Side-Channel Analysis: Combining (Quantitative) Statistical Analysis with Engineer's (Qualitative) Intuition

Werner Schindler

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Godesberger Allee 185–189
53175 Bonn, Germany
Werner.Schindler@bsi.bund.de

Keywords: Side-channel analysis, mathematical statistics.

Extended Abstract

In the last decade side-channel cryptanalysis (timing attacks, power attacks, electromagnetic radiation attacks, cache-based attacks etc.) has become an important branch in cryptology. The goal of any side-channel attack is to extract key-dependent information from the leakage signal. For simplicity, we focus on power attacks in the following although our considerations could be adapted to radiation attacks.

The 'classical' approach are DPA attacks, which require only little set-up work. On the negative side their efficiency is limited since they are usually based on a simple but restrictive model (e.g. Hamming weight model), which the attacked device often does not fulfil. A further problem is how to weight the side channel information gained at several instants.

Template attacks interpret power measurements as values taken on by random variables. 'Classical' template attacks do not employ any model assumptions besides that the power signals measured at time instants $t_1 < t_2 < \ldots < t_m$ are jointly normally distributed. In the following x denotes the relevant part of the plaintext (e.g. a byte that affects a particular S-box), z the masking value and k the relevant subkey. To estimate the density $f_{x,z,k}(\cdot)$ of the m-dimensional random vector

$$(I_{t_1}(x,z,k), \ldots, I_{t_m}(x,z,k)). \qquad (1)$$

in the profiling phase (aka characterization phase) the adversary performs power measurements with an identical training device for each triple $(x,z,k) \in \{0,1\}^p \times \mathcal{M} \times \{0,1\}^s$. Since normal distributions are uniquely determined by their mean vector and their covariance matrix it suffices to estimate these values. In the attacking phase (aka key extraction phase) the adversary performs measurements at the target device. In absence of masking he simply substitutes these measurement vectors into (products of) the empirical densities derived in the profiling phase and decides for the that subkey which yields the maximum value (maximum likelihood principle). In the presence of masking the adversary instead substitutes these values into a 'density mixture' $\bar{f}_{x,k}(\cdot) := \sum_z f_{x,z,k}(\cdot)/|\mathcal{M}|$.

The attacking efficiency of template attacks is optimal. The bottleneck yet is the profiling phase. The presence of masking techniques requires gigantic workload, in particular for strong hardware.

In our approach we use the natural (and hardly restrictive) assumption that

$$I_{t_j}(x,z,k) = h_{t_j}(x,z,k) + R_{t_j} \qquad (2)$$

where $h_{t_j}(x,z,k)$ quantifies the deterministic, subkey-dependent part of the leakage while R_{t_j} represents the (subkey-independent) noise at time t_j. Profiling falls

into two parts. At first we estimate h_{t_j} separately for all instants t_j (which does not lose any information!). In a second step we estimate the covariance matrix of the noise vector $(R_{t_1}, \ldots, R_{t_m})$. Key extraction works as for template attacks.

Note that estimating $h_{t_j}(x, z, k)$ for all triples (x, z, k) essentially yields the template attack. Instead, we only perform 2^s estimation processes, one for each admissible subkey k. For fixed subkey k the function $h_{t_j}(\cdot, \cdot, k)$ may be interpreted as an element in an $2^p|\mathcal{M}|$-dimensional subspace $\mathcal{F}$. In place of the exact function h_{t_j} we aim at its image $h^*_{t_j}$ under an orthogonal projection onto a low-dimensional subspace $\mathcal{V}_{t_j}$. The clou is that $h^*_{t_j}$ minimizes the expectation of a particular function on $\mathcal{V}_{t_j}$. Hence $h^*_{t_j}$ can be determined without knowing its pre-image h_{t_j}, moving statistics from the high-dimensional vector space $\mathcal{F}$ to the low-dimensional subspace $\mathcal{V}_{t_j}$, and thus reducing the number of measurements in the profiling phase to a small fraction. The suitability of the approximator $h^*_{t_j}$ depends on the choice of the subspace $\mathcal{V}_{t_j}$ which is selected under consideration of the target implementation, or more precisely, based on the qualitative understanding of significant reasons for subkey-dependent side-channel leakage (engineer's task!). Considering only the impact of (dis-)charging the bus lines in an 8-bit achitecture, for instance, yields a 9-dimensional subspace, also considering possible cross-talk phenomena of neighboured lines enlarges this subspace by 7 dimensions. Large coefficients of $h^*_{t_j}$ with regard to a particular vector space basis imply that this 'direction' has considerable impact on the side-channel leakage while small coefficients imply that these components are negligible. This constitutes a further advantage of our approach since one learns which reasons have significant impact on the side-channel leakage, supporting constructively a re-design of the device (if necessary). The adversary clearly may try different subspaces, adding or removing basis vectors ([3]).

In [2] a basic version of our approach (not considering masking countermeasures) was introduced and also experimentally verified. In [1] extensive experiments were performed to compare this basic version with template attacks; for the used device the number of measurements in the profiling phase could be reduced down to 2% (compared to template attacks) with our basic version, even then with a tolerable loss of efficiency in the key extraction phase. In the presence of masking countermeasures the efficiency gain in the profiling phase is even by one order of magnitude larger. We note that variants of template attacks have been proposed which employ strong model assumptions at least in intermediate steps (e.g. Hamming weight model) which reduces the number of measurements in the profiling phase but (presumably) also the attacking efficiency. We mention that exemplary measurements at a masked implementation performed at the chair of Christof Paar (by K. Lemke-Rust) confirmed the effectiveness of the stochastic approach in the presence of masking.

References

1. B. Gierlichs, K. Lemke, and C. Paar: Templates vs. Stochastic Methods. In: L. Goubin, M. Matsui (eds.): Cryptographic Hardware and Embedded Systems — CHES 2006, Springer, 15–29. Lecture Notes in Computer Science 4249, Berlin, 2006.
2. W. Schindler, K. Lemke, and C. Paar: A Stochastic Model for Differential Side Channel Analysis. In: J.R. Rao, B. Sunar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2005, 30–46. Springer, Lecture Notes in Computer Science 3659, Berlin, 2005.
3. W. Schindler: Advanced Stochastic Methods in Side Channel Analysis on Block Ciphers in the Presence of Masking. Submitted.