

Multi-channel Attacks

Dakshi Agrawal, Josyula R. Rao, and Pankaj Rohatgi

IBM Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
{[agrawal](mailto:agrawal@us.ibm.com),[jrrao](mailto:jrrao@us.ibm.com),[rohatgi](mailto:rohatgi@us.ibm.com)}@us.ibm.com

Abstract. We introduce *multi-channel attacks*, i.e., side-channel attacks which utilize multiple side-channels such as power and EM *simultaneously*. We propose an adversarial model which combines a CMOS leakage model and the maximum-likelihood principle for performing and analyzing such attacks. This model is essential for deriving the optimal and very often counter-intuitive techniques for channel selection and data analysis. We show that using multiple channels is better for template attacks by experimentally showing a three-fold reduction in the error probability. Developing sound countermeasures against multi-channel attacks requires a rigorous leakage assessment methodology. Under suitable assumptions and approximations, our model also yields a practical assessment methodology for net information leakage from the power and all available EM channels in constrained devices such as chip-cards. Classical DPA/DEMA style attacks assume an adversary weaker than that of our model. For this adversary, we apply the maximum-likelihood principle to such design new and more efficient single and multiple-channel DPA/DEMA attacks.

Keywords: Side-channel attacks, Power Analysis, EM Analysis, DPA, DEMA.

1 Introduction

1.1 The Problem

Recent research in side-channel attacks has validated and reinforced the observation that sensitive information can leak from cryptographic devices via a multitude of channels. The seminal work of [9,8] describing leakages in timing and power channels was followed by the work of [10,7,1] showing leakages via electromagnetic (EM) emanations. The work of [1] shows that even a single EM probe can yield multiple EM signals via demodulation of different carriers. Further, different EM carriers carry different information and some EM leakages exceed leakages in the power channel. All these channels provide a rich source of information for a determined adversary.

While it seems plausible that side-channel attacks can be significantly improved by capturing multiple side-channel signals such as the various EM channels and possibly the power channel, a number of questions remain. Which

side-channel signals should be collected? How should information from various channels be combined? How can one quantify the advantage of using multiple channels? These issues are especially relevant to an attacker since a significant equipment cost is associated with capturing each additional side-channel signal. Furthermore, in some situations, the detection risk associated with the additional equipment/probes required to capture a particular side channel has to be weighed against the benefit provided by that channel.

1.2 Contributions

To address these issues, we present a formal adversarial model for multi-channel analysis using the power and various EM channels.¹ Our model is based on a leakage model for CMOS devices and concepts from the Signal Detection and Estimation Theory. This formal model can be used to assess how an adversary can *best exploit* the wide array of signals available to him. In theory, this model can also deal with the problem of *optimal* channel selection and data analysis. However, in practice, a straight-forward application of this model can sometimes be infeasible. We show a judicious choice of approximations that renders the model useful for most practical applications.

Formulating such an adversarial model has numerous pitfalls. Ideally, the model should capture the strongest possible multi-channel attacks on an implementation of a cryptographic algorithm involving secret data. While such a model is easy to define, using it to assess vulnerabilities and create attacks will shift the focus from multi-channel information leakage to the specifics of the algorithm and implementation.

To refocus the attention on information leakage from multiple side-channels, we will only consider *elementary leakages*, i.e., information leaked during *elementary operations* of CMOS devices. This allows us to deal with information leakage aspects of multiple channels while not losing sight of the goal of evaluating entire implementations. In fact, it can be shown that the leakage in an entire computation is just the composition of elementary leakages from all of its elementary operations[2].

We introduce an adversarial model that is based on this view of elementary leakages of CMOS devices and is phrased in terms of the maximum likelihood testing of hypotheses. The model provides a formal way of comparing efficacies of various signal selection and processing techniques that can be used by a resource limited adversary.

Applying the model to the problem of signal selection, we find that the optimal strategies for picking even two best side-channels from a set of possibilities can be complex and counter-intuitive. For instance, picking the two channels with the best signal-to-noise ratios is quite often sub-optimal. The model also shows how to best combine information from multiple channels. This can be viewed as a generalization of template attacks [4] to the case of multiple channels. We provide experimental evidence to show that multi-channel based template attacks

¹ Combining the timing and power channel is already known, e.g., [11].

are superior to their single channel counterparts. Specifically, for a smart-card S^2 , we show that template attacks that use both an EM channel and a power signal are superior to attacks that use only a single channel.

Our model for multi-channel attacks is also valuable for the designers of cryptographic implementations since they need to know the amount of leakage from multiple sensors to select the appropriate level of countermeasures. We describe a methodology for assessing *any type of leakage* in an *information-theoretic* sense. The methodology permits the computation of bounds on the best error probability achieved by an all-powerful adversary. While such an assessment is impractical for arbitrary devices, it is feasible for the practically important case of chipcards with small word lengths.

One drawback of our model is assumption of a very powerful adversary who has full knowledge of the characteristics of the target device and is capable of performing attacks similar to template attacks on the device. In practice, such attacks are tedious to mount and often adversaries don't have knowledge about the device. Thus, DPA-style attacks continue to be important due to their simplicity and immediate applicability to unknown implementations. Using the maximum likelihood testing as a basis, we show how current single channel DPA-style attacks can be greatly improved and how multiple-channel DPA-style attacks can be designed. The key to these improvements is a relaxation of the maximum likelihood test which estimates the unknown parameters of the test on the fly. We provide empirical evidence to show that a better analysis can give a substantial reduction in the number of samples needed for a traditional DPA attack and even a better reduction factor when a multiple-channel DPA attack is carried out using a power and an EM channel with very similar leakage characteristics.

2 Adversarial Model

This section develops an adversarial model to formally address issues related to the leakage of information via multiple side-channel signals.

2.1 CMOS Side-Channel Elementary Leakages

In CMOS devices, all data processing is typically controlled by a “square-wave” shaped clock. Each clock edge triggers a short sequence of state changing events and corresponding currents in the data processing units. The events are transient and a steady state is achieved well before the next clock edge. At any clock cycle, all the events and resulting currents are determined by a comparatively small number of bits of the logic state of the device, i.e., one only needs to consider the state of *active* circuits during that clock cycle and not the entire state of the device. These bits, termed as *relevant bits* in [3], constitute the *relevant state* of the device.

² A pseudonym is used to protect vendor identity; S is a 6805-based sub-micron, double metal technology card with inbuilt noise generators

Signals found on side-channels such as power and EM result from the current flows within the device and are affected by the random thermal noise. As mentioned above, ideally, the current flows in a CMOS device are directly attributable to the relevant state of the device. However, in practice, there may be many very small leakage currents in the *inactive* parts of the circuit. These leakages can be approximated as a small Gaussian noise term having negligible correlation with any particular active part of the circuit.

Thus as a very good approximation, all side-channel emanations during a clock cycle carry information *only* about the events and the relevant state of the device that occurs during the clock cycle. This is strongly supported by the experimental results which show that algorithmic bits are significantly correlated to the power/EM signals *only* during the clock cycles when they are actively involved in a computation. Thus it is natural to model side-channel leakage from the CMOS devices in terms of the leakages of the relevant state that occur during each clock cycle. We term the operation performed by the device during each clock cycle as an *elementary operation* and define the corresponding leakage of the relevant state information from side-channels as an *elementary leakage*.

2.2 Adversarial Model for Elementary Leakages

Given the concept of elementary leakages, it is natural to formulate side-channel attacks in terms of how successful an adversary can be in obtaining information about the relevant state. For example, an adversary may be interested in the LSB of the data bus during a LOAD instruction. This has a natural formulation as a binary hypothesis testing problem for the adversary³. Such a formulation also makes sense as traditionally the binary hypothesis testing has been central to the notions of side-channel attack resistance and leakage immunity [3,5].

The adversarial model consists of two phases. The first phase, known as the *the profiling phase*, is a training phase for the adversary. He is given a training device identical to the target device, an elementary operation, two distinct probability distributions B_0 and B_1 on the relevant states from which the operation can be invoked and a set of sensors for monitoring side-channel signals. The adversary can invoke the elementary operation, on the training device, starting from any relevant state. It is expected that adversary uses this phase to prepare an attack.

In the second phase, known as the *the hypothesis testing phase*, the adversary is given the target device and the same set of sensors. He is allowed to make a *bounded number* of invocations to the same elementary operation on the target device starting from a relevant state that is drawn *independently* for each invocation according to exactly one of the two distributions B_0 or B_1 . The choice of distribution is unknown to the adversary and his task is to use the signals on the sensors to select the correct hypothesis (H_0 for B_0 and H_1 for B_1) about

³ In general, the adversary faces an M -ary hypothesis testing problem on functions of relevant state, for which results are straightforward generalizations of binary hypothesis testing.

the distribution used. The utility of the side-channels can then be measured in terms of the success probability achieved by the adversary as a function of the number of invocations.

2.3 Sophisticated Attack Strategies

Assume that an adversary acquires L statistically independent sets of sensor signals $\mathbf{O}_i, i = 1, \dots, L$. These L sets of signals may correspond to L invocations of an operation on the target device. Also assume that there are K equally likely hypotheses $H_k, k = 1, \dots, K$, on the origin of these signals. Let $p(\mathbf{O}|H)$ be the probability distribution of the sensor signals under the hypothesis H . Under these assumptions, the *maximum likelihood hypothesis* test is optimal and it decides in favor of the hypothesis H_k if

$$k = \operatorname{argmax}_{1 \leq k \leq K} \prod_{i=1}^L p(\mathbf{O}_i | H_k). \quad (1)$$

While the maximum likelihood test is optimal, it is usually impractical as an exact characterization of the probability distribution of the sensor signals \mathbf{O} may be infeasible. Such a characterization has to capture the nature of each of the sensor signals and the dependencies among them. This could further be complicated by the fact that, in addition to the thermal noise, the sensor signals could also display additional structure due to the interplay between properties of the device and those of the distributions of the relevant states. For example, if the hypothesis was on the LSB of a register while the device produced widely different signals only when the MSB was different, the sensor signals will display a bimodal effect attributable to the MSB. It turns out that in practice one can obtain near optimal results by making the right assumptions about the sensor signals. Such assumptions greatly simplify the task of hypothesis testing by requiring only a partial characterization of sensor signals.

The Gaussian Assumption. One such widely applicable assumption is the *Gaussian assumption* which states that under the hypothesis H , the sensor signal \mathbf{O} has a multivariate Gaussian distribution with mean μ_H and a covariance matrix Σ_H . A multivariate Gaussian distribution $p(\cdot|H)$ has the following form:

$$p(\mathbf{o}|H) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_H|}} \exp\left(-\frac{1}{2}(\mathbf{o} - \mu_H)^T \Sigma_H^{-1} (\mathbf{o} - \mu_H)\right), \quad \mathbf{o} \in \mathcal{R}^n, \quad (2)$$

where $|\Sigma_H|$ denotes the determinant of Σ_H and Σ_H^{-1} denotes the inverse of Σ_H .

The Gaussian assumption holds for a large number of devices and hypotheses encountered in the practice. In fact this assumption has been used successfully in the case of chip-cards [4].

It can be shown that under the Gaussian assumption, the maximum likelihood hypothesis testing for a single observation \mathbf{O} and two equally likely hypothesis H_0 and H_1 ⁴ simplifies to the following comparison:

$$(\mathbf{O} - \mu_{H_0})^T \Sigma_{H_0}^{-1} (\mathbf{O} - \mu_{H_0}) - (\mathbf{O} - \mu_{H_1})^T \Sigma_{H_1}^{-1} (\mathbf{O} - \mu_{H_1}) \geq \ln(|\Sigma_{H_1}|) - \ln(|\Sigma_{H_0}|) \quad (3)$$

where a decision is made in favor of H_1 if the above comparison is true, and in favor of H_0 otherwise.

In many cases of practical interest, noise in the sensor signals does not depend on the hypothesis, that is, $\Sigma_{H_0} = \Sigma_{H_1} = \Sigma_N$. In such cases, the following well-known result from the Statistics gives the probability of error in maximum-likelihood hypothesis testing [12]:

Fact 1 *For equally likely binary hypotheses, the probability of error in the maximum likelihood testing is given by*

$$P_\epsilon = \frac{1}{2} \operatorname{erfc}\left(\frac{\Delta}{2\sqrt{2}}\right) \quad (4)$$

where $\Delta^2 = (\mu_{H_1} - \mu_{H_0})^T \Sigma_N^{-1} (\mu_{H_1} - \mu_{H_0})$ and $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$. Note that Δ^2 has a nice interpretation as the optimal signal-to-noise ratio that an adversary can achieve under the Gaussian assumption.

In the rest of this section, we will present two applications of the theory discussed above. In the first application, a strategy for selecting multiple side channels is presented. In the second application, a template attack on multiple channels is devised.

2.4 Multiple Channel Selection

Consider a resource limited adversary who can select at most M channels for an attack. When viewed in terms of our model, this problem conceptually has a very simple solution: The adversary should choose those M channels that minimize his probability of error in the maximum likelihood testing.

This apparently simple technique can be quite subtle and tricky in practice. Clearly, in situations where a well-prepared adversary has nicely characterized/approximated signals from each of the channels under each hypothesis and the corresponding joint noise probability distribution between all the channels, the adversary can also calculate the error probability for each possible choice of M channels, at least for small M . For example, if the noise is Gaussian and independent of the hypothesis, then from Equation 4, since $\operatorname{erfc}(\cdot)$ decreases exponentially with Δ , the goal of an adversary limited to just two channels, would be to choose channels in such a manner, as to maximize the output signal-to-noise ratio Δ^2 .

⁴ Generalizations to multiple observations and more than two hypotheses are straightforward.

If instead of a rigorous approach, channels are selected by heuristic techniques, then the resulting selection could be sub-optimal for various subtle reasons. Firstly, different side-channels could leak different aspects of information relative to the hypotheses being tested and sometimes there could be value in combining channels which provide widely dissimilar information rather than combining those which provide similar but partial information. Secondly, even if many channels provide the same information, picking multiple channels from this set could still be valuable since that may be almost as good as having the ability to make multiple invocations of the device with the same data and collecting a single side-channel. Even for the case where only two side-channels can be selected, the optimal choice is quite tricky and subtle as shown by the example below where the naive approach of choosing the two signals with best signal-to-noise ratios is shown to be sub-optimal.

Example 1. Consider the case where an adversary can collect two signals $[O_1, O_2]^T$ at a single point in time, such that under the hypothesis H_0 , $O_k = N_k$, for $k = 1, 2$, and under the hypothesis H_1 , $O_k = S_k + N_k$. Assume that $\mathbf{N}_i = (N_1, N_2)^T$ has zero mean multivariate Gaussian distribution with

$$\Sigma_N = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Note that O_1 and O_2 have signal-to-noise ratios of S_1^2 and S_2^2 respectively. After some algebraic manipulations, we get

$$\Delta^2 = \frac{(S_1 + S_2)^2}{2(1 + \rho)} + \frac{(S_1 - S_2)^2}{2(1 - \rho)} \quad (5)$$

Now, consider the case of an adversary who discovers two AM modulated carrier frequencies which are close and carry compromising information, both of which have very high and equally good signal-to-noise ratios ($S_1 = S_2$) and another AM modulated carrier in a very different band with a lower signal-to-noise ratio. An intuitive approach would be to pick the two carriers with high signal-to-noise ratio. In this case $S_1 = S_2$ and we get, $\Delta^2 = 2S_1^2/(1 + \rho)$. Since both signals originate from carriers of similar frequencies, the noise that they carry will have a high correlation coefficient ρ , which reduces Δ^2 at the *output*. On the other hand, if the adversary collects one signal from a good carrier and the other from the worse quality carrier in the different band, then the noise correlation is likely to be lower or even 0. In this case:

$$\Delta^2 = \frac{(S_1 + S_2)^2}{2} + \frac{(S_1 - S_2)^2}{2} = S_1^2(1 + S_2^2/S_1^2) \quad (6)$$

It is clear that the combination of a high and a low signal-to-noise ratio signals would be a *better strategy* as long as $S_2^2/S_1^2 > (1 - \rho)/(1 + \rho)$. For example, if $\rho > 1/3$, then choosing carriers from different frequency bands with even half the signal-to-noise ratio results in better hypothesis testing. ■

Based on above analysis, in our experiments we routinely rejected a stronger channel which is colocated with another collected channel and chose a channel further away in the spectrum even if it had a lower signal-to-noise ratio.

2.5 Multi-channel Template Attacks

In [4], the power of using the maximum likelihood principle together with the Gaussian assumption was shown to be very effective in classifying successive bytes of an RC4 key using a *single* power side-channel signal. Expanding the template approach to multiple channel is straightforward. In the template attack, at any stage, the adversary uses an identical device to build exact characterizations for the signal and noise for each of the K possibilities he has to classify. Then he uses these characterizations to classify the one signal he is given from the target device. The first step in the template approach is the identification of those time instances (or indices of sample points) where the average signals for each of the K possibilities differ significantly. The second step is to compute the joint noise distribution of the channel at those points for each of the K possibilities. The third step is to classify the given signal into the K possibilities using the maximum likelihood testing.

For multiple-channels, the template attack is identical except that the signals from the multiple channels are concatenated together to yield a larger signal, i.e., for each invocation, a combined signal is created by concatenating the signals from the individual observed channels. Notice that the process of identifying the time instances and sample points could end up selecting somewhat different time slices for each channel, depending purely on the nature of leakage in each channel. The maximum likelihood testing will pick up information from all channels (possibly at different times) for classification.

To show that multiple channels help the classification process, we invoke an operation on the smart card S with two different input bytes and look at just 3 cycles during which the input was first processed. We collected EM and power samples simultaneously and evaluated how well the template attack could classify a single EM/power trace into the two hypotheses H_0 and H_1 for the input byte. We did this classification first using exactly one of the power/EM channels and then performed the classification using both channels simultaneously. Figures 1 shows the mean EM and Power signals for these hypothesis during these 3 cycles side by side.⁵ Fig 2 shows the error rate of our classification effort for inputs belonging to each hypothesis. One can clearly notice that using both channels simultaneously results in better classification compared to any single channel.

3 Leakage Assessment Methodology for Chipcards

The model developed in Section 2.2 can be used to derive a practical methodology for assessing information leakage from any L power and EM channels for simple CMOS devices such as 8-bit chipcards. Several key properties make such a methodology feasible. Firstly, for a fixed relevant state, the noise at any cycle is well-approximated by a Gaussian distribution. Thus, in the hypothesis testing phase, the problem becomes one of distinguishing between two distributions

⁵ The slight offset in time is due to delay of EM signals with respect to the power signal.

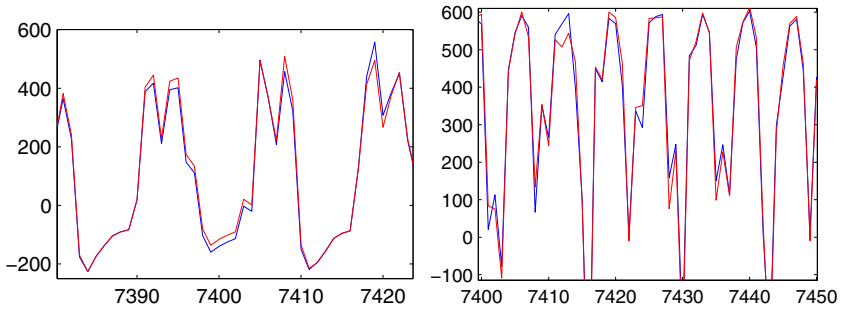


Fig. 1. Mean Power and EM signals during 3 cycles for two hypothesis

Correct Hypothesis	Error(Pwr)	Error(EM)	Error(EM+Pwr)
H0	9.5%	15.1%	2.8%
H1	20.1%	15.2%	6.6%

Fig. 2. Signal classification error using Power, EM and combination of Power and EM

B_0 and B_1 which are mixtures of Gaussians. Thus, if the number of relevant states (typically exponential in twice the word length) is small, each Gaussian in the collection can be profiled and the success probability for hypothesis testing can be computed. The problem of capturing leakages across multiple bands in multiple channels can be practically solved by splitting each channel into slightly overlapping bands upto a reasonable upper limit. Details of this assessment methodology with such practical considerations are given in the Appendix.

4 Single and Multi-channel DPA Attacks

In section 2, we assumed that the adversary had access to a test device identical to the target device and that he could carry out a profiling stage using the test device. In many circumstances, access to a test device may not be possible. In such cases, a DPA-style attack is preferred since it assumes no prior knowledge of device characteristics or implementation. In this section, we apply tools from the detection theory to optimize existing single channel DPA attacks and propose new multiple channel DPA attacks.

4.1 Improving DPA

In the traditional DPA attack, an adversary collects a set of N signals, $\mathbf{O}_i, i = 1, \dots, N$ emanating from a given channel. Assume that the signals are normalized to have zero sample average over all N signals. For each hypothesis H under consideration, the N signals are divided into two bins, termed the 0-bin and the 1-bin with $N_{H,0}$ and $N_{H,1}$ samples respectively. Let $\mu_{H,0}[j]$ and $\mu_{H,1}[j]$ be the

sample means of signals in the 0-bin and the 1-bin respectively for the hypothesis H . The next step in the DPA attack consists of computing the differences of sample means $\mu_H[j] = \mu_{H,0}[j] - \mu_{H,1}[j]$ for all hypotheses, and deciding in favor of the hypothesis H_i if $|\mu_{H_i}[j]|$ has the largest peak among all differences of means. In other words, the decision metric for the hypothesis H at time j is given by

$$M_H[j] = \left(\mu_{H,0}[j] - \mu_{H,1}[j] \right)^2, \quad (7)$$

and the decision is made in favor of the hypothesis H_i if for some value of j , say j_0 , $M_{H_i}[j_0] \geq M_H[j]$ for all H and j .

The traditional DPA attack and its variations have been successfully applied to attack several cryptographic implementations. However, by using the theory developed in the previous section, the effectiveness of traditional DPA can be increased significantly.

Before proceeding further, assume a void hypothesis H_v which corresponds to a random bifurcation of the N signals into the 0-bin and the 1-bin. Using the Gaussian assumption and Equation 3, the metric of a hypothesis H_i with respect to the null hypothesis at time j is given by

$$M_{H_i}[j] = \frac{\left(\mu_{H_i}[j] - E[\mu_{H_v}[j]] \right)^2}{V[\mu_{H_v}[j]]} - \frac{\left(\mu_{H_i}[j] - E[\mu_{H_i}[j]] \right)^2}{V[\mu_{H_i}[j]]} - \ln \left(\frac{V[\mu_{H_i}[j]]}{V[\mu_{H_v}[j]]} \right) \quad (8)$$

In order to compute this metric, we need the values of the following parameters: $E[\mu_{H_v}[j]$, $V[\mu_{H_v}[j]$, $E[\mu_H[j]]$, and $V[\mu_H[j]$. Since in the DPA attack, the adversary skips the profiling phase of the attack, (8) is not directly applicable. In such cases, the theory suggests that unknown parameters of the test equation be estimated directly from the collected signals. If the adversary uses a maximum-likelihood estimate of these parameters, then the resulting test is referred to as the generalized maximum-likelihood testing.

For the DPA attack, calculating the maximum likelihood estimate of the test parameters involves solving a set of nonlinear coupled equations. Therefore, instead of using the maximum-likelihood estimates of these parameters, we use sample estimates as follows: Let $\sigma_{H,0}^2[j]$ and $\sigma_{H,1}^2[j]$ be the sample variances of the signals in the 0-bin and the 1-bin respectively at time j for hypothesis H . We propose the following sample estimators⁶ of parameters in (8):

$$\begin{aligned} E[\mu_H[j]] &= \mu_H[j] \\ V[\mu_H[j]] &= \frac{\sigma_{H,0}^2[j]}{N_0} + \frac{\sigma_{H,1}^2[j]}{N_1} \end{aligned} \quad (9)$$

⁶ We omit the derivation of these estimators as the derivation is tedious and follows from straight-forward algebraic manipulations.

Sbox Hyp	Min Samples (Mean-diff)	Min Samples(Max-Likl)
S1,B3	640	350
S2,B3	630	210
S7,B3	110	40
S8,B3	130	90

Fig. 3. DPA results, mean-difference vs. approx. generalized maximum-likelihood

Substituting these in (8), we get the following formula for the metric:

$$M_{H_i}[j] = \frac{\left(\mu_{H_i}[j] - \mu_{H_v}[j]\right)^2}{\frac{\sigma_{H_v,0}^2[j]}{N_0} + \frac{\sigma_{H_v,1}^2[j]}{N_1}} - \ln\left(\frac{\frac{\sigma_{H_i,0}^2[j]}{N_0} + \frac{\sigma_{H_i,1}^2[j]}{N_1}}{\frac{\sigma_{H_v,0}^2[j]}{N_0} + \frac{\sigma_{H_v,1}^2[j]}{N_1}}}\right) \quad (10)$$

Intuitively, the signals in the 0-bin and 1-bin have similar distributions under the wrong hypothesis due to a random bifurcation of signals in the two bins. However, for the correct hypothesis, the distribution of signals in the 0-bin differs from the distribution of signals in the 1-bin. The traditional DPA attack only takes into account the differences in sample means. On the other hand, Equation 10 takes both the sample means and variances into account, and therefore may provide a better hypothesis test.

Figure 3 shows the results of applying this method to attacking the S-box lookup for a DES implementation. The first column shows the bit being predicted, the second shows the number of samples required for the correct key hypothesis to emerge as the winner under the traditional DPA metric while the third column shows the number of samples needed with the new metric. Clearly by using a better metric, our improvement in the DPA attack reduces the number of signals needed by a factor of 1.4–3.

4.2 Multi-channel DPA Attack

Multi-channel DPA attack is a generalization of the single-channel DPA attack. In this case, the adversary collects N signals, $\mathbf{O}_i, i = 1, \dots, N$. In turn, each of the signals \mathbf{O}_i is a collection of L signals collected from L side-channels. Thus, $\mathbf{O}_i = [\mathbf{O}_i^1, \dots, \mathbf{O}_i^L]^T$ where \mathbf{O}_i^l represents the i -th signal from the l -th channel. Note that all DPA style attacks treat each time instant independently and leakages from multiple channels can only be pooled together if they occur at the same time. Thus, in order for multi-channel DPA attacks to be effective, the selected channels must have very similar leakage characteristics.

The formulae for computing the metric for multi-channel DPA attack are generalizations of those for the single channel. The main difference is that the expected value of sample mean difference at time j under hypothesis H is a vector of length L , with the l -th entry being the sample mean difference of the l -th channel. Furthermore, the variance of the b -bin under hypothesis H at time j , is a covariance matrix of size $L \times L$ with the i, j -th entry being the correlation between signals from the i -th and j -th channels. Once again, as in the DPA

attack, the adversary does not have the luxury of estimating these parameters. Therefore, we substitute sample estimates for these parameters along the same lines as in Equation 9. We skip the cumbersome formulae and directly go to the results of multi-channel DPA attacks.

Figure 4 shows sample results of an attack on the S-box lookups in a DES implementation using the power channel together with an EM channel whose leakage is similar to the power channel. The first column shows the bit being predicted, the second shows the number of signals required for the correct key hypothesis to emerge as the winner using both channels with the multi-channel metric, the last two columns show the number of signals needed for the power and EM channels separately using the new DPA/DEMA metric. From this it is clear that the number of invocations needed for two channel attacks can be significantly less compared to single-channel attacks.

Sbox Hyp	Min Samples(Pwr+EM)	Min Samples(Pwr)	Min Samples(EM)
S1,B1	150	170	640
S1,B2	60	(>1000)	340
S1,B3	110	350	160
S2,B2	30	50	230
S2,B3	120	210	340
S4,B0	60	200	340
S6,B1	180	180	190
S7,B3	30	40	520
S8,B3	60	90	140

Fig. 4. Multi-Channel DPA-style attack using Power, EM and Power&EM. and EM

4.3 Future Work on Single/Multi-channel DPA/DEMA Attacks

It is well known to DPA/DEMA practitioners that for the correct hypothesis, the correlation signal with respect to time shows multiple peaks. However, current analysis techniques, including the ones presented here, do not combine information from peaks occurring at different time instances. This problem also manifests itself when combining various Power and EM channels since peaks on different channels may not coincide. One can also view the efficacy gap between template attacks and DPA attacks as a manifestation of the same problem.

We have started work which promises to bridge this gap. The main idea is to estimate the characteristics of useful peaks on the fly given only the collected signals (without using a training set) and apply techniques based on maximum-likelihood principle to identify the correct hypothesis.

References

1. Dakshi Agrawal, Bruce Archambeault, Josyula R. Rao and Pankaj Rohatgi. The EM Side Channel(s). Proceedings of CHES 2002, Springer, LNCS 2523, pp 29–45.

2. D. Agrawal, B. Archambeault, J. R. Rao and P. Rohatgi. The EM Side Channel(s): Attacks and Assessment Methodologies. See <http://www.research.ibm.com/intsec/emf-paper.ps>.
3. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao and Pankaj Rohatgi. Towards Sound Countermeasures to Counteract Power–Analysis Attacks. Proceedings of Crypto '99, Springer–Verlag, LNCS 1666, pp 398–412.
4. Suresh Chari, Josyula R. Rao and Pankaj Rohatgi. Template Attacks. Proceedings of CHES 2002, Springer, LNCS 2523, pp 13–28.
5. Jean–Sebastien Coron, Paul Kocher and David Naccache. Statistics and Secret Leakage. In the Proceedings of Financial Cryptography '00. Springer-Verlag, LNCS 1962, pp 157–173
6. L. Goubin and J. Patarin. DES and Differential Power Analysis. Proceedings of CHES '99, LNCS 1717, pp 158–172.
7. K. Gandolfi, C. Moutrel and F. Olivier. Electromagnetic Attacks: Concrete Results. Proceedings of CHES '01, LNCS 2162, pp 251–261.
8. P. Kocher. Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS and Other Systems. Advances in Cryptology-Crypto '96, Springer-Verlag, LNCS 1109, pp 104–113.
9. P. Kocher, J. Jaffe and B. Jun. Differential Power Analysis: Leaking Secrets. Advances in Cryptology — Proceedings of Crypto '99, Springer Verlag, LNCS 1666, pp. 388–397.
10. Jean–Jacques Quisquater and David Samyde. ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smart Cards. In Smart Card Programming and Security (E-smart 2001), LNCS 2140, pp.200-210,September 2001.
11. C. D. Walter and S. Thompson. Distinguishing Exponent Digits by Observing Modular Subtractions. In Progress in Cryptology- CT-RSA 2001, Springer, LNCS 2020, pp 192–207.
12. H. L. Van Trees. Detection, Estimation, and Modulation Theory, Part I. John Wiley & Sons. New York. 1968.

Appendix: Leakage Assessment for Chipcards

In this section, we address the question of whether one can assess and quantify the net leakage of information from multiple sensors. Can the information obtained by combining leakages from several (or even all possible) signals from available sensors be quantified regardless of the signal processing capabilities and computing power of an adversary?

Maximum likelihood testing is the optimal way to perform hypothesis testing. Thus, we use it to craft a methodology to assess information leakage from elementary operations. Our methodology takes into account signals extractable from all the given sensors across the entire EM spectrum. Results of such an assessment will enable one to bound the success probability of the optimal adversary for any given hypothesis.

Assume, that for a single invocation, the adversary captures the emanations across the entire electromagnetic spectrum from all sensors in an observation vector \mathbf{O} . Let Ω denote the space of all possible observation vectors \mathbf{O} . Since the likelihood ratio, $A(\mathbf{O})$ is a function of the random vector \mathbf{O} , the best achievable success probability, P_s , is given by:

$$P_s = \sum_{\mathbf{O} \in \Omega} I_{\{A(\mathbf{O}) > 1\}} p_{\mathbf{N1}}(\mathbf{O} - \mathbf{S}_1) + I_{\{A(\mathbf{O}) < 1\}} p_{\mathbf{N0}}(\mathbf{O} - \mathbf{S}_0) \quad (11)$$

where I_A denotes the indicator function of the set A , and $p_{\mathbf{N1}}(\mathbf{O} - \mathbf{S}_1)$ and $p_{\mathbf{N0}}(\mathbf{O} - \mathbf{S}_0)$ are noise distributions under the hypothesis 1 and 0.

When the adversary has access to multiple invocations, an easier way of estimating the probability of success/error involves a technique based on moment generating functions. We begin by defining the logarithm of the moment generating function of the likelihood ratio:

$$\mu(s) = \ln \left(\sum_{\mathbf{O} \in \Omega} p_{\mathbf{N1}}^s(\mathbf{O} - \mathbf{S}_1) p_{\mathbf{N0}}^{1-s}(\mathbf{O} - \mathbf{S}_0) \right) \quad (12)$$

The following is a well-known result from Information Theory:

Fact 2 Assume we have several statistically independent observation vectors⁷ $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_L$. For this case, the best possible exponent in the probability of error is given by the Chernoff Information:

$$C \stackrel{\text{def}}{=} - \min_{0 \leq s \leq 1} \mu(s) \stackrel{\text{def}}{=} - \mu(s_m) \quad (13)$$

Note that $\mu(\cdot)$ is a smooth, infinitely differentiable, convex function and therefore it is possible to approximate s_m by interpolating in the domain of interest and finding the minima. Furthermore, under certain mild conditions on the parameters, the error probability can be approximated by:

$$P_\epsilon \approx \frac{1}{\sqrt{8\pi L \mu''(s_m)} s_m (1 - s_m)} \exp(L \mu(s_m)) \quad (14)$$

Note that in order to evaluate (11) or (14), we need to estimate $p_{\mathbf{N0}}(\cdot)$ and $p_{\mathbf{N1}}(\cdot)$. In general, this can be a difficult task. However by exploiting certain characteristics of the CMOS devices, estimation of $p_{\mathbf{N0}}(\cdot)$ and $p_{\mathbf{N1}}(\cdot)$ can be made more tractable.

Practical Considerations

We will now outline some of the practical issues associated with estimating $p_{\mathbf{N0}}(\cdot)$ and $p_{\mathbf{N1}}(\cdot)$ for any hypothesis. The key here is to estimate the noise distribution for each cycle of each elementary operation and for each relevant state R that the operation can be invoked with. This results in the signal characterization, \mathbf{S}_R , and the noise distribution, $p_{\mathbf{NR}}(\cdot)$ which is sufficient for evaluating $p_{\mathbf{N0}}(\cdot)$ and $p_{\mathbf{N1}}(\cdot)$.

There are two crucial assumptions that facilitate estimating $p_{\mathbf{NR}}(\cdot)$: first, on chipcards examined by us the typical clock cycle is 270 nanoseconds. For such

⁷ For simplicity, this paper deals with *independent* elementary operation invocations. Techniques also exist for adaptive invocations.

devices, most of the compromising emanations are well below 1 GHz which can be captured by sampling the signals at a Nyquist rate of 2 GHz. This sampling rate results in a vector of 540 points per cycle per sensor. Alternatively, one can also capture all compromising emanations by sampling judiciously chosen and slightly overlapping bands of the EM spectrum. The choice of selected bands is dictated by considerations such as signal strength and limitations of the available equipment. Note that the slight overlapping of EM bands would result in a corresponding increase in the number of samples per clock cycle, however it remains in the range of 600-800 samples per sensor.

The second assumption, borne out in practice (see [4]), is that for a fixed relevant state, the noise distribution $p_{\mathbf{NR}}(\cdot)$ can be approximated by a Gaussian distribution. This fact greatly simplifies the estimation of $p_{\mathbf{NR}}(\cdot)$ as only about one thousand samples are needed to roughly characterize $p_{\mathbf{NR}}(\cdot)$. Moreover, the noise density can be stored compactly in terms of the parameters of the Gaussian distribution.

These two assumptions imply that in order to estimate $p_{\mathbf{NR}}(\cdot)$ for a fixed relevant state R , we need to repeatedly invoke (say 1000 times) an operation on the device starting in the state R , and collect samples of the emanations as described above. Subsequently, the signal characterization S_R can be obtained by averaging the collected samples. The noise characterization is obtained by first subtracting S_R from each of the samples and then using the Gaussian assumption to estimate the parameters of the noise distribution.

The assessment can now be used to bound the success of any hypothesis testing attack in our adversarial model. For any two given distributions B_0 and B_1 on the relevant states, the corresponding signal and noise characterizations, $S_0, S_1, p_{\mathbf{N0}}(\cdot)$, and $p_{\mathbf{N1}}(\cdot)$, are a *weighted sum* of the signal and noise assessments of the constituent relevant states S_R and $p_{\mathbf{NR}}(\cdot)$. The error probability of maximum-likelihood testing for a single invocation or its exponent for L invocations can then be bounded using (11) and (13) respectively.

We now give a rough estimate of the effort required to obtain the leakage assessment of an elementary operation. The biggest constraint in this process is the time required to collect samples from approximately one thousand invocations for each relevant state of the elementary operation. For an r -bit machine, the relevant states of interest are approximately 2^{2r} ; thus the leakage assessment requires time to perform approximately $1000 * 2^{2r}$ invocations. Assuming that the noise is Gaussian and that each sensor produces an observation vector of length 800, for n sensors the covariance matrix Σ_N has $(800 * n)^2$ entries. It follows that the computation burden of estimating the noise distribution would be proportional to $(800 * n)^2$. Such an approach is certainly feasible for an evaluation agency, from both a physical and computational viewpoint, as long as the size of the relevant state, r , is small. In our experiments, we found such assessment possible for a variety of 8-bit chipcards.