. Stochastic Methods.
In *CHES*, volume 4249 of *LNCS*, pages 15–29. Springer.
Yokohama, Japan.

[Guilley et al., 2008] Guilley, S., Chaudhuri, S., Sauvage, L., Hoogvorst, P., Pacalet, R., and Bertoni, G. M. (2008).
Security Evaluation of WDDL and SecLib Countermeasures against Power Attacks.
IEEE Transactions on Computers, 57(11):1482–1497.

References V

[Oswald et al., 2006] Oswald, E., Mangard, S., Herbst, C., and Tillich, S. (2006).
Practical Second-Order DPA Attacks for Masked Smart Card Implementations of Block Ciphers.
In Pointcheval, D., editor, *CT-RSA*, volume 3860 of *LNCS*, pages 192–207. Springer.

[Renauld et al., 2011] Renauld, M., Standaert, F., Veyrat-Charvillon, N., Kamel, D., and Flandre, D. (2011).
A formal study of power variability issues and side-channel attacks for nanoscale devices.
In Paterson, K. G., editor, *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, volume 6632 of *Lecture Notes in Computer Science*, pages 109–128. Springer.

References VI

[Souissi et al., 2010] Souissi, Y., Nassar, M., Guilley, S., Danger, J.-L., and Flament, F. (2010).

First Principal Components Analysis: A New Side Channel Distinguisher.

In Rhee, K. H. and Nyang, D., editors, *ICISC*, volume 6829 of *Lecture Notes in Computer Science*, pages 407–419. Springer.

[Standaert and Archambeau, 2008] Standaert, F.-X. and Archambeau, C. (2008).

Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages.

In *CHES*, volume 5154 of *Lecture Notes in Computer Science*, pages 411–425. Springer.

Washington, D.C., USA.

References VII

[Strobel et al., 2014] Strobel, D., Oswald, D., Richter, B., Schellenberg, F., and Paar, C. (2014).
Microcontrollers as (in)security devices for pervasive computing applications.
Proceedings of the IEEE, 102(8):1157–1173.