PROFILING DPA: EFFICIENCY AND EFFICACY TRADE-OFFS

Carolyn Whitnall, Elisabeth Oswald

carolyn.whitnall@bris.ac.uk
Department of Computer Science, University of Bristol

21st August 2013

OUTLINE

- What is profiled DPA? – an overview of the popular methods
- What makes a good power model? – our evaluation criteria
- How 'good' is good enough? – analysis of some example scenarios

SIDECHANNELANALALALYSIS*

* (By way of ‘wittily’ acknowledging my frequent pronunciation fails...)

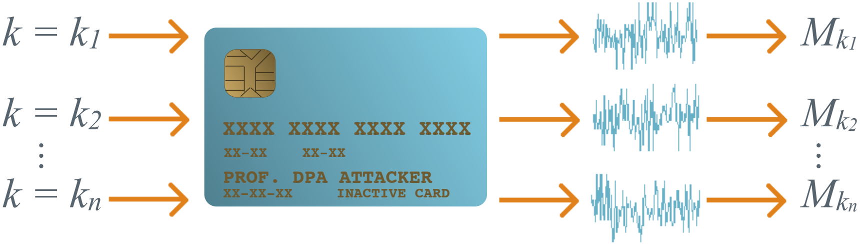
Profiled DPA

Profiling phase (supervised learning)

Attack phase (classification)

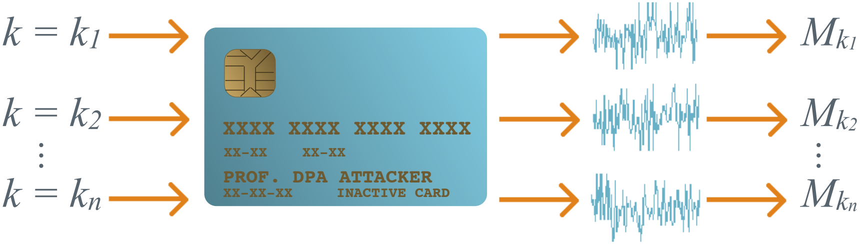
PROFILED DPA

PROFILING PHASE (SUPERVISED LEARNING)

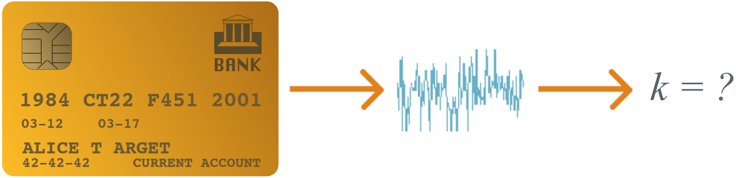


ATTACK PHASE (CLASSIFICATION)

PROFILING PHASE (SUPERVISED LEARNING)



ATTACK PHASE (CLASSIFICATION)



Two Typical Methods

'Classical' templates:

- Separate multivariate Gaussian models for each key-dependent value
- Covariance matrix estimated for each key-dependent value

Linear regression-based templates:

- Linear regression model fitted to the pooled data at each time point
- Covariance matrix estimated for pooled data (2nd, independent sample)

Choose the key hypothesis which maximises the log-likelihood of the observed traces.

OR (ignoring noise):

Choose the key hypothesis which maximises the correlation between the model fitted values and the observed traces.

Two Typical Methods

'Classical' templates:

- Separate multivariate Gaussian models for each key-dependent value
- Covariance matrix estimated for each key-dependent value

Linear regression-based templates:

- Linear regression model fitted to the pooled data at each time point
- Covariance matrix estimated for pooled data (2$^{\text{nd}}$, independent sample)

Choose the key hypothesis which maximises the log-likelihood of the observed traces.

OR (ignoring noise):

Choose the key hypothesis which maximises the correlation between the model fitted values and the observed traces.

Two Typical Methods

'Classical' templates:

- Separate multivariate Gaussian models for each key-dependent value
- Covariance matrix estimated for each key-dependent value

Linear regression-based templates:

- Linear regression model fitted to the pooled data at each time point
- Covariance matrix estimated for pooled data (2nd, independent sample)

Choose the key hypothesis which maximises the log-likelihood of the observed traces.

OR (ignoring noise):

Choose the key hypothesis which maximises the correlation between the model fitted values and the observed traces.

Two Typical Methods

'Classical' templates:

- Separate multivariate Gaussian models for each key-dependent value
- Covariance matrix estimated for each key-dependent value

Linear regression-based templates:

- Linear regression model fitted to the pooled data at each time point
- Covariance matrix estimated for pooled data (2^{nd}, independent sample)

Choose the key hypothesis which maximises the log-likelihood of the observed traces.

OR (ignoring noise):

Choose the key hypothesis which maximises the correlation between the model fitted values and the observed traces.

Two Typical Methods

'Classical' templates:

- Separate multivariate Gaussian models for each key-dependent value
- Covariance matrix estimated for each key-dependent value

Linear regression-based templates:

- Linear regression model fitted to the pooled data at each time point
- Covariance matrix estimated for pooled data (2^{nd}, independent sample)

Choose the key hypothesis which maximises the log-likelihood of the observed traces.

OR (ignoring noise):

Choose the key hypothesis which maximises the correlation between the model fitted values and the observed traces.

Linear Regression vs. Classical

Consider an 8-bit intermediate value target (e.g. AES S-box output)...

- Classical templates have *fixed complexity*: 2^m conditional mean vectors, 2^m covariance matrices.
- Linear regression has *adjustable complexity*: an intercept, coefficients on all the equation terms, and one covariance matrix.
  - Potentially large reduction in profiling traces needed (e.g. linear model expression requires only $m + 1$ coefficients).
  - Potentially substantial degradation in model quality if simplifying assumptions are not correct.
  - Higher-order terms in the model equation militate against model degradation but add to profiling data complexity.
- Linear regression models *coincide* with classical (in complexity and quality of deterministic part) once all possible monomial terms are included in the equation.

Linear Regression vs. Classical

Consider an 8-bit intermediate value target (e.g. AES S-box output)...

- Classical templates have *fixed complexity*: 2^m conditional mean vectors, 2^m covariance matrices.

- Linear regression has *adjustable complexity*: an intercept, coefficients on all the equation terms, and one covariance matrix.
  - Potentially large reduction in profiling traces needed (e.g. linear model expression requires only $m + 1$ coefficients).
  - Potentially substantial degradation in model quality if simplifying assumptions are not correct.
  - Higher-order terms in the model equation militate against model degradation but add to profiling data complexity.

- Linear regression models *coincide* with classical (in complexity and quality of deterministic part) once all possible monomial terms are included in the equation.

Consider an 8-bit intermediate value target (e.g. AES S-box output)...

- Classical templates have *fixed complexity*: 2^m conditional mean vectors, 2^m covariance matrices.

- Linear regression has *adjustable complexity*: an intercept, coefficients on all the equation terms, and one covariance matrix.
  - Potentially large reduction in profiling traces needed (e.g. linear model expression requires only $m + 1$ coefficients).
  - Potentially substantial degradation in model quality if simplifying assumptions are not correct.
  - Higher-order terms in the model equation militate against model degradation but add to profiling data complexity.

- Linear regression models *coincide* with classical (in complexity and quality of deterministic part) once all possible monomial terms are included in the equation.

Linear Regression vs. Classical

Consider an 8-bit intermediate value target (e.g. AES S-box output)…

- Classical templates have *fixed complexity*: 2^m conditional mean vectors, 2^m covariance matrices.

- Linear regression has *adjustable complexity*: an intercept, coefficients on all the equation terms, and one covariance matrix.
  - Potentially large reduction in profiling traces needed (e.g. linear model expression requires only $m + 1$ coefficients).
  - Potentially substantial degradation in model quality if simplifying assumptions are not correct.
  - Higher-order terms in the model equation militate against model degradation but add to profiling data complexity.

- Linear regression models *coincide* with classical (in complexity and quality of deterministic part) once all possible monomial terms are included in the equation.

Previous Work

Templates vs. Stochastic Methods, B. Gierlichs, K. Lemke-Rust, C. Paar. *CHES 2006, LNCS 4249: 15–29, Springer.*

- LR templates recover key with fewer (profiling) traces but classical achieve higher success rates once profiling sample is large.
- Analysis primarily experimental: true distributions unknown so difficult to comment on model quality.
- Tested scenarios limited and favourable to LR (close to HW).

How to Compare Profiled Side-Channel Attacks?, F.X. Standaert, F. Koeune, W. Schindler. *ACNS 2009, LNCS 5536: 485–498, Springer.*

- Information theoretic metric can be used to quantify model quality.
- Analysis geared more towards theory (establishing an evaluation framework).
- Tested scenarios limited to simulated HW leakage – LR has big advantage; comparative findings do not extend to general case.

Previous Work

Templates vs. Stochastic Methods, B. Gierlichs, K. Lemke-Rust, C. Paar. *CHES 2006, LNCS 4249: 15–29, Springer.*

- LR templates recover key with fewer (profiling) traces but classical achieve higher success rates once profiling sample is large.
- Analysis primarily experimental: true distributions unknown so difficult to comment on model quality.
- Tested scenarios limited and favourable to LR (close to HW).

How to Compare Profiled Side-Channel Attacks?, F.X. Standaert, F. Koeune, W. Schindler. *ACNS 2009, LNCS 5536: 485–498, Springer.*

- Information theoretic metric can be used to quantify model quality.
- Analysis geared more towards theory (establishing an evaluation framework).
- Tested scenarios limited to simulated HW leakage – LR has big advantage; comparative findings do not extend to general case.

Previous Work

Templates vs. Stochastic Methods, B. Gierlichs, K. Lemke-Rust, C. Paar. *CHES 2006, LNCS 4249: 15–29, Springer.*

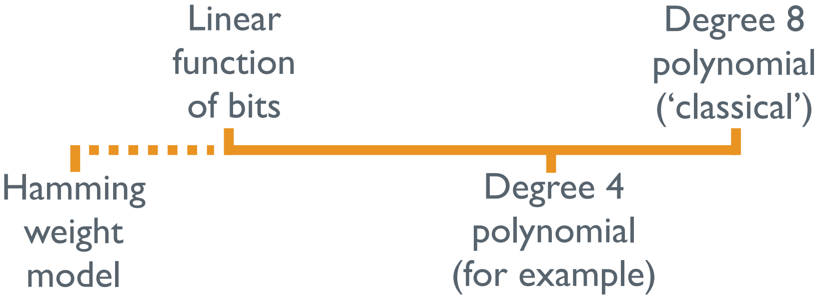
- LR templates recover key with fewer (profiling) traces but classical achieve higher success rates once profiling sample is large.
- Analysis primarily experimental: true distributions unknown so difficult to comment on model quality.
- Tested scenarios limited and favourable to LR (close to HW).

How to Compare Profiled Side-Channel Attacks?, F.X. Standaert, F. Koeune, W. Schindler. *ACNS 2009, LNCS 5536: 485–498, Springer.*

- Information theoretic metric can be used to quantify model quality.
- Analysis geared more towards theory (establishing an evaluation framework).
- Tested scenarios limited to simulated HW leakage – LR has big advantage; comparative findings do not extend to general case.

OUR CONTRIBUTION

- Explore trade-offs in a *wider range of scenarios*, including those *not* well-suited to low-degree approximations.
- *Theoretic* (rather than experimental) evaluation where possible.
- Hypothetical scenarios with *fully-specified leakage distributions* give concrete benchmarks for model quality/performance.



What Makes a Good Power Model?

1. Profiling complexity: the fewer traces needed to build the model, the better.
2. Goodness-of-fit: the closer the model is to the actual leakage distribution, the better.
3. DPA performance: the fewer the traces needed to recover the key from the target device, the better.

Measuring Profiling Complexity

- Difficult to measure theoretically: sample size formulae exist for simpler statistical problems but not for precise coefficient estimation.
- Empirical approach:
  - 1,000 repeat experiments on randomly drawn balanced samples
  - Gaussian noise at high (8) medium (1) and low (0.125) signal-to-noise ratios
  - Fit models of degree ranging from 1 through to 8
  - Count number of traces required to reach a certain threshold of precision

MEASURING GOODNESS-OF-FIT

- Find least squares solution $\{\hat{\beta}_0, \ldots, \hat{\beta}_p\}$ for the system of equations representing the regression in the absence of noise:

$$\{Y_v\}_{v \in \mathcal{V}} = \left\{ \sum_{j=0}^{p} \beta_j g_j(v) \right\}_{v \in \mathcal{V}}$$

- Compute *coefficient of determination* – proportion of variation in the leakage function which is accounted for by the model:

$$\rho\left(\left\{ \sum_{j=0}^{p} \hat{\beta}_j g_j(v) \right\}_{v \in \mathcal{V}}, \{Y_v\}_{v \in \mathcal{V}} \right)^2$$

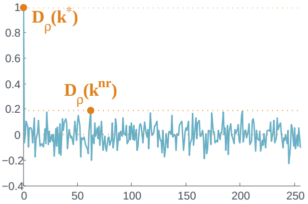
Model fitted values

Actual leakage

- Compute the theoretic correlation distinguishing vector under each model:

$$D_\rho(k) = \rho(Y, M_{LR}(V_k)) = \frac{\operatorname{cov}(Y, M_{LR}(V_k))}{\sqrt{\operatorname{var}(Y)}\sqrt{\operatorname{var}(M_{LR}(V_k))}}$$

- Use sample size formulae to calculate the number of traces required to distinguish the true key from the nearest rival:



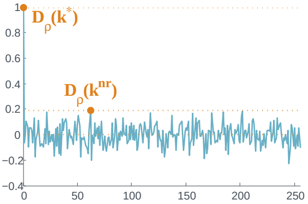
$$N^* = 3 + 8 \cdot \frac{z^2_{1-\alpha}}{\left(\ln \frac{1+D_\rho(k^*)}{1-D_\rho(k^*)} - \ln \frac{1+D_\rho(k^{nr})}{1-D_\rho(k^{nr})}\right)^2}$$

Measuring DPA Performance

- Compute the theoretic correlation distinguishing vector under each model:

$$D_\rho(k) = \rho(Y, M_{LR}(V_k)) = \frac{\mathrm{cov}(Y, M_{LR}(V_k))}{\sqrt{\mathrm{var}(Y)}\sqrt{\mathrm{var}(M_{LR}(V_k))}}$$

- Use sample size formulae to calculate the number of traces required to distinguish the true key from the nearest rival:



Quantile of the standard normal $N(0,1)$

$$N^* = 3 + 8 \cdot \frac{\boxed{z^2_{1-\alpha}}}{\left(\ln \frac{1+D_\rho(k^*)}{1-D_\rho(k^*)} - \ln \frac{1+D_\rho(k^{\mathrm{nr}})}{1-D_\rho(k^{\mathrm{nr}})}\right)^2}$$

α: "significance level"

Some Example Scenarios

Consider leakage of the form $L(v) + \varepsilon$, where $L(v)$ is the deterministic, data-dependent component which we will call the *leakage function* and $\varepsilon \sim N(0, \sigma_\varepsilon)$ is *additive Gaussian noise*. (The intermediate value v in our analysis is the AES S-box output.)

1. The leakage function is proportional to the *Hamming weight*, as motivated by typical behaviour of CMOS technology.
2. Adjacent wires interact so that the leakage is proportional to the *Hamming weight plus quadratic terms* involving adjacent bits of the intermediate value.
3. The leakage is a *highly nonlinear* function of the intermediate bits such as that arising from hardware implementations of AES.

SOME EXAMPLE SCENARIOS

Consider leakage of the form $L(v) + \varepsilon$, where $L(v)$ is the deterministic, data-dependent component which we will call the *leakage function* and $\varepsilon \sim N(0, \sigma_\varepsilon)$ is *additive Gaussian noise*. (The intermediate value v in our analysis is the AES S-box output.)

1. The leakage function is proportional to the *Hamming weight*, as motivated by typical behaviour of CMOS technology.
2. Adjacent wires interact so that the leakage is proportional to the *Hamming weight plus quadratic terms* involving adjacent bits of the intermediate value.
3. The leakage is a *highly nonlinear* function of the intermediate bits such as that arising from hardware implementations of AES.

Some Example Scenarios

Consider leakage of the form $L(v) + \varepsilon$, where $L(v)$ is the deterministic, data-dependent component which we will call the *leakage function* and $\varepsilon \sim N(0, \sigma_\varepsilon)$ is *additive Gaussian noise*. (The intermediate value v in our analysis is the AES S-box output.)

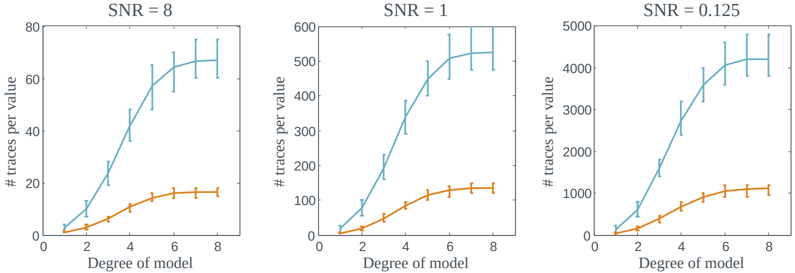
1. The leakage function is proportional to the *Hamming weight*, as motivated by typical behaviour of CMOS technology.
2. Adjacent wires interact so that the leakage is proportional to the *Hamming weight plus quadratic terms* involving adjacent bits of the intermediate value.
3. The leakage is a *highly nonlinear* function of the intermediate bits such as that arising from hardware implementations of AES.

Some Example Scenarios

Consider leakage of the form $L(v) + \varepsilon$, where $L(v)$ is the deterministic, data-dependent component which we will call the *leakage function* and $\varepsilon \sim N(0, \sigma_\varepsilon)$ is *additive Gaussian noise*. (The intermediate value v in our analysis is the AES S-box output.)

1. The leakage function is proportional to the *Hamming weight*, as motivated by typical behaviour of CMOS technology.
2. Adjacent wires interact so that the leakage is proportional to the *Hamming weight plus quadratic terms* involving adjacent bits of the intermediate value.
3. The leakage is a *highly nonlinear* function of the intermediate bits such as that arising from hardware implementations of AES.

Profiling Complexity



- Affects all leakage scenarios similarly.
- Sample sizes to estimate maximum degree polynomials are around 30 times more than those to estimate linear polynomials.
- Little change in complexity between degree 6 and degree 8 models.
- Reasonable savings only possible at degree 5 or lower.
- Sample size increases as signal decreases but relationship between models of different degree is consistent.

Low Degree Leakages

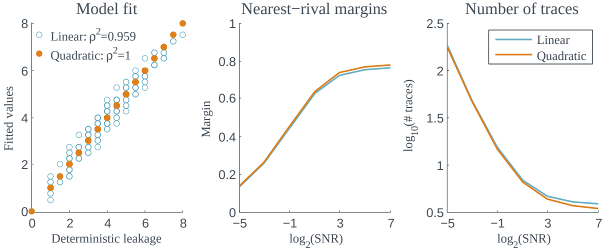
Hamming weight leakage:

- Perfectly approximated by a linear model function.
- Performs equivalently to ‘classical’ models.

Hamming weight leakage:

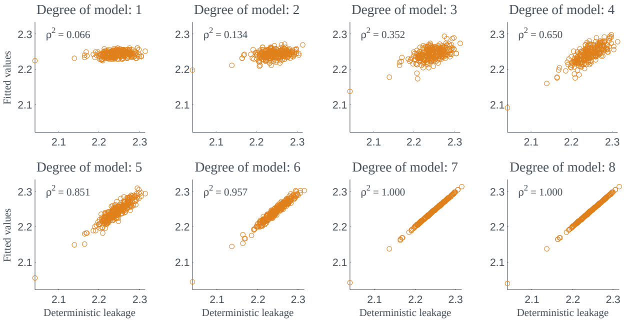
- Perfectly approximated by a linear model function.
- Performs equivalently to ‘classical’ models.

Leakage with adjacent interactions:



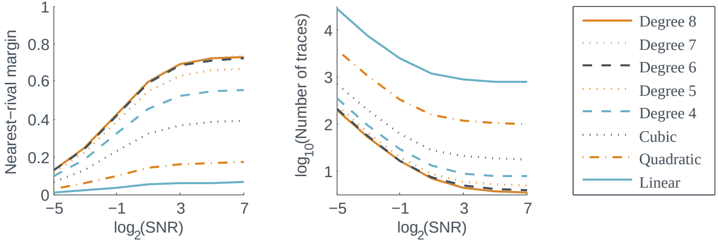
- Closely approximated by a linear model function.
- Performance only marginally diminished.

Toggle Count-Based Leakage:



- Linear model inadequate to approximate the leakage – captures just 6% of the variation.
- Degree 4 model accounts for about two thirds of the variation, with less than half the number of parameters required for the classical model.

Toggle Count-Based Leakage:



- Very little difference in distinguishing power between the degree 5 and classical models.
- Linear and quadratic models are able to recover the key, but by very small margins and requiring lots of traces – over a hundred times as many in the case of the linear model.
- Degree 4 model requires around twice as many traces.

Summary Table

			Adjacent interactions		Toggle count-based	
Model	**#Params**	**Profiling complexity**	**Model fit**	**Attack complexity**	**Model fit**	**Attack complexity**
HW	–	0	0.88	1.2–1.3	0.04	930–1,270
Deg. 1	9	0.03	0.96	1.0–1.1	0.06	136–220
Deg. 2	37	0.13	1	1	0.13	19–29
Deg. 3	93	0.33	1	1	0.35	3.6–5.2
Deg. 4	163	0.63	1	1	0.65	1.7–2.2
Deg. 5	219	0.83	1	1	0.85	1.2–1.4
Deg. 6	247	0.90	1	1	0.96	1.0–1.1
Deg. 7	255	1	1	1	1	1
Deg. 8	256	1	1	1	1	1

Summary Table

Experiments suggest the formula overstates the sample size in the case of highly-degraded models (further work needed).

			Adjacent interactions		Toggle count-based	
Model	**#Params**	**Profiling complexity**	**Model fit**	**Attack complexity**	**Model fit**	**Attack complexity**
HW	–	0	0.88	1.2–1.3	0.04	930–1,270
Deg. 1	9	0.03	0.96	1.0–1.1	0.06	136–220
Deg. 2	37	0.13	1	1	0.13	19–29
Deg. 3	93	0.33	1	1	0.35	3.6–5.2
Deg. 4	163	0.63	1	1	0.65	1.7–2.2
Deg. 5	219	0.83	1	1	0.85	1.2–1.4
Deg. 6	247	0.90	1	1	0.96	1.0–1.1
Deg. 7	255	1	1	1	1	1
Deg. 8	256	1	1	1	1	1

Conclusion

- Linear regression is an excellent alternative to classical profiling when the true leakage function is simple.
- Over-simplified assumptions when the leakage is complex can substantially diminish attack performance.
- Device evaluation perspective:
  - Classical profiling remains the best way to test for vulnerability against the strongest possible adversary.
- Attacker perspective:
  - In our example, degree 4 models offer a promising trade-off between profiling and attack complexity.
  - Even minimal profiling can substantially *increase* attack performance relative to standard assumptions (such as Hamming weight leakage) when those assumptions do not hold.

Conclusion

- Linear regression is an excellent alternative to classical profiling when the true leakage function is simple.
- Over-simplified assumptions when the leakage is complex can substantially diminish attack performance.
- Device evaluation perspective:
  - Classical profiling remains the best way to test for vulnerability against the strongest possible adversary.
- Attacker perspective:
  - In our example, degree 4 models offer a promising trade-off between profiling and attack complexity.
  - Even minimal profiling can substantially *increase* attack performance relative to standard assumptions (such as Hamming weight leakage) when those assumptions do not hold.

Conclusion

- Linear regression is an excellent alternative to classical profiling when the true leakage function is simple.
- Over-simplified assumptions when the leakage is complex can substantially diminish attack performance.
- Device evaluation perspective:
  - Classical profiling remains the best way to test for vulnerability against the strongest possible adversary.
- Attacker perspective:
  - In our example, degree 4 models offer a promising trade-off between profiling and attack complexity.
  - Even minimal profiling can substantially *increase* attack performance relative to standard assumptions (such as Hamming weight leakage) when those assumptions do not hold.

Thank you for listening!

Any questions?