The EM Side–Channel(s):Attacks and Assessment Methodologies

Dakshi Agrawal Bruce Archambeault Josyula R. Rao
Pankaj Rohatgi
IBM Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
email: {agrawal,barch,jrrao,rohatgi}@us.ibm.com

Abstract

We present a systematic investigation of the leakage of compromising information via electromagnetic (EM) emanations from chipcards and other devices. This information leakage differs substantially from and is more powerful than the leakage from other conventional side-channels such as timing and power. EM emanations are shown to consist of a multiplicity of compromising signals, each leaking somewhat different information. Our experimental results confirm that these signals could individually contain enough leakage to break cryptographic implementations and to defeat countermeasures against other side-channels such as power. Using techniques from Signal Detection Theory, we also show that generalized and far more devastating attacks can be constructed from an effective pooling of leakages from multiple signals derived from EM emanations.

The magnitude of EM exposure demands a leakage assessment methodology whose correctness can be rigorously proved. We define a model that *completely* and quantitatively bounds the information leaked from multiple (or all available) EM side-channel signals in CMOS devices and use that to develop a practical assessment methodology for devices such as chipcards.

1 Introduction

Side–channel cryptanalysis has been used successfully to attack many cryptographic implementations [10, 11]. Most of the publicly available literature on side–channels deals with attacks based on timing or power. While it is rumored that there is a large body of classified literature on exploiting leakages due to electromagnetic (EM) emanations, there is scant information about this in the public domain.

With the recent declassification of portions of the TEMPEST documents [7], initial reports by J. J. Quisquater[12] and publication of some EM attacks [9], an awareness of the potential of the EM side–channel is developing. However, some basic questions remain unanswered. For instance, what are the causes and types of EM emanations? How does information leaked via EM emanations compare with leakages from other side–channels? What implementations are vulnerable to EM side–channel attacks? Can the EM side–channel overcome countermeasures designed to provide protection against other side–channel attacks? Given the set of EM emanations available to an adversary, is it

possible to bound the net information leaked in an *information theoretic* sense, i.e., independent of the computational and signal processing abilities of the adversary? With questions such as these in mind, we conducted a systematic investigation of EM side–channel leakage from chipcards. In this paper, we address each of these basic questions.

A crucial insight from our investigation is that the output of even a *single* EM sensor consists of multiple compromising signals of different types, strengths and information content. Of these, high amounts of compromising information is usually found in very low energy signals. It is therefore imperative that signals be separated early in the acquisition process to avoid loss of low energy signals due to precision limits of signal capturing equipment. Here, we differ from [9] where signals are obtained from different micro–antenna positions, but nevertheless the output of the sensor is treated as a single signal. While careful positioning of a micro–antenna can acquire and emphasize some low energy signals, without further signal separation, some high information, low energy signals can still be overwhelmed by adjacent lower information, higher energy sources.

We describe the basis for this insight in Section 2. We first discuss the causes and types of various EM signals that we observed using several sensors. These include several unexpected signals that are easy to overlook if one does not suspect their existence. For instance, despite extensive work with power signals, researchers had so far missed the very faint, but far more compromising amplitude modulated EM signals present in the power line and other conductors attached to the chipcard. We then describe the experimental equipment, techniques used to extract compromising signals and empirical evidence that confirms that different signals carry different information.

Since EM emanations permit the use of several conductive and radiative sensors each containing multiple signals, an adversary has a wide array of compromising signals at his disposal. In situations where the power side–channel is not available, e.g., the power supply is filtered or the attack has to be mounted from a distance, the advantages of using EM are obvious. The more pertinent question is whether this abundance of signals provides any practical advantage in using the EM side–channel in situations where the traditional side–channels such as power are also available. This question is especially appropriate given the high cost of some EM equipment.

In Section 3, we answer this question in the affirmative. We demonstrate that even low cost EM equipment which can collect only one signal at a time is quite effective against fielded devices. For all devices, we obtained signals which were amenable to attacks such as simple and differential electromagnetic attacks (SEMA and DEMA [12]). We were able to successfully attack cryptographic implementations of block ciphers such as DES, public key schemes such as RSA on chipcards and SSL Accelarators and even proprietary algorithms such as the prescribed COMP128 GSM authentication scheme. More interestingly, in many devices we could obtain EM signals in which some leakages were excessive, i.e., they had a far superior signal to noise ratio than in the power signal [1]. These excessive leakages form the basis of *devastating* attacks against fielded systems including those resistant to power analysis attacks. We present such attacks against two major classes of power analysis countermeasures [11, 1, 6] implemented on a test system [2]. We show that the attack against the secret–sharing countermeasures of [1, 6] is powerful enough to work even in the case when the code of a protected implementation is unknown.

Despite their effectiveness, our low–cost attacks provide only a glimpse of what is possible. We collected only one signal at a time and followed the intuitive strategy of using the signal source with

[1] Most devices have classes of "bad" instructions with excessive leakages.

[2] A test system was chosen to avoid disclosing weaknesses of commercially deployed systems.

the best signal to noise ratio. A better funded adversary would deploy several sensors[3] to collect multiple signals. More importantly, he will use techniques from Signal Detection Theory to launch far more sophisticated and devastating attacks. As we show, such attacks can be highly efficient requiring greatly reduced number of samples: more than an order of magnitude less than traditional schemes that employ simple intuitive statistics such as the mean signal and the L_2–norm. The theory also prescribes better adversarial strategies. For instance, we show that if only two sensor signals can be collected, the intuitive strategy of picking the two signals with the best signal–to–noise ratio can be sub–optimal. Therefore, for the purpose of determining vulnerabilities and for devising countermeasures, it is essential to have a formal model to understand and assess how such an adversary can *best exploit* the wide array of signals available from the sensors that he can deploy.

Formulating such an adversarial model has numerous pitfalls. Ideally, the model should capture the strongest side–channel attacks possible on an implementation of a cryptographic algorithm involving secret data. While it is easy to define such a model, using it to assess vulnerabilities will inevitably move the focus from information leakage from sensors to the analysis of algorithm and implementation specific attacks that various adversaries with a wide spectrum of capabilities could employ.

To refocus the attention on information extractable from the sensors, our model will focus only on *elementary leakages*, i.e., information leaked during *elementary operations* of CMOS devices. Such a focus is not limiting in the information–theoretic sense as we prove that all side–channel information leakage in a computation can be viewed as a composition of elementary leakages from all of its elementary operations. The model of elementary leakages maps directly to first order differential side–channel attacks (attacks in which different cycles of each computation are considered independently, such as DEMA) on implementations. These elementary leakages also serve as basic building blocks for the design and analysis of more sophisticated, algorithm and implementation specific attacks.

In Section 4, we describe the adversarial model in terms of hypothesis testing. The model provides a formal way of comparing efficacies of various signal selection and processing techniques that can be applied by resource limited adversaries in terms of success probabilities achieved. It also provides a framework to quantify and bound the information leakage from the sensors in terms of the best error probability achieved by an all–powerful adversary. Using this model, we describe some sophisticated and counter-intuitive strategies, based on Signal Detection Theory, that a resource limited adversary can use to launch devastating attacks. Next, we describe a methodology based on the same Theory that addresses the issue of 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. We then show (Theorem 1) that such an assessment also *completely* captures side–channel leakages in arbitrary attacks on implementations.

2 EM Signals and Leakages

This section describes the origin and types of various EM signals that we have observed[4]. We provide empirical evidence that confirms that different signals carry different information and describe the experimental equipment and techniques used to extract various compromising signals.

[3] In practice, any adversary will be limited by the number and types of sensors that he can deploy.

[4] While there is an obvious overlap with the declassified TEMPEST documents (NACSIM 5000) [13], we only describe what we have verified in our investigations.

2.1 Origin of EM Emanations

EM emanations arise as a consequence of current flows within the control, I/O, data processing or other parts of a device. These flows and resulting emanations may be *intentional* or *unintentional*. Each current carrying component of the device not only produces its own emanations based on its physical and electrical characteristics but also affects the emanations from other components due to coupling and circuit geometry.

Of these numerous emanations, those induced by data processing operations carry the most compromising information. In CMOS devices, ideally, current only flows when there is a change in the logic state of a device. In addition, 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, the events and resulting currents are determined by a small number of bits of the logic state of the device, i.e., one only needs to consider the *active* circuits during that clock cycle. These bits, termed as *relevant bits* in [1], constitute the *relevant state* of the device. These currents result in many compromising emanations in several unintended ways. Such emanations carry information about the currents and hence the events and relevant state of the device. In practice, CMOS devices are not ideal there may be many very small leakage currents in *inactive* parts of the circuit as well. These can be approximated as a small gaussian noise term having negligible correlation with any particular inactive part of the circuit.[5]

Since each active component of the device produces and induces various types of emanations, these multiple emanations provide multiple views of events unfolding within the device. Views emphasizing different active components can be obtained by using different types and positions of sensors [9, 12] or even by focusing on different types of emanations that can be captured by a single sensor as we will show in this paper. This is in sharp contrast to the power side–channel where there is only a single aggregated view of net current inflow. The presense of multiple views make the EM side–channel(s) much more powerful than the power side–channel.

2.2 Types of EM Emanations

There are two broad categories of EM emanations:

1. Direct Emanations: These result from *intentional* current flows. Many of these consist of short bursts of current with sharp rising edges which result in emanations observable over a wide frequency band. Often, components at the higher frequencies prove more useful to the attacker due to overwhelming noise and interference prevalent in the lower frequency bands. In a complex circuits, isolating direct emanations can be difficult and may require use of tiny field probes positioned very close the signal source and/or special filters so as to minimize interference from other signal sources; getting good results may necessisate having to decapsulate the chip packaging [9, 12].

2. Unintentional Emanations: Increased miniaturization and complexity of modern CMOS devices results in electrical and electromagnetic coupling between components in close proximity. Such couplings, while inconsequential from the perspective of a circuit designer, provide a rich source of compromising emanations to the attacker. These emanations manifest themselves as *modulations* of carrier signals generated, present or "introduced" within the device. One strong source of carrier

[5]This fact is well known in DPA literature where expermentally it is observed that algorithmic bits are significanty correlated to the total current only during the times when they are actively involved in a computation.

signals is the harmonic–rich "square–wave" clock signal[6] propagated throughout the device. Other sources include communication related signals. Some of the ways in which modulation occurs include:
a. Amplitude Modulation: Non-linear coupling between a carrier signal and a data signal results in the generation and emanation of an Amplitude Modulated (AM) signal. The data signal can be extracted by using a receiver tuned to the carrier frequency and then performing AM demodulation.

b. Angle Modulation: Coupling of circuits also results in Angle Modulated Signals (FM or Phase modulation). For instance, while signal generation circuits should ideally be completely decoupled from data processing circuits, this is rarely achieved in practice. For example, if these circuits draw upon limited energy source, the generated signal, very often, is angle modulated by the data signal. The data signal is recoverable by angle demodulation of the generated signal.

Exploiting modulated carriers can be easier and more effective that trying to work with direct emanations. Firstly, some modulated carriers could have substantially better propogation than direct emanations, permitting effective EM attacks without resorting to invasive techniques and attacks that can be performed at a distance. E.g, all attacks decribed in this paper *do not* require any intrusive/invasive techiques or fine grained positioning of probes. Secondly, careful field probe positioning cannot separate two sources of direct emanations in close proximity, whereas such sources may be easily separable due to their differing interaction with the carriers present in the vicinity.

2.2.1 Emperical Results

We now present emperical results which illustrate the types of emanations discussed above.
Experiment 1: Direct Emanations: We used a relatively recent smart card that we call smartcard A (to protect vendor identity[7]), which we programmed to enter a 13 cycle infinite loop, running on the externally supplied 3.68MHz clock. A handmade near-field probe (a small metal plate attached to a co-axial cable) was placed close to the plastic at the back of smart card, near the chip. The signal was amplified using a wideband amplifier (0.1-500Mhz) and 500K sample points (representing approx 284 iterations of the loop) were collected with an 8-bit, 500Mhz digital scope. In the time domain, this baseband (band centered at 0Mhz) *direct emanations* signal, looked like a differentiated form of the external clock and provided *no* visual indication of a loop execution. The situation can be best analysed in the frequency domain. The signal received by the probe consists of the signal of interest, i.e., a periodic signal corresponding to a loop iteration at a frequency 283Khz (3.68Mhz/13), and other signals from the chip and its vicinity such as the clock (periodic with freq 3.68Mhz) and other aperiodic noise. Capturing the received signal with a limited resolution (8-bit) scope further introduces quantization noise. Figure 1 plots the FFT of the captured baseband signal (where the Y-axis is the magnitude[8] and the X-axis is the frequency in KHz). The large spikes below 100 MHz are the high energy harmonics of the clock signal and tiny spikes sprinkled between them are other types of direct and unintentional emanations which are of interest. Very little signal is noticeable above

[6]Theoretically a perfectly symmetric, perfectly square signal consists of the fundamental frequency and all the odd harmonics with progressively diminishing strengths. In practice, the clock signal is always imperfect.

[7]Smartard A is 6805-based, uses 0.6 micron triple metal technology with an optional variable internal clock as one defence against DPA.

[8]In all our figures, signal or FFT magnitudes should be treated as relative quantities, since we don't track the absolute quantities involved as the signals typically undergo analog amplification/processing steps before being captured by our scopes with 8-bit or 12-bit resolution. We typically set the scope sensitivity so that the received signal covers most of the available 8-12 bit dynamic range.

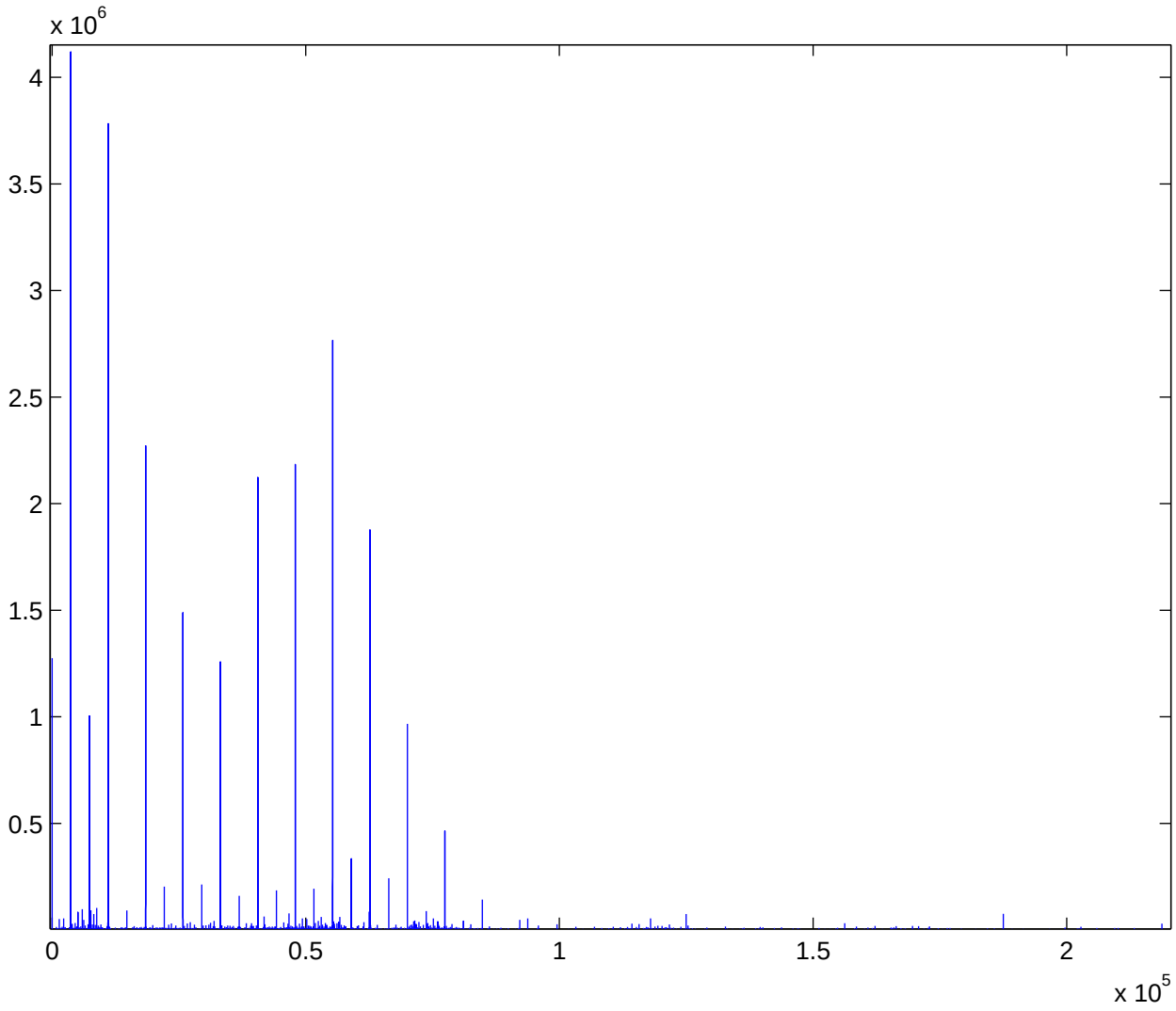

Figure 1: FFT of baseband signal from Experiment 1 with Smartcard A

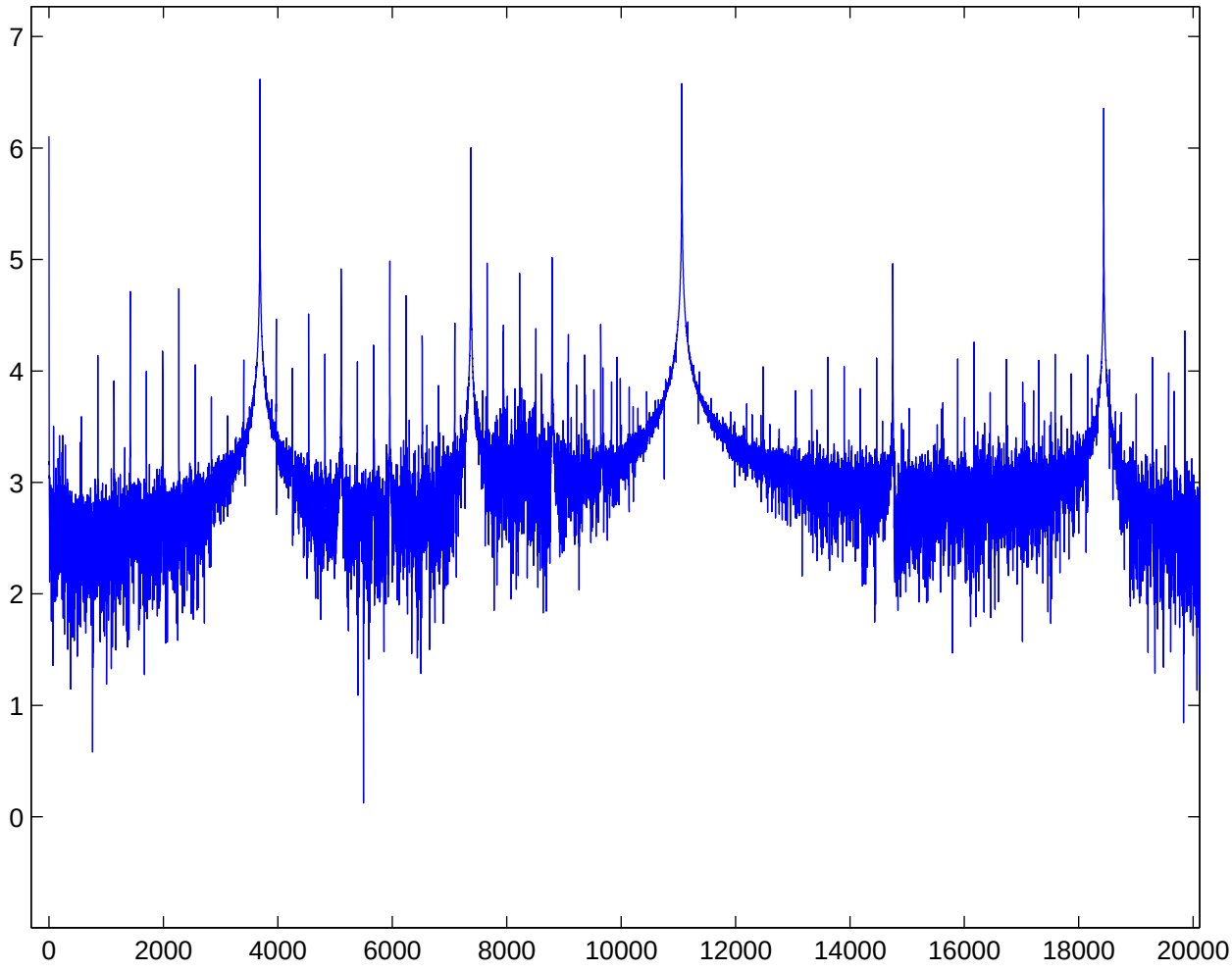

Figure 2: Log of FFT in the region 0-20Mhz from Experiment 1 with Smartcard A

125 MHz, essentially because the signal strenghts here are much lower than in the 0-100Mhz band and these signals have been overwhelmed by quantization noise. In the linear scale, the fact that the card is executing a 13-cycle loop is not apparent. On a log (base 10) scale, zooming into the region from 0 to 20Mhz, in Figure 2 one can indeed see the signal of interest at 283Khz and its harmonics, interspersed between the clock signal and its harmonics. Note that the use of a large time window (284 iterations of the loop) helps in detecting these periodic signals since the aperiodic noise from the chipcard, environment and quantization gets reduced due to averaging. Since direct emanations of interest are more than an order of magnitude smaller than other interfering signals, exploiting them in the presence of quantization noise will require the use of smaller, better and carefully positioned probes (possibly after decapulating the chip) to emphasize these signals and to reduce interference from other signals. In addition a specially designed comb filter or some other method could be used to suppress the contribution from the clock signal. Since the results of using this approach have already been publicized [9, 12], we focus mainly on a different approach that we employed based on exploiting *unintentional emanations*.

Experiment 2: Unintentional AM emanations: Next we took the same setup as in Experiment 1, but now the output of the probe was sent to an AM receiver, tuned to the 41'st clock harmonic at 150.88 Mhz with a band of 50Mhz. The demodulated output from the receiver was sampled with a 12-bit 100Mhz scope [9] and 100K sample points representing approximately 284 iterations were collected. Figure 3 (where the Y-axis is the magnitue and the X-axis is the frequency in KHz) plots the FFT of this signal. Now, even in this *linear* scale, the signal of interest, i.e., the 283KHz signal corresponding to the loop and its harmonics is clearly visible among the signal corresponding to the clock harmonics and the loop structure was also visible in the time-domain. Notice that these greatly improved results were obtained using the same sensor setting as in Experment 1, and with the same number of loop iterations. Note that we are also operating in a part of the spectrum which did seem to even have much signal according to Figure 1; the signals in this band were overwhelmed by the quantization noise introduced when we attempted to capture the baseband signal.

Experiment 3: Unintentional Far–Field AM Emanations: We examined emanations from an Intel–based server containing a commercial, PCI bus based SSL accelerator S[10]. We programmed the server to repeatedly invoke S to perform a 2048 bit exponentiation with a single–nibble exponent. Severl AM modulate carriers, including those at multiples of the 33Mhz PCI clock frequency and other weaker intermodulated carriers were found. In the near field, sevaral AM-demodulated carriers leaked information about the internal structure of the exponentiation even from a single sample, thus enabling a variety of attacks. Some of the stronger information bearing signals, e.g., carriers at multiples of the 33MHz PCI clock propagate to distances upto forty feet but the quality of the received information degrades as a function of the distance (due to inverse-square law) and the bandwidth (due to the thermal noise floor) being used. Figure 5 plots a signal (amplitude vs. time in ms) captured by a log–periodic antenna 15 feet away using the 299MHz carrier and 1MHz bandwidth. The invocations of S is clearly visible as a region from 7.55ms to 7.91ms where the amplitude goes below -1000. At this resolution, the macro structures within the exponentiation are already visible. At higher resolutions, there is enough information to enable template attacks [2].

Experiment 4: Unintentional Angle Modulated emanations: Next we turned on the variable internal clock DPA protection mechanism in Smartcard A and kept everything else, including the loop and the

[9] One advantage of using low sampling rate is that higher precision sampling equipment is available

[10] S is rated to perform 200, 1024-bit CRT based RSA private key ops/s.

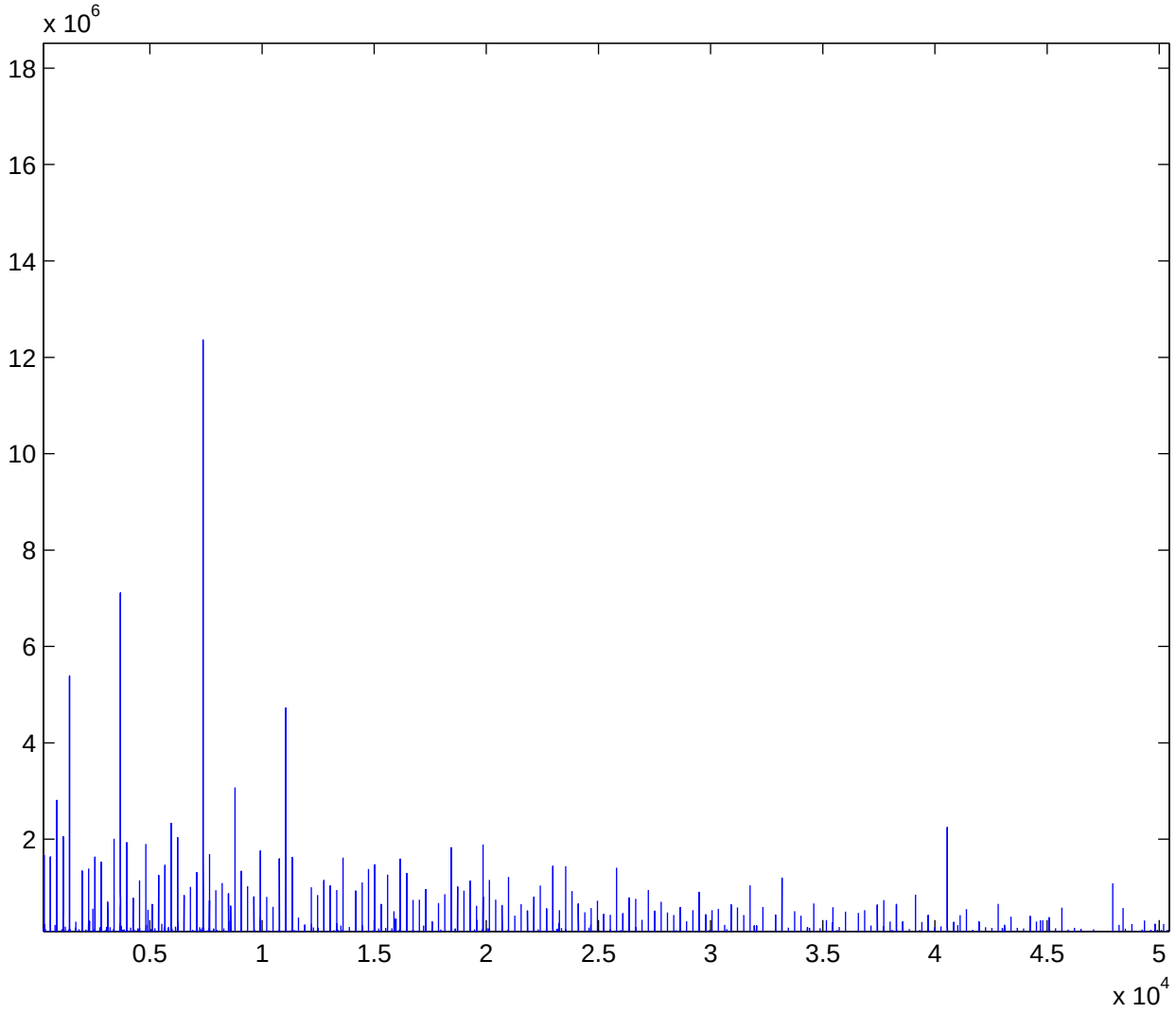

Figure 3: FFT of demodulated signal (150.88 Mhz carrier, 50Mz band) in Experiment 2 with Smartcard A

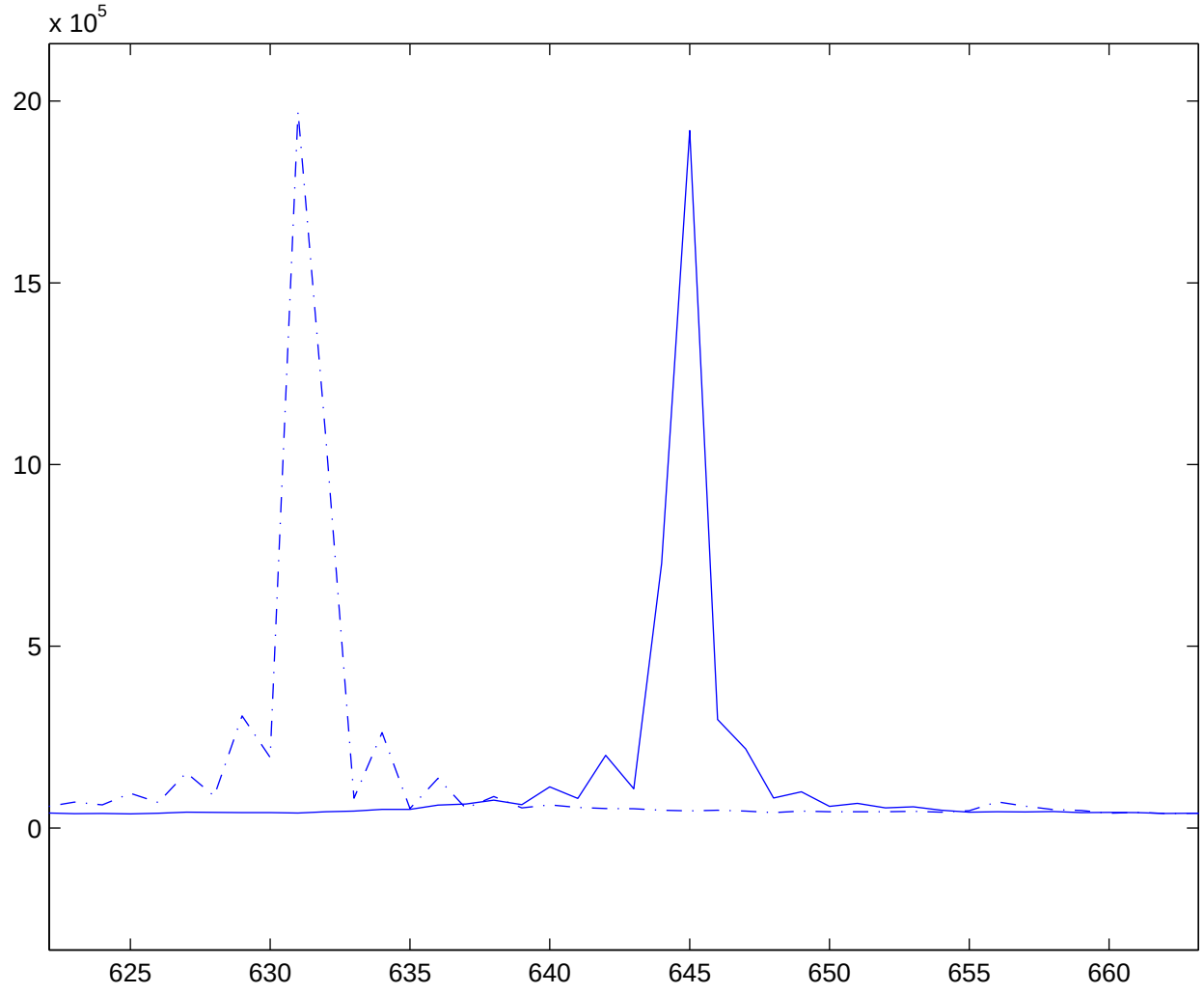

Figure 4: Two FFTs showing difference in loop frequency for LSB=0 and LSB=1 for smartcard A

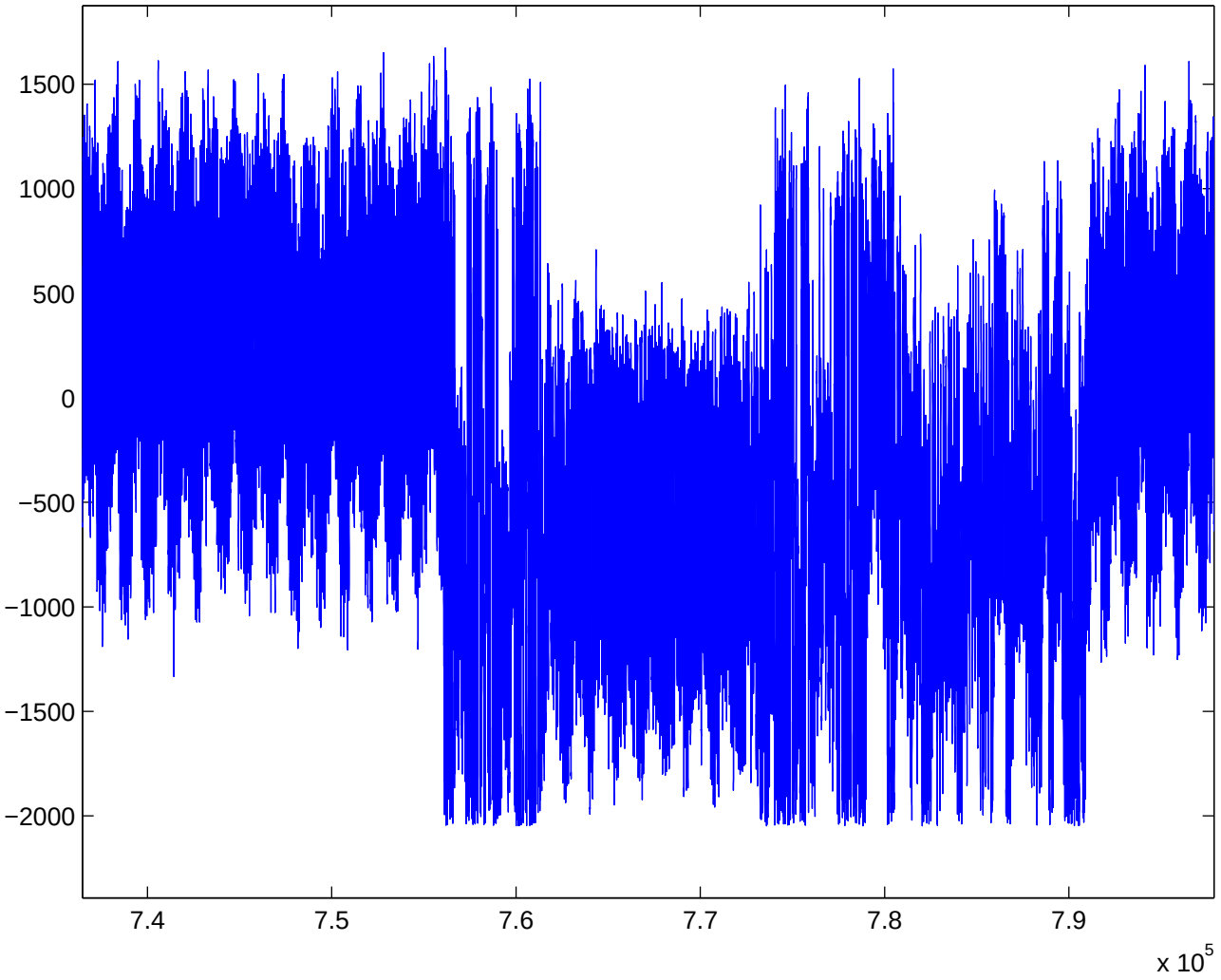

Figure 5: EM Signal from SSL Accelerator S

sensor position the same as Experiment 1. One of the instructions in the 13-cycle loop was to load a user supplied byte B from RAM to accumulator. We found a carrier where by AM-demodulation one could clearly see the card executing a loop. We experimented with different values of the byte B and made the following surprising observation: The average frequency of the 13-byte loop was dependent on the least significant bit (LSB) of B but not on other bits. This is shown in Figure 4, where is the magnitude of FFT of the EM output for two different cases (with the X axis being frequency in KHz). The first case (shown by a broken line) shows the loop frequency with the LSB of B being 1 and in the second case (shown by a solid line) the loop frequency executes with LSB of B being 0. When the LSB is 1 the loop runs slower. This is due to some coupling between the LSB of the data line and the circuitry generating the internal clock. Although the clock timing itself varies very often, when there is a 1 bit on the line, we found that this intrinsic variation gets biased towards slowing down the clock for a couple of subsequent cycles. We speculate that this is becuause the clock circuitry draws energy from the same source as some other circuitry affected by LSB. Thus, angle demodulation, e.g., FM demodulation, turns out to be a good avenue for attacking smartcard A using LSB based hypothesis, effectively transforming a countermeasure into a liability. Another advantage of such angle demodulation based attacks is that the internal clock signal is very strong and observable at a distance.

2.3 Information Leakage across EM Spectrum

In this section we provide experimental evidence to reinforce a central theme of this paper, i.e., the output of even a single wideband EM sensor logically consists of multiple EM signals each carrying qualitatively different compromising information. In addition, certain leakages can be substantially superior in some EM signals as compared to the power consumption signal.

Whereas the presence and locations of certain types of EM signals (e.g., angle modulated carriers, intermodulated carriers etc) are very device dependent, our experiments show that universally, AM carriers at harmonics of the clock frequency are a rich and easily accessible source of compromising information. For smart cards, since the fundamental freqency is so low, the intermediate harmonics are usually the best. Lower harmonics suffer from excessive noise and interference and higher harmonics tend to have extremely low signal strength[11].

In this section, we examine the leakage of information from four types of signals obtained from a smartcard, which we call smartcard B (to protect vendor identity[12]), while it performed DES in software with no power analysis countermeasures, except for the internal noise generators being on. The smartcard ran on the 3.68Mhz external clock. Three of these signals were obtained by AM demodulating the output of a near field probe placed as in Experiment 1, at three different intermediate carrier frequencies (50Mhz bands around 188Mhz, 224.5 and 262Mhz). The fourth signal was the power consumption signal. All signals were collected by a 12-bit, 100Mhz digital scope.

It is well known that plotting the results of a differential side channel attack launched against a bit value used in a computation is a good way to assess the leakage of the bit [11]. This is not surprising since this plot is essentially the difference between the average of all signals in which the bit is 1 and the average of all signals in which the bit is 0, plotted against time. At times in the computation where this bit is not involved or at points in the computation where this bit is involved but that information does not leak in the side-channel, the value of the difference would be small and not noticeable. At

[11]This is because clock edges are not very sharp in practice.

[12]Smartcard B is a 6805-based, 0.7micron, double metal technology card with inbuilt noise generators.

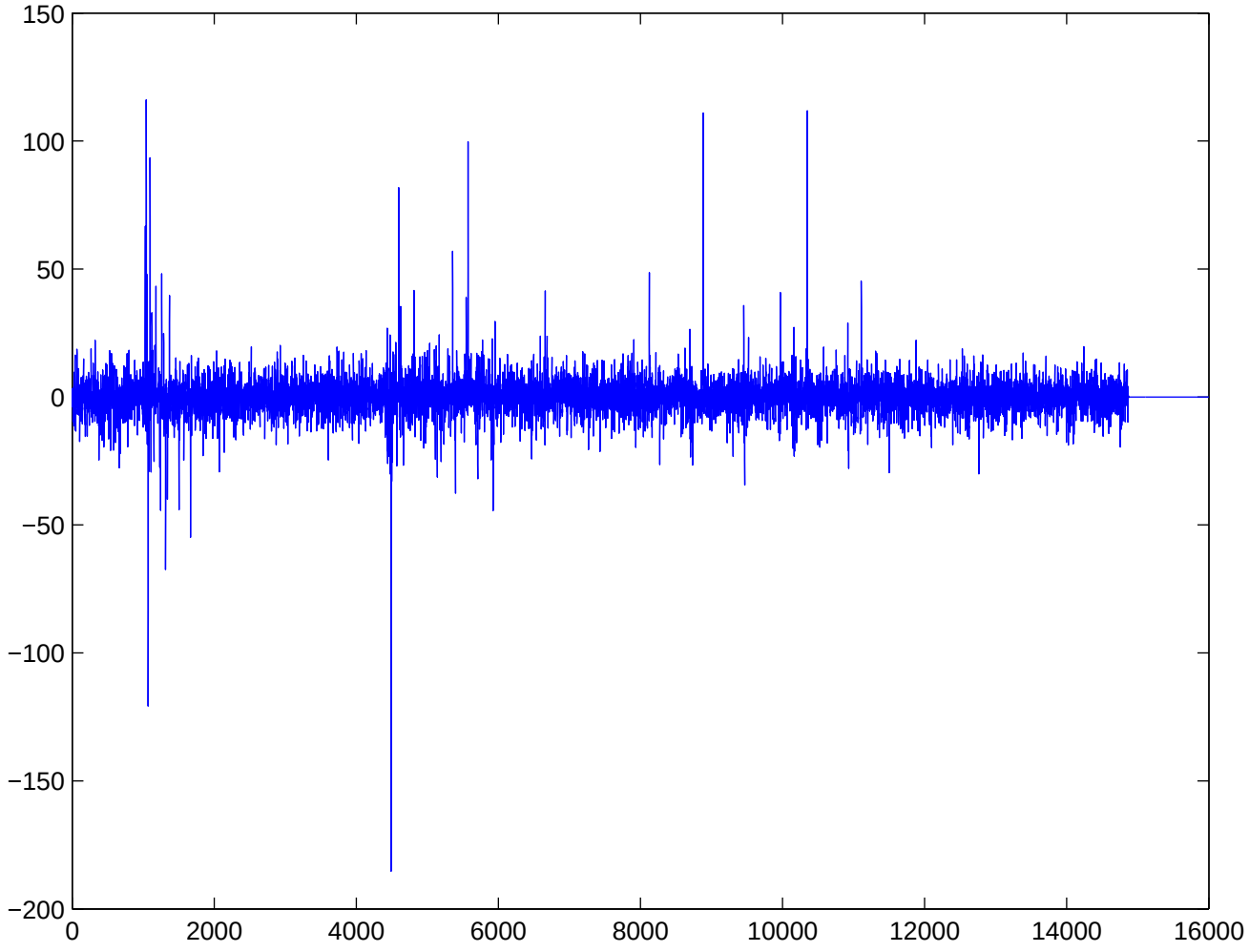

Figure 6: DEMA attack on DES on smartcard B using the 224.5 Mhz carrier

points where the bit is used in the computation *and* this information leaks in the signal, this difference is likely to be large or noticeable.

Figures 6, 7, 8, and 9 show the results of a differential side-channel attack on an S-box output bit in the first cycle of the DES implementation, using the four different signals. Figures 8, 6, and 7 are for the EM signals and Figure 9 is for the power signal. All figures are aligned in time, the X-axis shows elasped time in 10ns units (due to 100Mhz sampling) and the Y-axis shows the difference in the averages of signals with bit=0 and bit=1 for 2000 invocations of DES with random inputs. Even at this resolution it is clear that the leakage results are qualitatively different from each other. There are some gross similarities between the EM leakages in Figures 6 and 7 and between the power leakage in Figure 9 and the EM leakage in Figure 8.

These leakages can be viewed by plotting them all together. Figures 10, 11, 12 show some of the regions in such a plot. Each leakage is plotted in a different line-style, with the power leakage being a solid line and the 3 EM leakages plotted in different broken-line styles (188Mhz with a dotted line, 224.5Mhz with a dashed line 262Mhz with alternate dot and dashes). It is clear from these figures that even though the signals fall into two gross classes at the macro level, there are significant differences

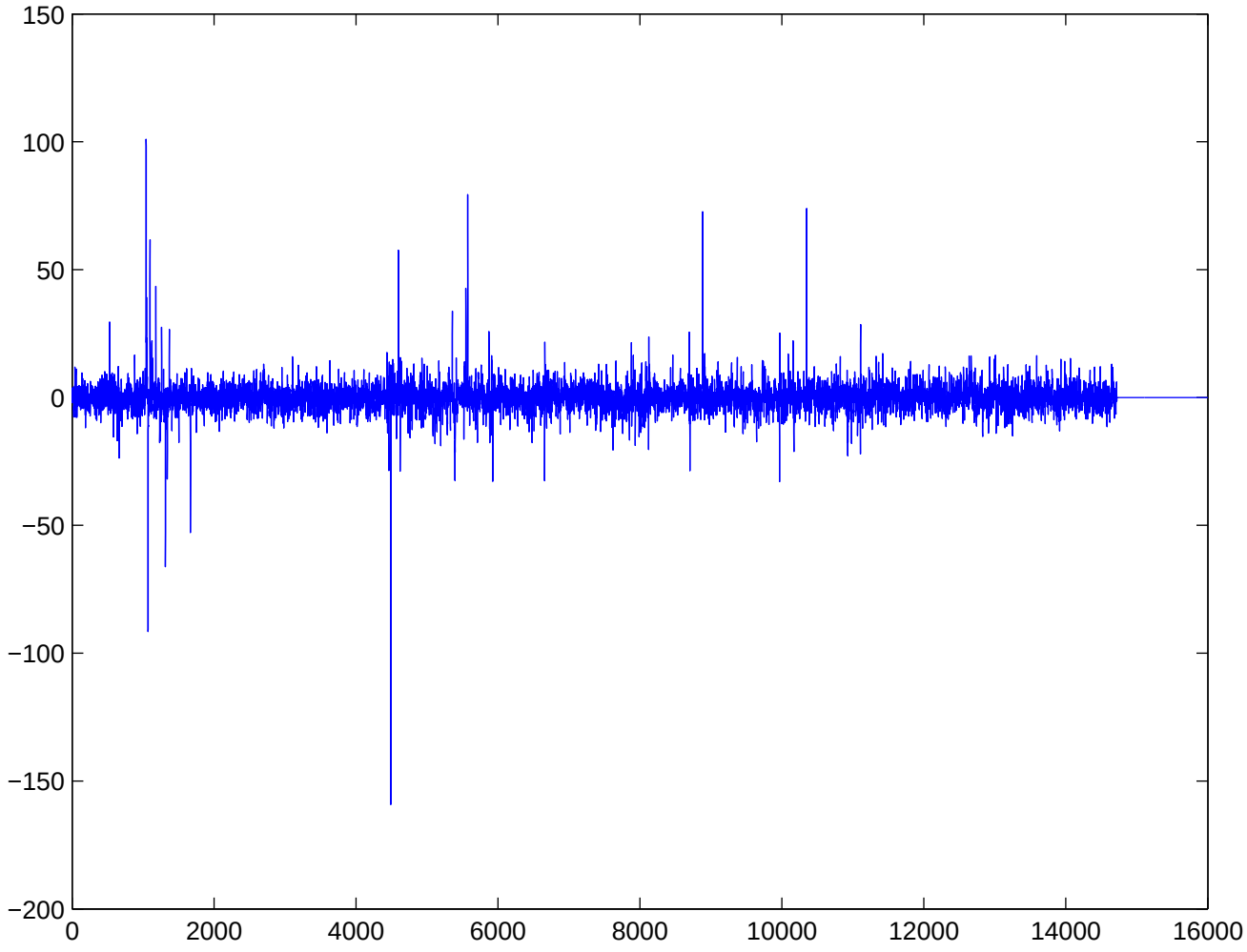

Figure 7: DEMA attack on DES on smartcard B using the 262Mhz carrier

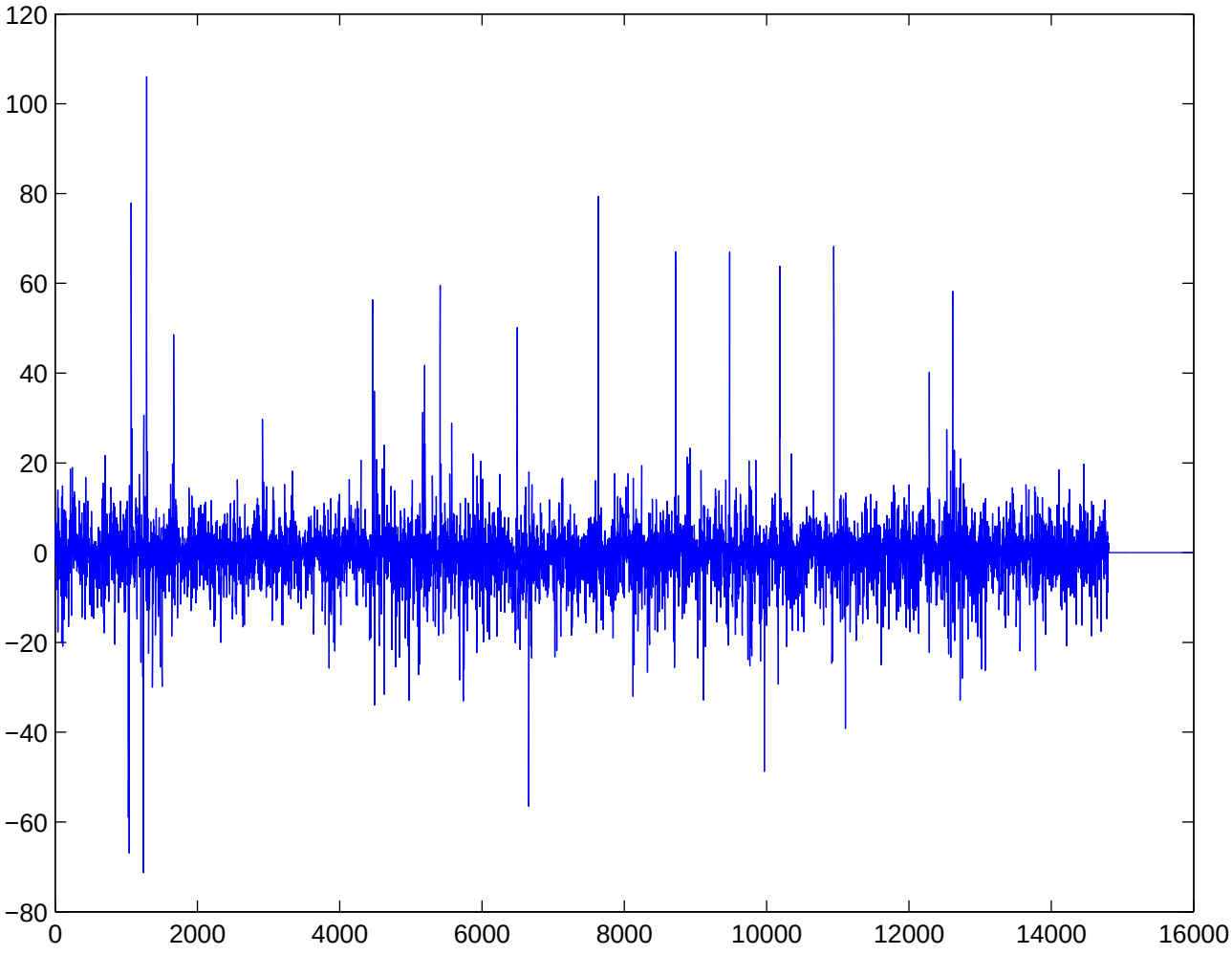

Figure 8: DEMA attack on DES on smartcard B using the the 188Mhz carrier

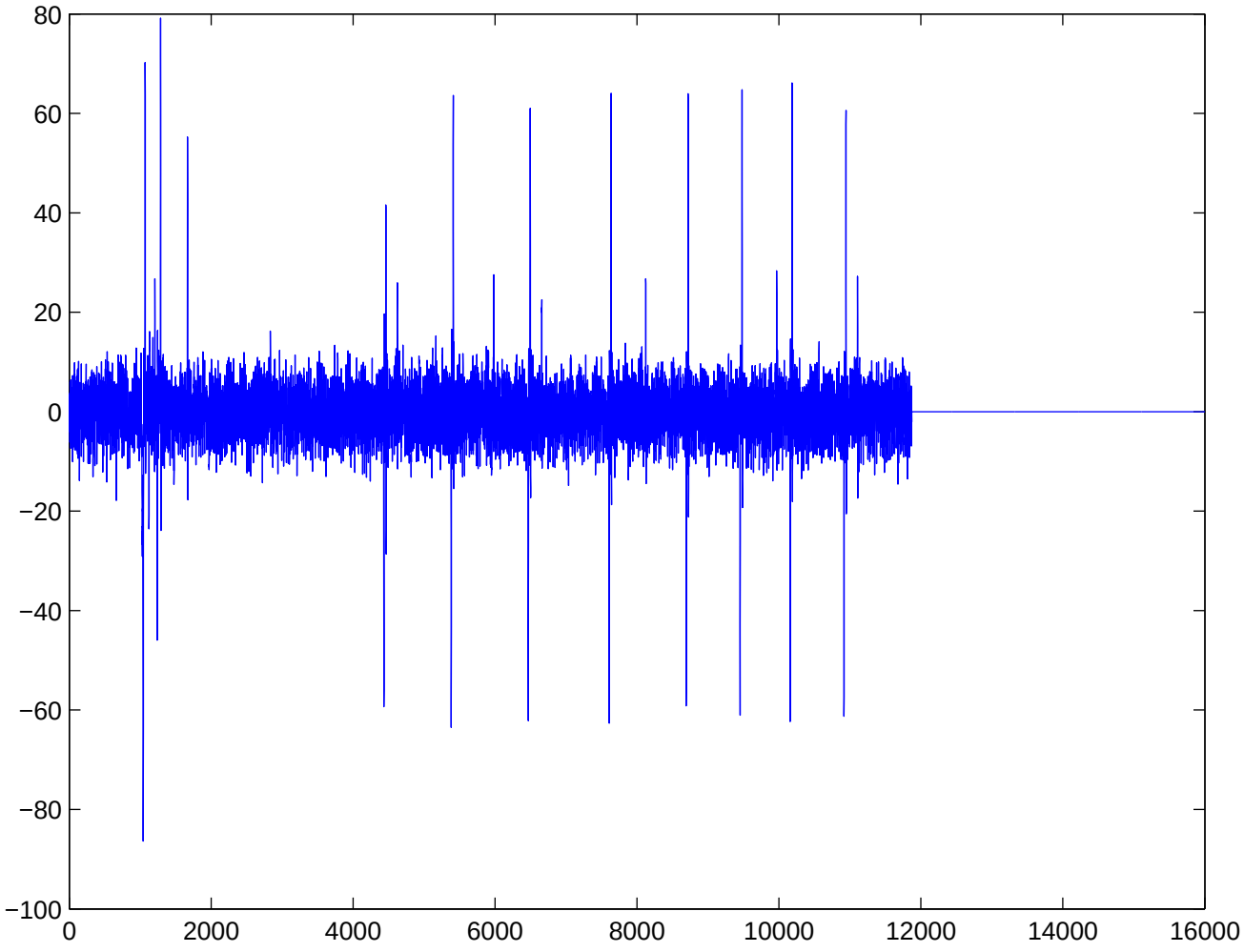

Figure 9: DPA attack on DES on smartcard B

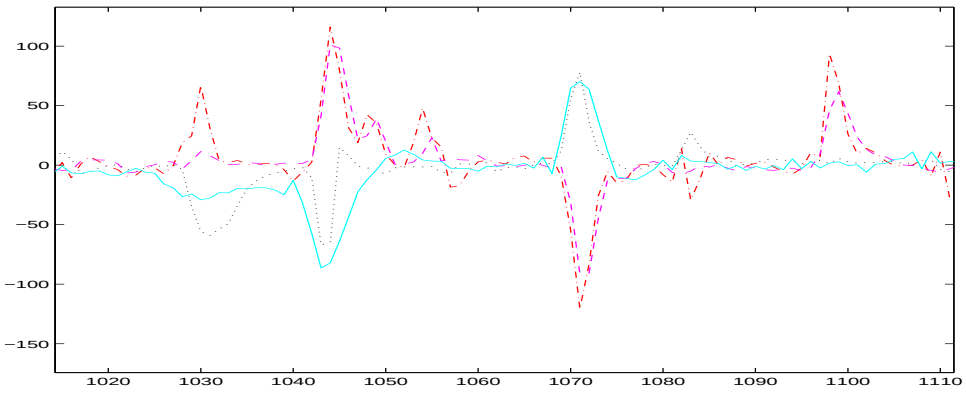

Figure 10: Comparison of DEMA/DPA Leakages at region 1020–1110

even between signals within a class at a cycle level (see Figure 10). Moreover, there are leakages which appear in EM signals (and sometimes excessively so), which do not appear in the power signal (see Figure 11). Such leakages are due to a what we will later term as a "bad" instruction. There are also leakages which are large in power, but low in some (but not all) EM signals (see Figure 12).

2.4 Propagation and Capture of EM signals

EM signals propagate via radiation and conduction, often by a complex combination of both to eventually emerge from the device. This naturally suggests the usage of two classes of sensors to capture the signals that emerge. Radiated signals are best captured by strategically placing near field probes or antennas around the device.

For best results the probes/antennas should be as close as possible or at least in the "near-field", i.e., no more that a wavelength away[13], although a few of the emanations can also be captured from larger distances.

Conductive emanations consist of faint currents found on all conductive surfaces or lines attached to the device. Sometimes, these currents ride on top of stronger, intentional currents flowing within the same conductors. Capturing these emanations requires current probes similar to those used for power analysis and subsequent signal processing to extract them from the stronger signals. In fact, if the researchers experimenting with power analysis attacks were to re-analyze the raw signals from their current probes, they will discover that apart from the relatively low frequency, high amplitude power consumption signal, there are faint higher frequency AM modulated carriers representing conductive emanations as well. For example, Figure 13 shows the current signal during 3 rounds of DES in smartcard B captured with 12-bit, 100Mhz scope (amplitude on Y axis, X axis is elasped time in 10ns units). However, the same power line is also a conductor entering the chip and hence also carries faint currents due to conductive EM emanations. Conductive EM emanations at low frequencies will get lost amongst the larger currents driving the card, but it is quite easy to obtain several EM emanations at higher frequencies. Figure 14 shows one such EM signal extracted from the same power line by AM demodulation of an intermediate frequency carrier during 3 rounds of DES, captured by a 12-bit, 100Mhz scope.

We have found that the most effective near field probes are those made of a small plate of a highly

[13]For emanations at 300 MHz, the wavelength is 1 meter.

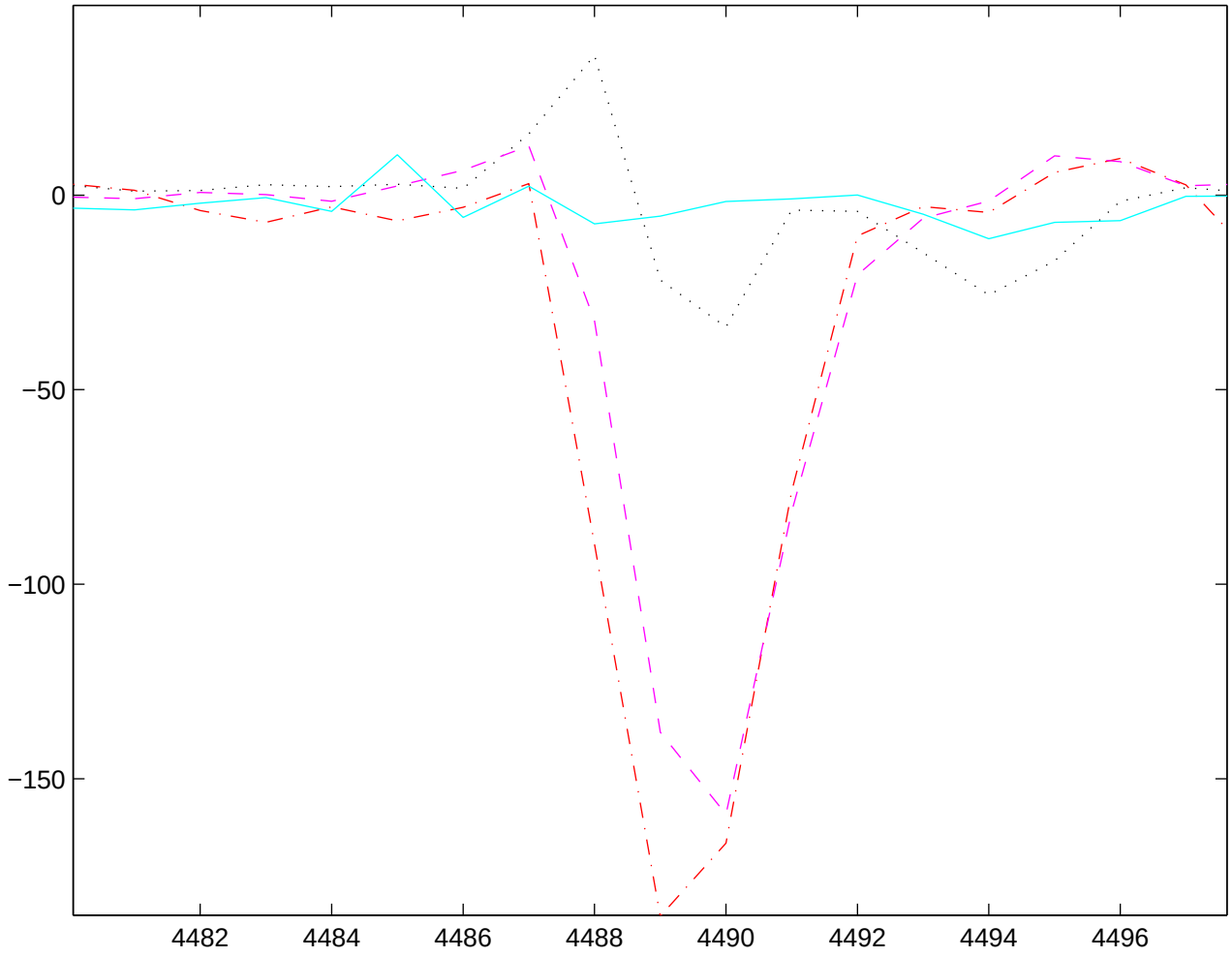

Figure 11: Comparison of DEMA/DPA Leakages at region 4482–4496

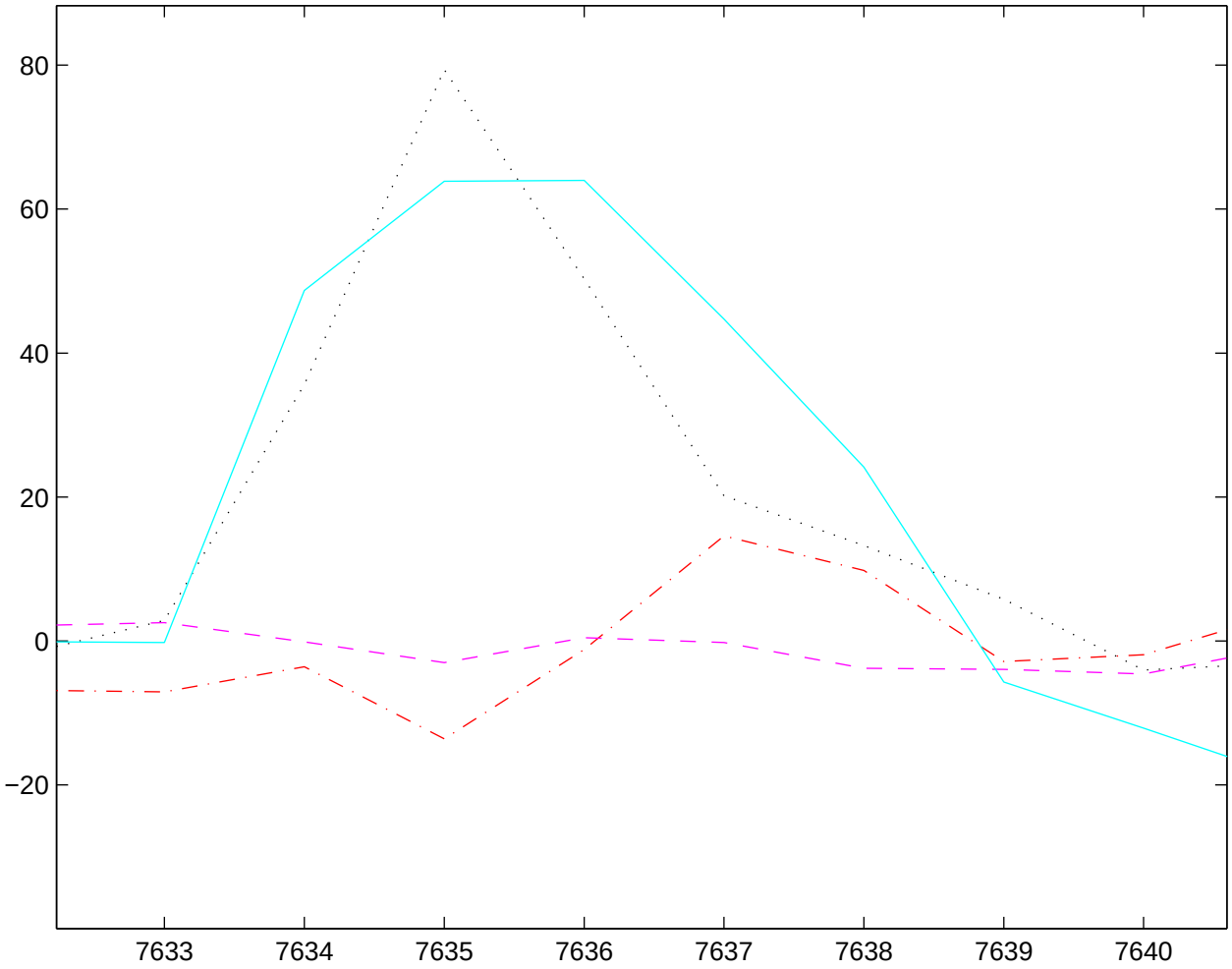

Figure 12: Comparison of DEMA/DPA Leakages at region 7633–7640

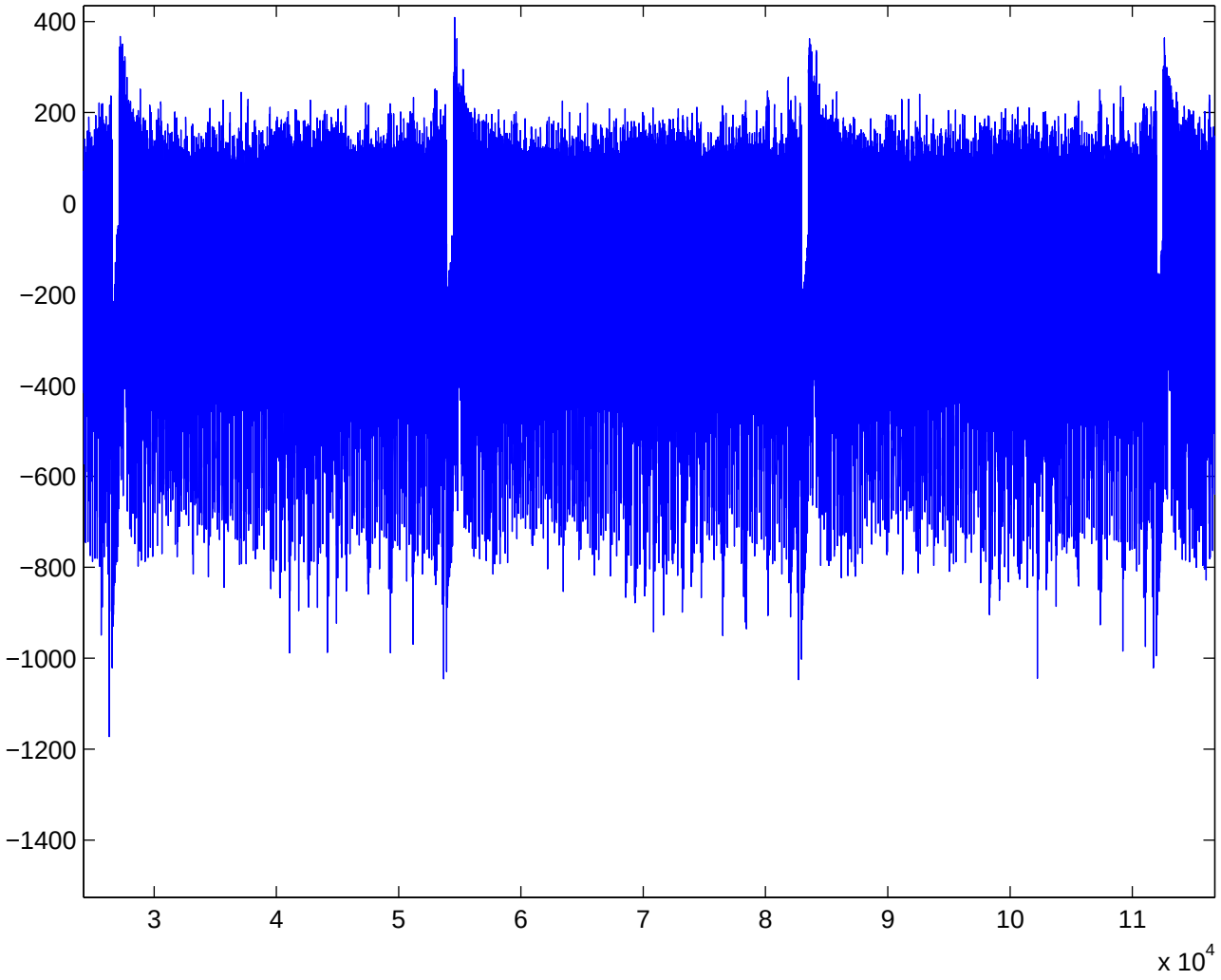

Figure 13: Baseband Power Signal for 3 rounds of DES on smartcard B

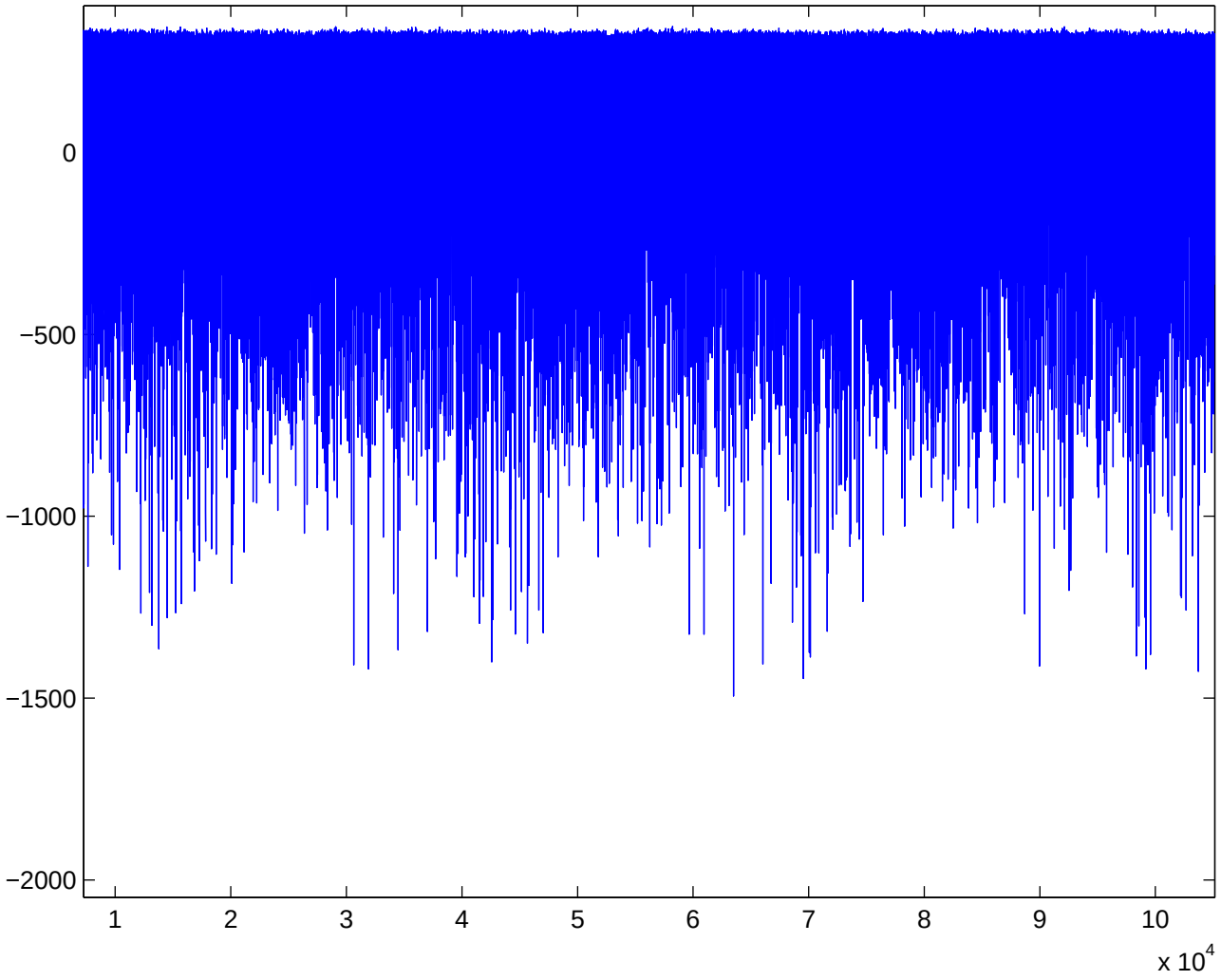

Figure 14: EM Signal on Power Line for 3 rounds of DES on smartcard B

conducting metal like silver or copper. In the far field, we have used both wideband antennas such as biconical antennas for lower frequencies and log-periodic for higher frequencies. In some cases, once a useful carrier is identified, one can also hand-craft narrowband Yagi antennas to improve its reception from a distance. Although, it is better to shield the equipment from ambient EM emanations, this shielding does not have to be elaborate; it is far more productive to make sure is no strong source of the interfering emissions (in the band of interest) located close the the device to be attacked.

The emanations captured by each sensor need to be processed to extract compromising information. For direct emanations, filters may suffice. For unintentional emanations, which manifest themselves as modulations of carrier signals of various frequencies, a wide bandwidth (preferably tunable) receiver/demodulator is convenient. Examples of such receivers include the R-1550 Receiver from Dynamic Sciences [8] and the 8617 Receiver from Watkins–Johnson. Typically, these receivers have a wide frequency range (upto 1GHz) and bandwidth from 40Hz to 200MHz depending on the options. If such equipment is not available, one alternative is to use cheaper wideband radio receivers which have intermediate frequency output (e.g., ICOM 7000 or 8500), or to even construct using commonly available low noise electronic equipment such as signal generators, mixers, band pass filters etc a hetrodyning system which will down-convert the band of interest on to a lower intermediate frequency.

The filtered/demodulated signal from a receiver or the intermediate frequency output of a receiver/hetrodyning circuit can be captured by equipment identical to that used for power analysis attacks, such as a digital sampling board and/or oscilloscope, which can have high precision since sampling rate does not need to be very high. For intermediate frequency output, subsequent processing such as additional filtering and demodulation will have to be done in software.

Obtaining multiple EM signals generally requires multiple receivers and signal capturing equipment but if very high precision signal capturing equipment is available then one may be able use a smaller number of receivers and signal capturing equipment by performing signal separation in software after capturing emanations within a very wide band. Equipment such as spectrum analyzers are also useful for quickly identifying carriers and potentially useful emanations. A useful rule-of-thumb is to expect strong carriers at odd harmonics of the clock.

3 EM Attacks With Low Cost Equipment

We now present some simple yet remarkably powerful attacks that use low cost equipment. These attacks require *only one* receiver and the capability to capture *only one* signal at a time. Even with this restricted setup, one can perform extensive experimentation with a device to identify and assess a large number of EM signals for their leakage characteristics. This enables the design of attacks exploiting each type of observed leakages. The simple and intuitive attack strategy is to pick an EM signal with the most leakage relative to noise.

As the reader may have anticipated, we found several EM signals for each tested device where the classical side-channel attacks such as Simple Electromagnetic Attacks (SEMA) and Differential Electromagnetic Attacks (DEMA) [12] could be performed on algorithms like DES. Section 3.1 illustrates results of such attacks on a chipcard. While these attacks are interesting, by themselves, they do not justify why EM side-channel(s) should be used in preference to others. A good justification would be if EM leakages were somehow superior, or better still, if they could be used to break implementations secure against power and timing attacks. Therefore in Section 3.1 we also show that it is indeed the case for some intructions that we term "bad instructions" and in Section 3.2 we show how these can be used to defeat power analysis countermeasures.

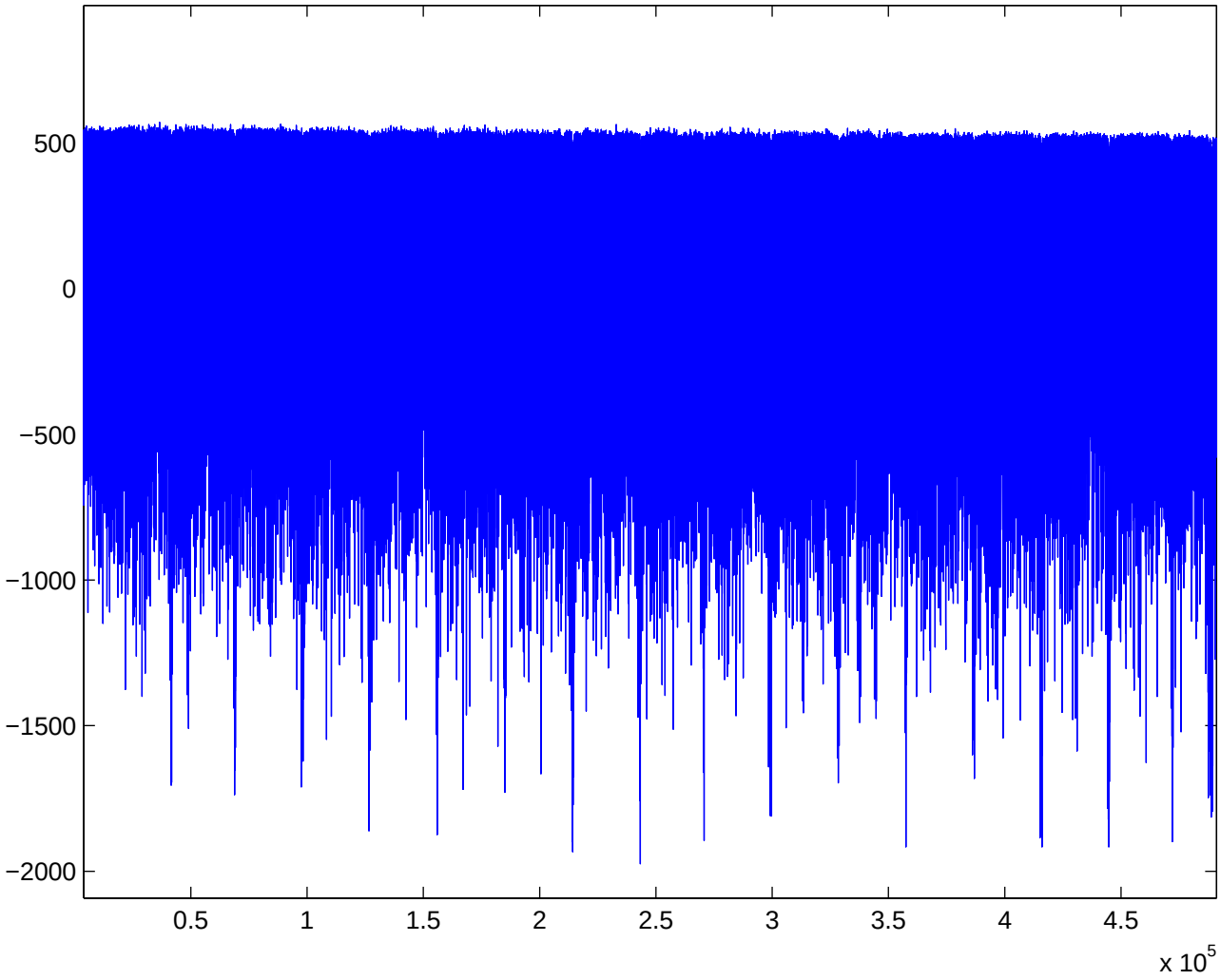

Figure 15: EM Signal showing 16 DES rounds from smartcard B (AM demod 262Mhz carrier, 50Mhz band

3.1 SEMA, DEMA and Bad Instructions

The terms Simple and Differential Electromagnetic Attacks, abbreviated as SEMA and DEMA, were introduced by Jean-Jacques Quisquater at numerous rump session talks at Eurocrypt '00, Crypto '00 and CHES '00. In this section, we describe some of the SEMA and DEMA results we obtained on certain chipcards. Our SEMA attack on a chipcard will be based on a "bad instruction", i.e., an instruction which leaks much more information in an EM signal than in the power signal. Before describing the attack, we briefly discuss the information present in many compromising EM signals.

3.1.1 Information in Compromising EM Signals

Just as in the power, a compromising EM signal contains information about the computation done on the chipcard at various levels of granularity. Consider the setup where smartcard B is performing DES in the setup described in Section 2.3 and the EM signal during the computation captured by a 12-bit 100Mhz scope after AM demodulating the probe signal at the 262Mhz carrier with 50Mhz

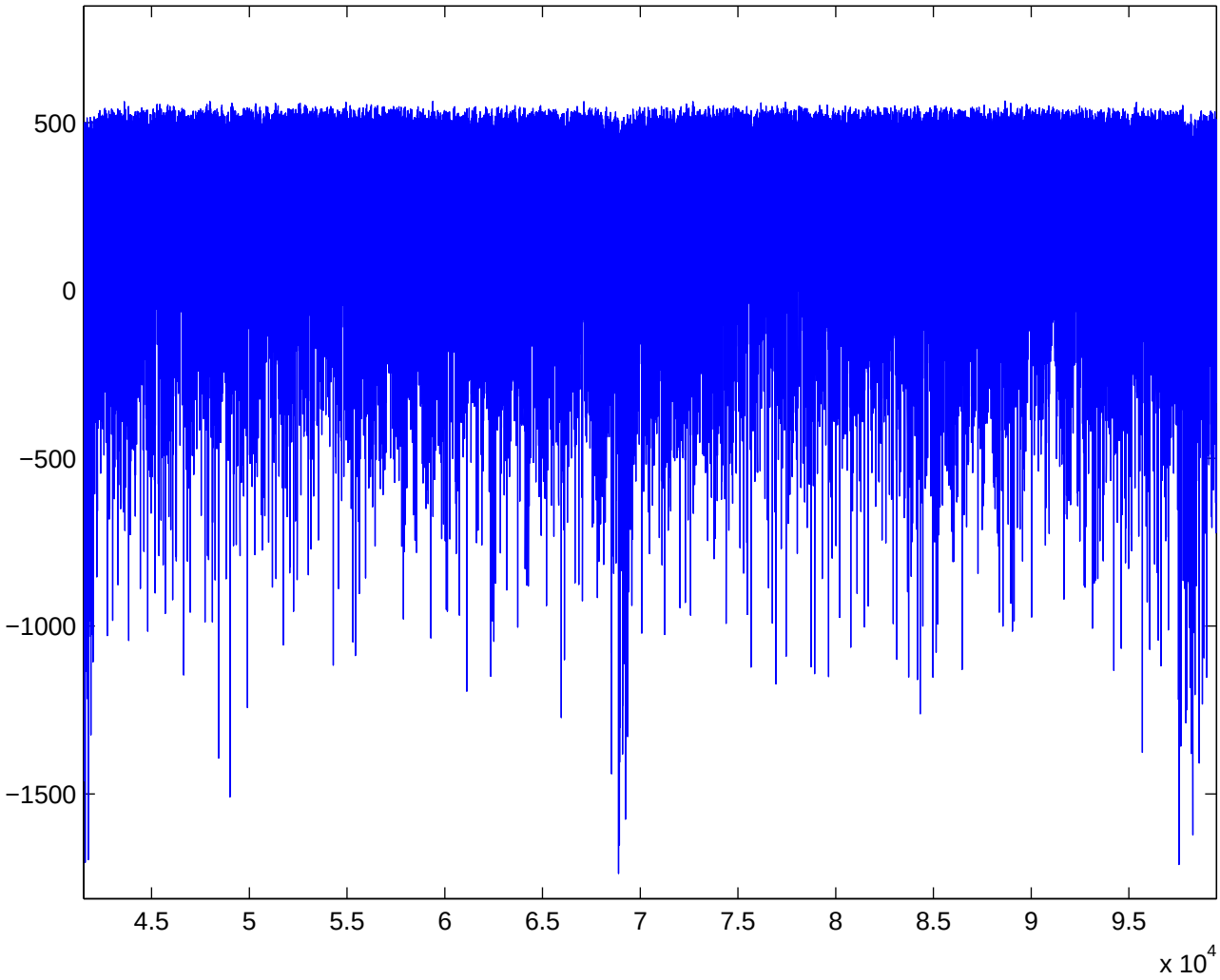

Figure 16: EM Signal of two rounds of DES from smartcard B

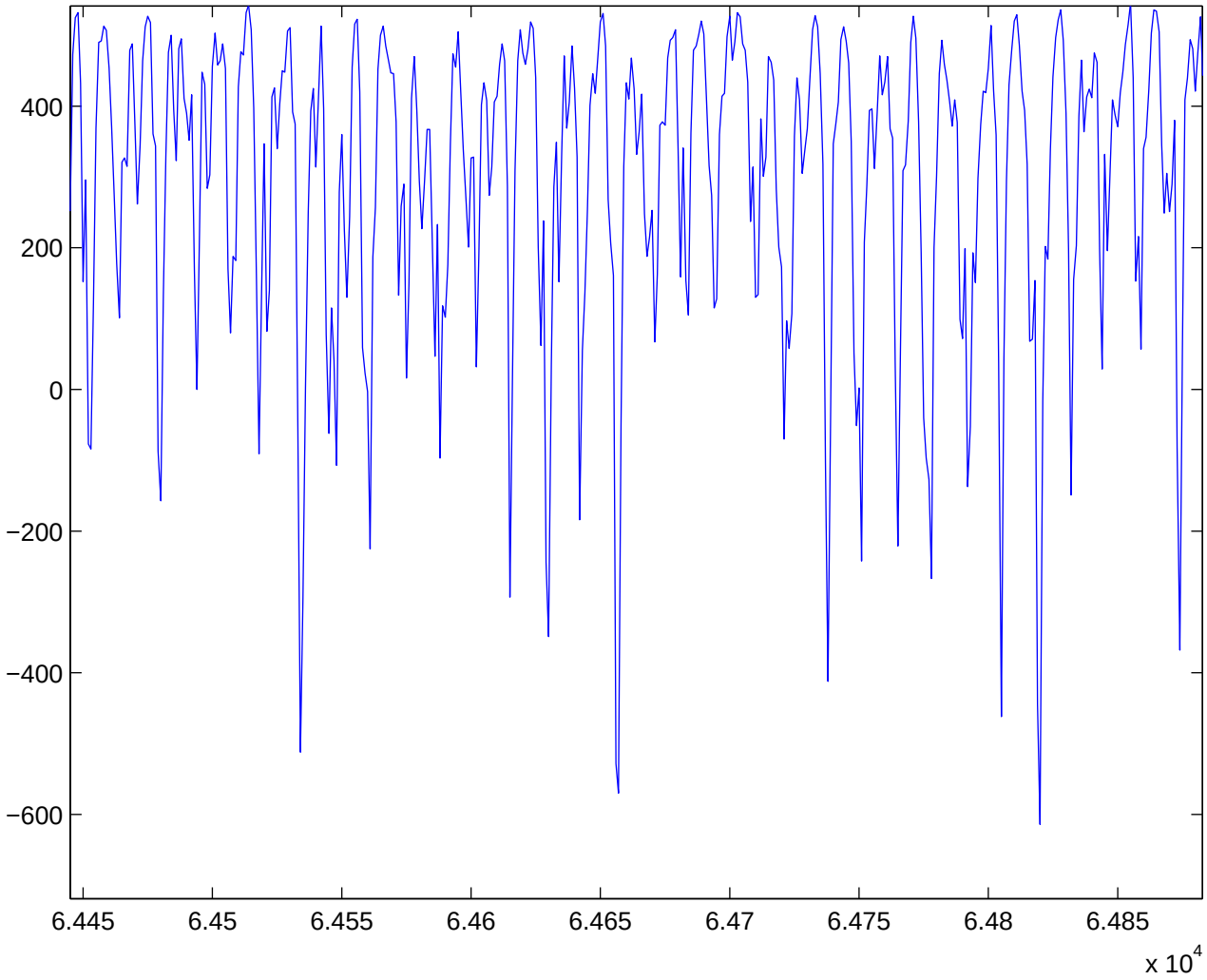

Figure 17: EM Signal showing cycle level information from smartcard B

band. Figures 3.1.1,3.1.1 and Figure 3.1.1, shows the caputured signal (plotted as amplitude vs time in 10ns units) at different time scales. At a macroscopic level, in Figure 3.1.1 one can see the overall structure of the computation, i.e, 16 similar sized structures terminated by sharp negative peaks for each of the 16 rounds of DES. At an intermediate level of resolution, Figure 3.1.1 shows two rounds of DES. In each round, there are three regions with different densities of negative peaks. These regions can be contrasted to the regions visible in Figure 13. At a microscopic level, one can see emanations at the clock cycle level as shown in Figure 3.1.1.

3.1.2 SEMA

In a SEMA attack, an adversary is able to extract compromising information from a single EM sample. If a computation makes use of conditional branches based on secret information, then on a compromising EM signal, this can be observed as relative shifts in the distances between major computational structures. In some cases, these shifts may be sufficient to reveal the branch taken, which in turn confirms the value of the secret information. This is analogous to what has already been demonstrated for power samples [11]. Thus conditional statements in the code could provide valuable opportunities for both SPA and SEMA.

In our opinion, the interesting case is where SEMA attacks are successful in extracting information whereas SPA attacks fail. This is possible if an EM signal for some instruction leaks more information than the power side-channel. The following experiment which shows the existence of such "bad instructions" confirms this possibility.

Bad Instructions:

In the following set of figures, we consider smartcard B in which the internal noise generators had been turned off. In such a setting, we observed that an instruction which tests a bit of a byte in the memory leaks information about the tested bit from even a *single* signal sample in the EM side-channel but not in the power side-channel. Each figure plots the amplitudes of two signals with respect to time (in 10ns units).

Figure 18 shows two EM signals in which the bits tested are both 0, In both figures the data was collected by a 12bit, 100Mhz scope after demodulating at the 262Mhz carrier. This is seen as a low value in both the signals at the point 18915. Figure 19 shows two EM signals in which one of the bits tested is 0 and the other is 1. This is seen as a low value in one of the signals and a high value in the other at the point of interest which in this case is 18780. The corresponding figures for the power side-channel are shown in Figures 20 (at point 19260) and 21 (at point 18990) respectively. The power signal levels, at the corresponding points where the EM emanations differed widely, are very close. This was also verified by taking averages of 1000 power samples. The experiment once again confirmed that the averaged signal at the point of interest was almost identical for the 0 and 1 bit.

Even with noise enabled, it was possible to classify the bit value correctly with high probabilty by using only a few samples (20–30). Section 4.2.1 illustrates this with an example. This example also shows the value of EM side-channels—this bit value had no easily observable leakage in the power side-channel and even statistical attacks required several thousand samples.

Since the last four figures were obtained by considering individual samples and a skeptical reader may remain unconvinced, we now present statistical observations over 2000 samples for three cycles of the same "bad" bit-test instruction as opposed to the approximately 1.5 cycles illustrated in the last four figures.

Figure 22 shows two signals corrresponding to mean EM signals (each taken over 1000 samples)

obtained during the execution of a "bad" bit-test instruction when the tested bit was 0 and 1 respectively. As can be seen, the mean signals differ significantly at a few places, namely those at which the tested bit influences the EM emanations.

Figure 23 shows two signals corresponding to the standard deviation of the EM signals (each taken over 1000 samples) obtained during the execution of a "bad" bit-test instruction when the tested bit was 0 and 1 respectively. It turns out that while the means and standard deviations of the EM signals are sufficient to distinguish the tested bit with non-negligible probability (as was shown in Section 4.2.1), the corresponding values for the power signals are not.

3.1.3 DEMA

The analogy for differential power analysis (DPA) is DEMA. DEMA results from 3 different EM signals are presented in Section 2.3.

3.2 Defeating Power Analysis Countermeasures

In [11], a suggested countermeasure to power analysis is to use only those instructions whose power leakage is not excessive and to refresh sensitive information, such as a key, after each invocation in a non-linear fashion. This forces the adversary is extract a substantial portion of the key from a *single invocation* since incomplete key information does not help in subsequent invocations. Another class of countermeasures based on splitting all sensitive information into shares was proposed in [1, 6]. The basic idea here is that uncertainty in the information about each share is exponentially magnified in proportion to the number of shares. The key to breaking both classes of countermeasures is to identify "bad" instructions which leak much more information in some EM side-channels than the traditional side-channels. Such instructions if used in power-analysis resistant implementations would subvert the very assumptions supporting their resistance. Similarly, excessive leakage in intstructions dealing with shares would diminish the very uncertainity being magnified by the shares leading to a compromise of sensitive information.

For all chip cards that we examined, there were several such instructions. In our investigations, we did not find any instruction that leaked in the power side-channel but did not leak in some EM side-channel. This can happen if all critical parts of a chipcard are well-shielded but the power signal is not. We feel that this is unlikely since a designer who shields EM emanations so well is also likely

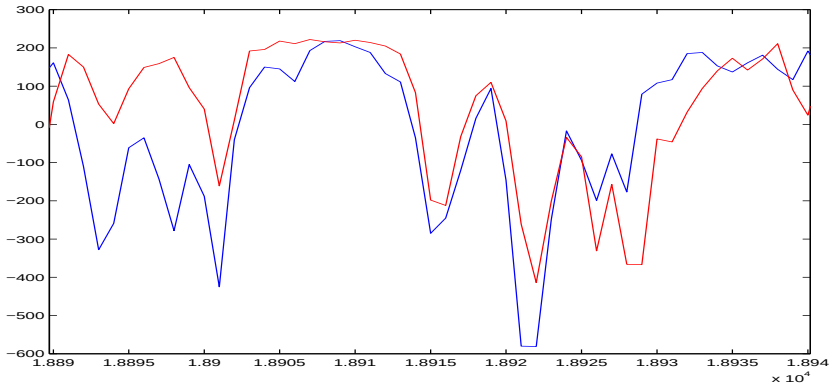

Figure 18: Two EM Signals where tested bits are 0 (seen as low values at 18915)

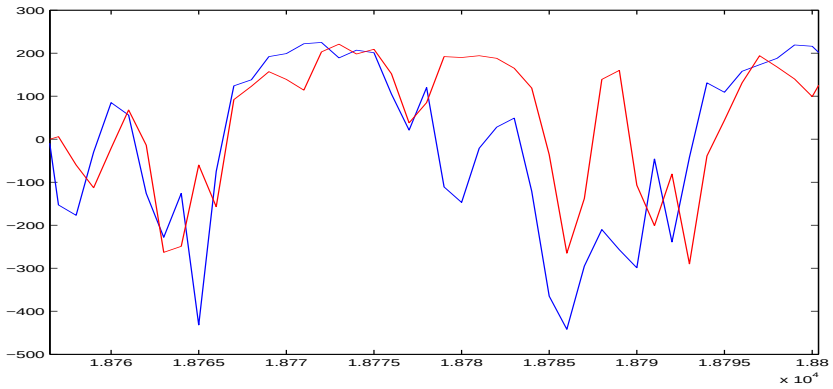

Figure 19: Two EM Signals where tested bits are 0 and 1 (seen as low and high values at 18780)

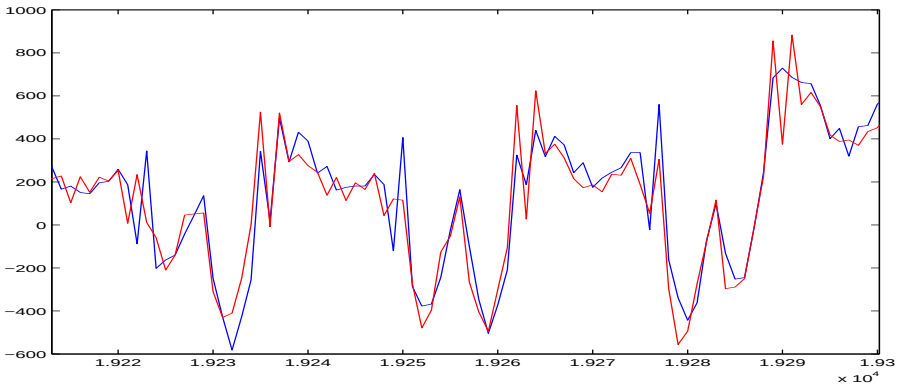

Figure 20: Two Power Signals where tested bits are 0 at 19260

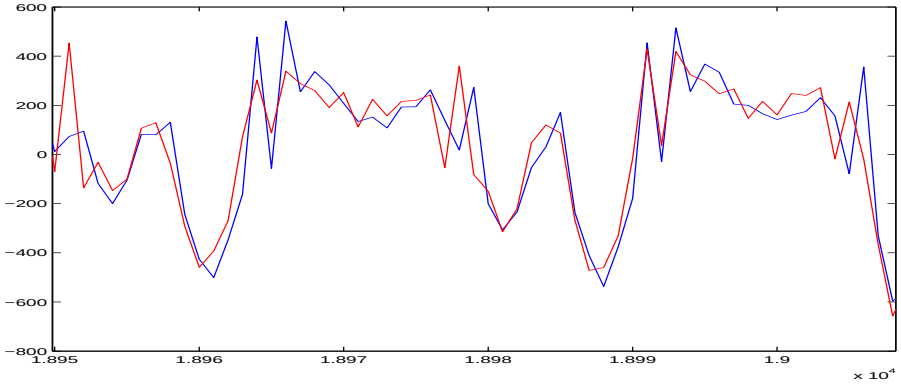

Figure 21: Two Power Signals where tested bits are 0 and 1 at 18990

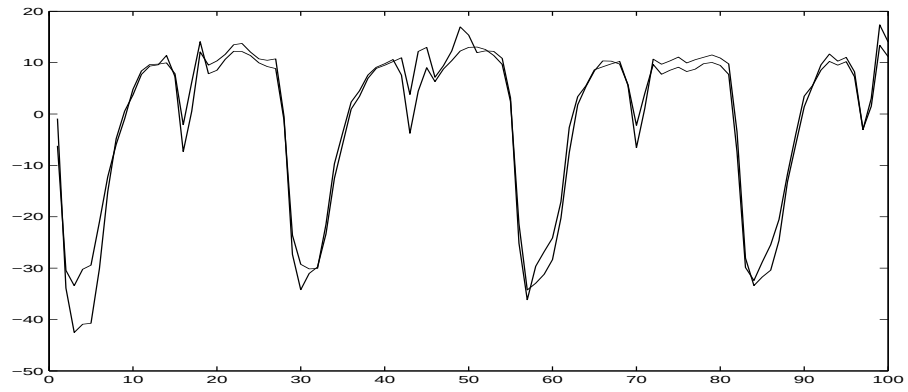

Figure 22: Mean signals over 3 cycles for a bad instruction where the computation differs only in one bit

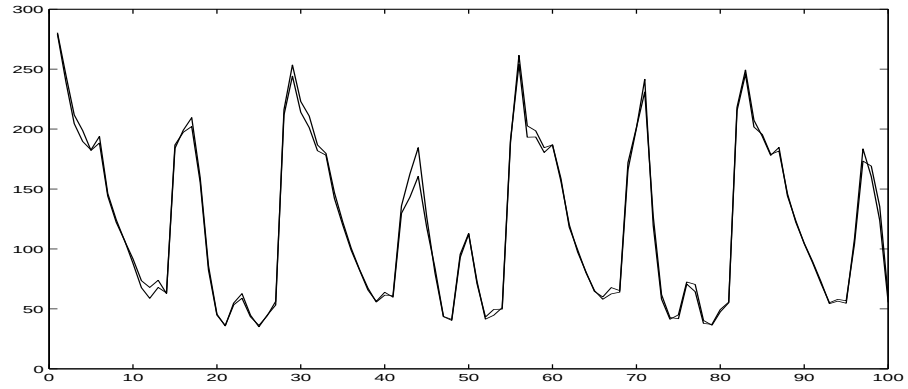

Figure 23: Standard deviations of signals over 3 cycles for a bad instruction where the computation differs only in one bit

to protect against power signal leakages. In the absence of shielding, we *believe* that a power signal leakage implies leakage in EM emanations due to the physics of the semiconductor devices.

The bit test instruction is a very useful instruction for implementing algorithms, such as DES, which involve bit level permutations. For example, it can be used for key expansion and P-permutation. Based on its low leakage characteristics in power, there is no reason for not using such a useful instruction in power analysis resistant implementations. If this instruction is used for this purpose on this card with noise disabled, then a SEMA attack would be sufficient to extract the DES key regardless of which class of countermeasures [11, 1, 6] was used. However, if noise was enabled, then the countermeasure of [11] may still work while feasible higher order statistical attacks would still defeat the countermeasures of [1, 6].

3.3 Higher Order EM Attacks on Secret-Sharing

DPA countermeasures based on secret-sharing schemes choose an appropriate value for the number of shares based on the leakage characteristics and the desired level of resistance against higher order power analysis [11], in terms of the number of samples required to break the implementation. If a leakage is superior in an EM signal, then the number of samples for the corresponding higher order

EM attack will be substantially lower.

To verify this, we implemented a two-way XOR-based secret sharing scheme for bits on the chipcard where the bit test instruction (described in the previous section) leaks more information on an EM signal. This sample code split the input bits into pairs of shares and tested the values of the bits of the shares using the bit test instruction. We confirmed that DPA and DEMA did not work, i.e., no single point in the power/EM signal correlated with any of the input bits. We then performed a second order DEMA attack using 500 EM signals. Specifically, we defined a statistical measure on the signal at two points where the two shares of a bit were being tested. We observed a significant difference in the measure for the case where a zero bit was shared as opposed to where a one bit was shared. No such difference was observed with five thousand power samples. We illustrate these results in the next section which deals with a more general case.

A valid criticism of the above experiment is that we had the benefit of knowing exactly where the shares were being manipulated and also the bad instructions being used. In real life, it is highly unlikely that an adversary would have any knowledge of code in the card. Thus the attack will only be useful in practice if it can deal with unknown code.

3.3.1 Attacks on Unknown Code

Suppose we are given a chipcard containing an unknown k-way secret-sharing based DPA protected code for a known algorithm. Further assume that "bad" instructions have already been identified and some of these instructions are used to manipulate shares. These, of course, are necessary conditions for EM attacks to be more effective than power attacks. Let us also assume that it is also possible using signal processing to remove execution sequence and variable clock randomization that has been added as countermeasures to complicate alignment of signals and each signal can be realinged into a cannonical execution sequence[14]

The value of k is usually small. For simplicity, assume that k is 2: the attack generalizes for slightly larger k. Fix a reasonable limit L on the number of EM samples that can be collected. We now show that if k is small and if with knowledge of the code we could have broken the protected code using L samples, then this attack can break the unknown protected code with $O(L)$ samples.

In case of a two-way split, a first step is to identify the two locations where the shares of algorithmic quantity are being manipulated using bad instructions in the computation. If code-execution randomization can be removed using signal processing, then this can, in principle be done for many algorithms. Knowing the algorithm, one can provide two different inputs such that the value of the variable is different for these inputs while most of the other variables are the same within the window of interest. For example, in DES one could choose two inputs which differ only on chosen 1 bit, so that only the output of a single S-box is affected. This way we can try to discover where the shares of the output of the S-box are manipulated in the cannonical execution sequence.

Take L EM samples for each of these two different inputs. If the exact locations were known then there is second order statistic, S, that can be applied to the signal at these two locations to distinguish between the two different inputs, thus enabling hypothesis testing.

Without location information, one can only assume that the two locations are an integral number, D, of clock cycles apart. So the strategy is to compute the statistic S for each point on the signal with respect to a corresponding point D cycles away. This exercise is done for both sets of inputs for all reasonable values of D. If the shares of the variable are not manipulated at distance D, then the

[14]In our experience this has been quite feasible, especially since one does not need perfect canonical realignment, only an alignment which is correct with a reasonable probability.

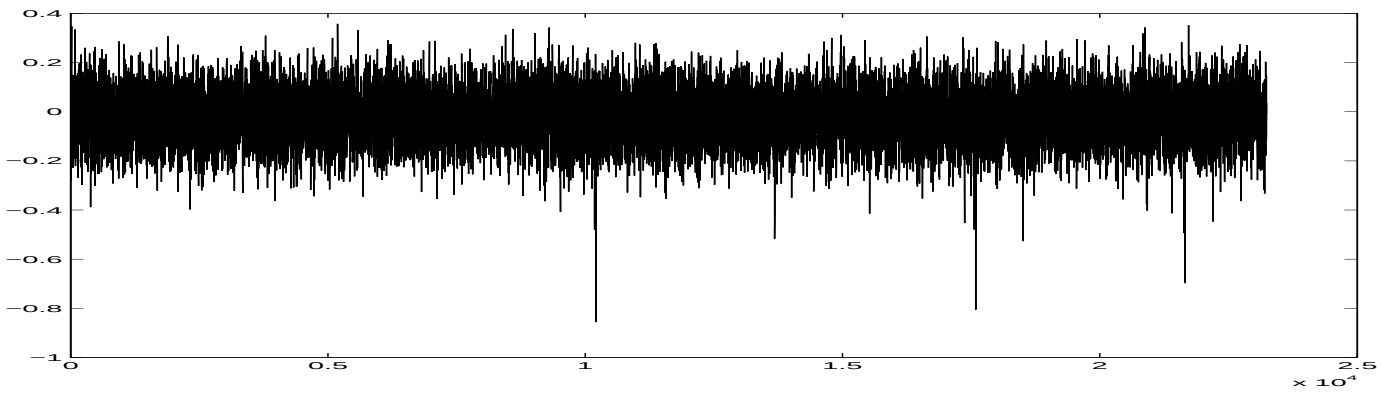

Figure 24: Difference in correlation statistics for $D = 40$, $L = 500$

values of the statistic S at all points will be similar for the two inputs. However, for the right value of D, there will be a significant difference in S exactly at the point where the first share is manipulated and thus the exact location of each share is revealed.

A practical optimization is to choose the two inputs so that multiple variables are different. Then the above exercise will yield candidate locations for the shares for these variables. Once share locations are identified, second (or higher) order attacks can be applied as if the code were known.

We illustrate this attack on the test implementation with the bit test instruction mentioned earlier. In the implementation, the shares of one of the input bits were tested 40 cycles apart. Section 3.1 shows that when a bit is 1, the signal at the bit test instruction is high and when the bit is 0, the signal is low. For a 2-way bit split using an XOR-scheme, the shares of a 0 bit will be (0, 0) or (1, 1) with equal probability and the shares of a 1 bit would be (0, 1) or (1, 0) with equal probability. This suggests that a good statistic S is the correlation coefficient between the corresponding signal points where the shares of bits are being tested. S will be positive when the bit is 0 and negative when the bit is 1.

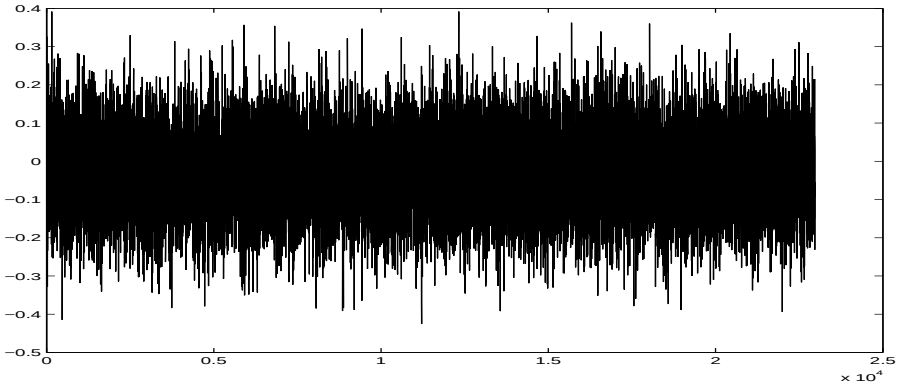

Figure 25: Difference in correlation statistics for $D = 50$, $L = 500$

We experimented with $L = 500$, for two different inputs, which differed in exactly three bits. Figure 24 shows the difference in the statistic S when the distance D is 40, plotted against elasped time in 10ns units. The three significant negative peaks were confirmed to be at exactly the points

where the first shares of the three bits (that differ) were being manipulated. No peaks were found when D differed from 40, e.g, Figure 25 shows the case for $D = 50$. No peaks were seen even for $D = 40$ for five thousand power signals, showing that higher order DPA does not work with five thousand signals. Our findings also show that this EM attack even works with $L = 200$.

4 Adversarial Model, Attacks and Leakage Assessment

This section develops and uses an adversarial model to formally address issues relating to leakage of information via multiple side-channel signals using several sensors. In particular, we address questions such as: Given a limit on signal collection capabilities, which signals should an adversary choose? 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? For reasons explained in the Introduction, this model does not deal with specific algorithms and implementations. Instead, it focuses solely on the *elementary leakages* of information about relevant states in each cycle of each *elementary operation* of CMOS devices via sensors to different adversaries including unbounded ones.

4.1 Adversarial Model

In CMOS devices, compromising side-channel signals during each clock cycle depend *solely*[15] on the *relevant state* at the beginning of the cycle and some random thermal noise (see Section 2). It is therefore natural to formulate questions about the information leakage in terms of 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 adversary[16]. Such a formulation also makes sense as binary hypothesis testing has traditionally been central to the notions of side-channel attack resistance and leakage immunity defined and used earlier [1, 3].

The adversarial model consists of two phases. The first phase, known as the *the emanation 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 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 the distribution used. The utility of the side-channel can then be measured in terms of the success probability the adversary can achieve as a function of the number of invocations allowed.

[15] In practice, this is a very good first approximation. Minor deviations due to residual second order effects can also be addressed as they are transient.

[16] 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.

4.2 Sophisticated Adversarial Strategies

Hypothesis testing is a well researched area in fields such as Information Theory, Statistics and Signal Detection Theory. Techniques from these areas provide several sophisticated strategies that even an adversary with limited resources can use. We now describe one such technique which is optimal in theory, but may require some assumptions and approximations to implement in practice.

In the emanation profiling phase, the adversary builds a statistical profile of the signals available from sensors for each hypothesis H_0 and H_1. For each hypothesis, the adversary performs K independent experiments in which the relevant state is drawn according to distributions B_0 and B_1. In each experiment, he collects a vector $\mathbf{O}$ of n signals (i.e., $\mathbf{O} = [O_1, \ldots, O_n]^T$) from the sensors. [17].

As a result, for each hypothesis, he obtains a collection, $\mathbf{O}_i = [O_{i1}, \ldots, O_{in}]^T$, $i = 1, \ldots, K$, of signal vectors. He then computes the *average sensor signal* denoted by $\mathbf{S} = \{S_1, \ldots, S_n\}$, where $S_j = \sum_{i=1}^{K} O_{ij}/K$. Subtracting the average sensor signal $\mathbf{S}$ from the observation vectors $\mathbf{O}_1, \ldots, \mathbf{O}_K$, produces K *noise signal vectors* $\mathbf{N}_1, \ldots, \mathbf{N}_K$. The next step is to derive the statistical characterization of the noise signal under this hypothesis using $\mathbf{N}_1, \ldots, \mathbf{N}_K$ as samples. It is well known that if the value of K is large enough, a *complete* statistical characterization of the noise signals in form of a probability density function $p_{\mathbf{N}}(\cdot)$ can be obtained at least in theory [5]. The results of the emanation profiling phase are the signal characterizations, $\mathbf{S}_0$ and $\mathbf{S}_1$, and the noise probability density characterizations, $p_{\mathbf{N0}}(\cdot)$ and $p_{\mathbf{N1}}(\cdot)$ for the two hypotheses.

In the hypothesis testing phase. the adversary acquires L sets of sensor signals $\mathbf{O}_i, i = 1, \ldots, L$. It is well-known that in order to minimize the probability of error in hypothesis testing, the adversary should conduct the *maximum likelihood test*. The likelihood ratio $\Lambda(\mathbf{O}_1, \ldots, \mathbf{O}_L)$ is given by $\prod_{i=1}^{L} p_{\mathbf{N1}}(\mathbf{O}_i - \mathbf{S}_1)/p_{\mathbf{N0}}(\mathbf{O}_i - \mathbf{S}_0)$. The adversary decides in favor of H_1 if $\Lambda(\mathbf{O}_1, \ldots, \mathbf{O}_L) > 1$, and decide in favor of H_0 otherwise.

While the approach thus far is optimal, it may be impractical as an exact characterization of the noise probability density $p_{\mathbf{N}}(\cdot)$ may be infeasible. Such a characterization has to capture the nature of each of the noise signals and the dependencies between them. This could further be complicated by the fact that, in addition to thermal noise, the noise signals could also display additional structure due to the interplay between properties of the device and those of the distributions. 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 noise signals will display a bimodal effect attributable to the MSB. Noise characterization, however is a well studied problem, and there exist a rich set of techniques which allow one to obtain near optimal results by making the right assumptions about the noise. Such assumptions greatly simplify the task by permitting the use of only partial characterizations of noise.

4.2.1 The Gaussian Assumption

One widely applicable assumption is the *Gaussian assumption* which states that the noise vector $\mathbf{N}$ has a multivariate Gaussian distribution with zero mean and a covariance matrix Σ_N. Such an assumption approximately holds for a large number of devices and hypotheses encountered in practice. Use of such multivariate statistics provides far better results, such as greatly reduced error probabilities for hypothesis testing, as compared to current side channel attack techniques. Such modeling also permits comparison of different multiple signal selection strategies. This analysis often results in counterintuitive multi-signal selection strategies: e.g., sometimes it is better to use signals with low

[17] Practical issues regarding the composition of these signals and the value of n will be described later.

signal-to-noise ratio and low noise correlation than signals with high signal-to-noise ratio but high noise correlation.

Figure 26 illustrates the noise distribution at a single point in the execution of the "bad" instruction shown in Figures 22 and 23. As can be seen the noise distribution is well approximated by a Gaussian distribution. It should be noted that sometimes a better approximation to the noise density is be obtained by a mixture of Gaussian densities. In other cases, analysis based on such an assumption is still useful as it provides valuable insights and often, improvements by more refined models may not be significant. We also mention that there is a significant body of techniques dealing with non-Gaussian noise, whose description is beyond the scope of this paper.

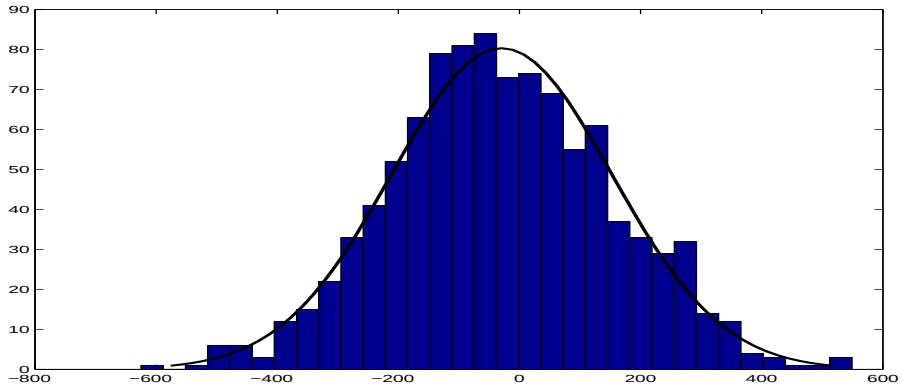

Figure 26: Noise distribution of an EM signal at a single point (point 5 in 3 cycles for a bad instruction) showing an approximately Gaussian distribution

A multivariate Gaussian distribution $p_{\mathbf{N}}(\cdot)$ has the following form:

$$p_{\mathbf{N}}(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_N|}} \exp(-\frac{1}{2}\mathbf{n}^T \Sigma_N^{-1}\mathbf{n}), \quad \mathbf{n} \in \mathcal{R}^n \tag{1}$$

where $|\Sigma_n|$ denotes the determinant of Σ_N and Σ_N^{-1} denotes the inverse of Σ_N. For simplicity, we assume that the noise densities for the two hypotheses, $p_{\mathbf{N0}}(\mathbf{n})$ and $p_{\mathbf{N1}}(\mathbf{n})$ are the same and denoted by $p_{\mathbf{N}}(\mathbf{n})$. With these assumptions, the hypothesis test based on the likelihood ratio for a single observation[18] simplifies to: The adversary decides in favor of H_1 if $(\mathbf{S}_1 - \mathbf{S}_0)^T \Sigma_N^{-1} \mathbf{O} \geqslant \tau$, and in favor of H_0 otherwise, where $\tau = \frac{1}{2}(\mathbf{S}_1^T \Sigma_N^{-1} \mathbf{S}_1 - \mathbf{S}_0^T \Sigma_N^{-1} \mathbf{S}_0)$ (see [14]). Note that under the Gaussian assumption, only the covariance matrix Σ_N needs to be determined for maximum likelihood testing. The following well-known result from the Detection Theory is now applicable [14].

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

$$P_\epsilon = \frac{1}{2}\,\mathrm{erfc}\Big(\frac{\Delta}{2\sqrt{2}}\Big) \tag{2}$$

where $\Delta^2 = (\mathbf{S}_1 - \mathbf{S}_0)^T \Sigma_N^{-1} (\mathbf{S}_1 - \mathbf{S}_0)$ *and* $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$. *Note that* Δ^2 *has a nice interpretation as the optimal signal-to-noise ratio that an adversary can achieve under the Gaussian assumption.*

[18] Generalizations to multiple observations are straightforward.

Example 1 To show the power of maximum-likelihood testing, consider a chipcard with a "bad" instruction, such as the one described in Section 3. EM signals and their statistics for such an instruction are shown in Figures 22 and 23 respectively in Section 3.1.2. It is clear from the Figure 22 that the maximum difference in the mean signals occurs at point 5. A simple approach to binary hypothesis testing would be to compare a given observation to the mean signals at this point and decide in favor of the mean signal which is closer to the observation. In an experiment consisting of 2000 such observations, this approach results in a probability of error of 49.7% which is marginally better than just guessing.

Clearly, the probability of error can be reduced by considering all the points in the observation. A simple approach again is to compute the Euclidean distance between the observation and the mean signals and decide in favor of the closer signal. In an experiment consisting of 2000 such observations, this approach results in a probability of error of 48.8% which is somewhat better than using just one point. However, for the maximum likelihood testing, the probability of error for the same observations turns out to be 44.7%. This is in excellent agreement with the probability of error of 45.5% predicted by (2). This agreement validates our Gaussian assumption for this particular instruction and the chipcard.

It should be kept in mind that we have achieved this bias just by looking at *one* bit in *one* invocation of a "bad" instruction. It can be shown that the error probability for maximum likelihood testing decreases exponentially in the number of invocations. This implies that the error probability can be made arbitrarily small with only a modest increase in the number of invocations, say, 20–30.

Since erfc$(\cdot)$ decreases exponentially with Δ, the goal of a resource-limited adversary would be to choose signals in such a manner, as to maximize the output signal-to-noise ratio Δ^2. As shown below, this goal differs from the naive approach of choosing the signals with best signal-to-noise ratios *and then* feeding them into a signal processing unit to test a hypothesis.

Example 2 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 $\mathbf{N} = (N_1, N_2)^T$ 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

$$\Delta^2 = \frac{(S_1 + S_2)^2}{2(1+\rho)} + \frac{(S_1 - S_2)^2}{2(1-\rho)} \tag{3}$$

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:

$$\Delta^2 = \frac{(S_1 + S_2)^2}{2} + \frac{(S_1 - S_2)^2}{2} = S_1^2(1 + S_2^2/S_1^2) \tag{4}$$

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. ∎

Results obtained by using the Gaussian assumption are useful in analyzing several situations. For instance, suppose we have only analyzed the signal-to-noise ratios of two channels and not the noise correlations between them. Suppose an adversary collects a signal from each of these channels by invoking an operation. We can still compute a lower bound on his error probability by assuming that $\rho = 0$, since his error probability will only increase as the correlation coefficient ρ increases. In general, even for non-Gaussian noise distributions, a lower bound on the error probability can be obtained by assuming that multiple sets of sensor signals are statistically independent.

4.3 Leakage Assessment Methodology for Chipcards

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, $\Lambda(\mathbf{O})$ is a function of the random vector $\mathbf{O}$, the best achievable success probability, P_s, is given by:

$$P_s = \sum_{\mathbf{O} \in \Omega} I_{\{\Lambda(\mathbf{O})>1\}} p_{\mathbf{N1}}(\mathbf{O} - \mathbf{S}_1) + I_{\{\Lambda(\mathbf{O})<1\}} p_{\mathbf{N0}}(\mathbf{O} - \mathbf{S}_0) \tag{5}$$

where I_A denotes the indicator function of the set A.

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:

$$\mu(s) = \ln\Big(\sum_{\mathbf{O} \in \Omega} p^s_{\mathbf{N1}}(\mathbf{O} - \mathbf{S}_1) p^{1-s}_{\mathbf{N0}}(\mathbf{O} - \mathbf{S}_0) \Big) \tag{6}$$

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

Fact 2 *Assume we have several statistically independent observation vectors*[19] $\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_L$. *For this case, the best possible exponent in the probability of error is given by the* Chernoff Information*:*

$$C \stackrel{\text{def}}{=} - \min_{0 \leqslant s \leqslant 1} \mu(s) \stackrel{\text{def}}{=} - \mu(s_m) \tag{7}$$

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:

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

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

Note that in order to evaluate (5) or (8), 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.

4.4 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 (see Theorem 1) 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 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, is that for a fixed relevant state, the noise distribution $p_{\mathbf{NR}}(\cdot)$ can be approximated by a Gaussian distribution. In Section 4.2.1, we provide experimental evidence to validate this assumption for CMOS devices. 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 (5) and (7) 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.

4.5 Completeness Theorem for Elementary Leakages

Finally, we conclude this paper with a theorem to justify the completeness of our assessment process for gauging side-channel vulnerabilities not just at a cycle level but vulnerabilities of entire implementations.

Our model and vulnerability assessment deals with the leakage of information about the relevant state of elementary operations at the cycle level. Thus it can directly be used to assess cycle by cycle leakage of relevant state information from the side-channels for an implementation, where the cycles are treated independently. This is the case for the classical differential side-channel attacks also known as first-order attacks. However, any implementation which consists of multiple operations will have very strong dependencies between the relevant states that occur across multiple cycles spread throughout the computation. These dependencies are the basis of higher-order differential side-channel attacks like the ones we describe in the paper. Thus, as specified in [1], assessing vulnerabilities for such implementations can be modeled as a hypothesis testing exercise to distinguish between two possible *joint* distributions of the relevant states in the computation, using the *joint* signal distributions introduced in the side-channel.

On the surface, it may seem that the proposed assessment of vulnerabilities, which takes into account only the cycle by cycle information available from sensor signals about the relevant state, may not be comprehensive. In this section, we prove a completeness theorem (Theorem 1) that shows that this is not the case, i.e., collecting cycle by cycle information about the relevant states is sufficient for the assessment of vulnerabilities of entire implementations.

Theorem 1 (Completeness of Elementary Leakages) *Let H be a hypothesis on the relevant state distribution of an entire implementation. Let $\mathcal{O}$ be any observation vector for the implementation and $T(\mathcal{O})$ be the cycle-by-cycle sufficient statistics obtained using $\mathcal{O}$ and our assessment. Then,*

$$I(H;\mathcal{O}) = I(H;T(\mathcal{O}))$$

That is, all information for hypothesis testing of implementations is in the elementary leakages.

Our proof makes heavy use of information theoretic arguments, and we refer the reader to Cover and Thomas [4] for the background material used here from Information Theory.

Let $\mathcal{R} = [\mathbf{R}_1, \ldots, \mathbf{R}_K]^T$ denote relevant states during an attack that lasts K cycles. Assume that $\mathcal{R}$ is distributed according to a density in the family of probability density functions $\{p_h(\cdot)\}$ indexed by $h \in \Omega$. The set Ω consists of different hypotheses under test. Let $\mathcal{O} = [\mathbf{O}_1, \ldots, \mathbf{O}_K]^T$ be the vector of observations during the attack. As discussed earlier, due to the CMOS circuitry, observations in each clock cycle depend only on the relevant state in that clock cycle. Mathematically, this can be expressed by saying that the conditional distribution of $\mathbf{O}_i$ depends only on $\mathbf{R}_i$ and is conditionally independent of the hypothesis h, that is,

$$\Pr(\mathbf{O}_i|\mathcal{R}) = \Pr(\mathbf{O}_i|\mathbf{R}_i), \quad \text{and} \tag{9}$$

$$\Pr(\mathbf{O}_i, h|\mathbf{R}_i) = \Pr(\mathbf{O}_i|\mathbf{R}_i)\Pr(h|\mathbf{R}_i). \tag{10}$$

Using (9) and (10), it is easy to check that given $\mathcal{R}$, the observations $\mathcal{O}$ and the hypothesis h are statistically independent:

$$\Pr(\mathcal{O}, h|\mathcal{R}) = \Pr(\mathcal{O}|\mathcal{R})\Pr(h|\mathcal{R}) \tag{11}$$

Let $T(\mathbf{O}_i)$ be a *sufficient statistic* for $\mathbf{R}_i$ derived from $\mathbf{O}_i$, that is,

$$I(\mathbf{R}_i;\mathbf{O}_i) = I(\mathbf{R}_i;T(\mathbf{O}_i)) \qquad \text{for all distributions on } \mathbf{R}_i. \tag{12}$$

Here $I(X;Y)$ denotes the *mutual information* between X and Y [4]. Let $T(\mathcal{O})$ denote the *cycle-by-cycle* sufficient statistics given by $T(\mathcal{O}) = [T(\mathbf{O}_1),\ldots,T(\mathbf{O}_K)]^T$. The following theorem asserts that as far as the determination of hypothesis h is concerned, the information contained in the cycle-by-cycle sufficient statistics is the same as the information contained in the whole observation vector $\mathcal{O}$.

Theorem 2 *For the setup described above, regardless of the statistical distribution of hypothesis h, the mutual information between h and $\mathcal{O}$ is equal to the mutual information between h and $T(\mathcal{O})$:*

$$I(h;\mathcal{O}) = I(h;T(\mathcal{O})) \tag{13}$$

Proof. Recall that $T(\mathbf{O}_i)$ is a sufficient statistics for $\mathbf{R}_i$ derived from $\mathbf{O}_i$, that is,

$$I(\mathbf{R}_i;\mathbf{O}_i) = I(\mathbf{R}_i;T(\mathbf{O}_i)) \qquad \text{for all distributions on } \mathbf{R}_i. \tag{14}$$

Since the conditional distribution of $\mathbf{O}_i$ only depends on $\mathbf{R}_i$, by using the chain rule of mutual entropies, we can show that $I(\mathcal{R};\mathcal{O}) = \sum_{j=1}^{K} I(\mathbf{R}_i;\mathbf{O}_i)$ and $I(\mathcal{R};T(\mathcal{O})) = \sum_{j=1}^{K} I(\mathbf{R}_i;T(\mathbf{O}_i))$. As a result

$$I(\mathcal{R};\mathcal{O}) = I(\mathcal{R};T(\mathcal{O})) \qquad \text{for all distributions on } \mathcal{R}. \tag{15}$$

By the chain rule of mutual entropies,

$$I(\mathcal{R},h;\mathcal{O}) = I(h;\mathcal{O}) + I(\mathcal{R};\mathcal{O}|h) \tag{16}$$

$$= I(\mathcal{R};\mathcal{O}) + I(h;\mathcal{O}|\mathcal{R}) \tag{17}$$

The conditional independence of $\mathcal{O}$ and h given $\mathcal{R}$ (see (11)), implies that $I(h;\mathcal{O}|\mathcal{R}) = 0$, and,

$$I(h;\mathcal{O}) = I(\mathcal{R};\mathcal{O}) - I(\mathcal{R};\mathcal{O}|h). \tag{18}$$

Since $T(\mathcal{O})$ is a function of $\mathcal{O}$, the conditional independence of $\mathcal{O}$ and h given $\mathcal{R}$ also implies the conditional independence of $T(\mathcal{O})$ and H given $\mathcal{R}$. Using this fact and the chain rule of mutual entropies we can deduce that

$$I(h;T(\mathcal{O})) = I(\mathcal{R};T(\mathcal{O})) - I(\mathcal{R};T(\mathcal{O})|h). \tag{19}$$

An examination of (13), (15), (18), and (19) shows that it suffices to prove

$$I(\mathcal{R};\mathcal{O}|h) = I(\mathcal{R};T(\mathcal{O})|h) \tag{20}$$

Intuitively, the above equality holds since the value of h only determines the distribution of $\mathcal{R}$, and (15) holds regardless of the distribution of $\mathcal{R}$. In the rest of this proof, we will formally show that (20) holds.

We start by expanding $I(\mathcal{R},\mathcal{O})$ for the case when $\mathcal{R}$ is distributed according to $p_\omega(\cdot)$ for some $\omega \in \Omega$.

$$\begin{aligned} I(\mathcal{R},\mathcal{O}) &= -\sum_{\mathcal{O}} p(\mathcal{O})\log p(\mathcal{O}) + \sum_{\mathcal{O},\mathcal{R}} p(\mathcal{R},\mathcal{O})\log p(\mathcal{O}|\mathcal{R}) \\ &= -\sum_{\mathcal{O}}\Big(\sum_{\mathcal{R}} p(\mathcal{O}|\mathcal{R})p_\omega(\mathcal{R})\Big)\log\Big(\sum_{\mathcal{R}} p(\mathcal{O}|\mathcal{R})p_\omega(\mathcal{R})\Big) + \sum_{\mathcal{O},\mathcal{R}} p(\mathcal{O}|\mathcal{R})p_\omega(\mathcal{R})\log p(\mathcal{O}|\mathcal{R}) \\ &= H^\omega(\mathcal{O}) - H^\omega(\mathcal{O}|\mathcal{R}) \qquad \text{for all } h \in \Omega \end{aligned} \tag{21}$$

Here $H^\omega(\cdot)$ denotes the entropy when the distribution of $\mathcal{R}$ is given by $p_\omega(\mathcal{R})$. Using a similar argument, we can deduce that

$$I(\mathcal{R}, T(\mathcal{O})) = H^\omega(T(\mathcal{O})) - H^\omega(T(\mathcal{O})|\mathcal{R}) \qquad \text{for all } \omega \in \Omega$$

Since $I(\mathcal{R}, \mathcal{O}) = I(\mathcal{R}, T(\mathcal{O}))$ for all distribution of $\mathcal{R}$, it follows that

$$H^\omega(\mathcal{O}) - H^\omega(\mathcal{O}|\mathcal{R}) = H^\omega(T(\mathcal{O})) - H^\omega(T(\mathcal{O})|\mathcal{R}) \qquad \text{for all } \omega \in \Omega \tag{22}$$

The following calculation shows that in fact, $H^\omega(\mathcal{O}|\mathcal{R}) = H(\mathcal{O}|\mathcal{R}, h = \omega)$ for all $\omega \in \Omega$.

$$\begin{aligned} H(\mathcal{O}|\mathcal{R}, h = \omega) &= -\sum_{\mathcal{O},\mathcal{R}} p(\mathcal{O}, \mathcal{R}|h = \omega) \log p(\mathcal{O}|\mathcal{R}, h = \omega) \\ &= -\sum_{\mathcal{O},\mathcal{R}} p(\mathcal{R}|h = \omega) p(\mathcal{O}|\mathcal{R}) \log p(\mathcal{O}|\mathcal{R}) \\ &= -\sum_{\mathcal{O},\mathcal{R}} p_\omega(\mathcal{R}) p(\mathcal{O}|\mathcal{R}) \log p(\mathcal{O}|\mathcal{R}) \\ &= H^\omega(\mathcal{O}|\mathcal{R}) \end{aligned} \tag{23}$$

A similar calculation shows that $H^\omega(T(\mathcal{O})|\mathcal{R}) = H(T(\mathcal{O})|\mathcal{R}, h = \omega)$, $H^\omega(\mathcal{O}) = H(\mathcal{O}|h = \omega)$, $H^\omega(T(\mathcal{O})) = H(T(\mathcal{O})|h = \omega)$ for all $\omega \in \Omega$. We can now finish the proof as follows

$$\begin{aligned} I(\mathcal{R}; \mathcal{O}|h) &= H(\mathcal{O}|h) - H(\mathcal{O}|\mathcal{R}, h) \\ &= \sum_{\omega \in \Omega} p(h = \omega) \Big(H(\mathcal{O}|h = \omega) - H(\mathcal{O}|\mathcal{R}, h = \omega) \Big) \\ &= \sum_{\omega \in \Omega} p(h = \omega) \Big(H^\omega(\mathcal{O}) - H^\omega(\mathcal{O}|\mathcal{R}) \Big) \\ &= \sum_{\omega \in \Omega} p(h = \omega) \Big(H^\omega(T(\mathcal{O})) - H^\omega(T(\mathcal{O})|\mathcal{R}) \Big) \qquad (24) \\ &= \sum_{\omega \in \Omega} p(h = \omega) \Big(H(T(\mathcal{O})|h = \omega) - H(T(\mathcal{O})|\mathcal{R}, h = \omega) \Big) \\ &= H(T(\mathcal{O})|h) - H(T(\mathcal{O})|\mathcal{R}, h) \\ &= I(\mathcal{R}; T(\mathcal{O})|h) \end{aligned}$$

where (24) follows from (22).

5 Countermeasures

Due to the presence of several unexpected EM leakages, a comprehensive EM vulnerability assessment has to be an integral part of any effort to develop countermeasures against EM attacks on specific implementations. Such countermeasures fall into two broad categories: *signal strength reduction* and *signal information reduction*. Techniques for signal strength reduction include circuit redesign to reduce egregious unintentional emanations and the use of shielding and physically secured zones to

reduce the strength of compromising signals available to an adversary relative to ambient thermal noise. Techniques for signal information reduction rely on the use of randomization and/or frequent key refreshing within the computation [10, 11, 1, 6] so as to substantially reduce the effectiveness of statistical attacks using the available signals.

6 Acknowledgments

We owe a debt of gratitude to Helmut Scherzer for key components that enabled us to explore EM attacks further, in particular, help with the experimental setup, data collection and analysis tools and with smart card programming. This work has benefited greatly from interactions with Emmanuel Yashchin for his suggestions on the use of various statistics and statistical tools. Last but not the least, we would like to thank Elaine and Charles Palmer for several useful comments and their constant encouragement.

References

[1] Suresh Chari, Charanjit S. Jutla, Josyula R. Rao and Pankaj Rohatgi. Towards Sound Countermeasures to Counteract Power–Analysis Attacks. Advances in Cryptology — Proceedings of Crypto '99, Springer–Verlag, LNCS 1666, August 1999, pages 398–412.

[2] S. Chari, J. R. Rao and P. Rohatgi. Template Attacks, To appear in Proceedings of CHES '02.

[3] Jean–Sebastien Coron, Paul Kocher and David Naccache. Statistics and Secret Leakage. In the Proceedings of Financial Cryptography '00. Lecture Notes In Computer Science, Springer–Verlag, 2000.

[4] T.M. Cover and J. A. Thomas. Elements of Information Theory. John WIley & Sons Inc, 1991.

[5] H. Cramer. Mathematical Models of Statistics. Princeton University Press. 1946.

[6] L. Goubin and J. Patarin. DES and Differential Power Analysis. Proceedings of the Workshop on Cryptographic Hardware and Embedded Systems, CHES '99, LNCS 1717, August 12–13, 1999, Worcester, MA, pages 158–172.

[7] NSA Tempest Series `http://cryptome.org/#NSA--TS`.

[8] Dynamic Sciences International Inc. See `http://www.dynamic--sciences.com/r1550.html`.

[9] K. Gandolfi, C. Mourtel and F. Olivier. Electromagnetic Attacks: Concrete Results. In the Proceedings of the Workshop on Cryptographic Hardware and Embedded Systems 2001 (CHES 2001), LNCS 2162 Paris, France, May 2001, pp 251–261.

[10] P. Kocher. Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS and Other Systems. Advances in Cryptology-Crypto '96, Lecture Notes in Computer Science # 1109, pp 104–113.

[11] P. Kocher, J. Jaffe and B. Jun. Differential Power Analysis: Leaking Secrets. Advances in Cryptology — Proceedings of Crypto '99, Springer Verlag, LNCS 1666, pages 388–397. One version of the paper is available online at `http://www.cryptography.com/dpa/technical/index.html`.

[12] 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), Cannes, France, LNCS 2140, pp.200-210,September 2001.

[13] The complete unofficial TEMPEST web page. Available at http://www.eskimo.com/~joelm/tempest.html.

[14] H. L. Van Trees. Detection, Estimation, and Modulation Theory, Part I. John Wiley & Sons. New York. 1968.