Power Analysis Part II.b

Cryptographic Hardware for Embedded Systems

ECE 3894

Fall 2019

Assoc. Prof. Vincent John Mooney III

Georgia Institute of Technology

©Georgia Institute of Technology, 2018-2019
Reading


• All figures in this lecture are from the aforementioned manuscript.

• NOTE: this lecture, Power Analysis Part II.b, is an updated superset of Power Analysis Part II.a
Questions Answered by This Lecture

• If one measures the energy consumption or power of a microchip, what do the power traces reveal?
• What is the statistical methodology used to reveal the information claimed to have been learned?
Power Analysis

• Cryptographic operations executed on microchips exhibit variations in energy consumption / power
  • Note that power is a rate of energy consumption, i.e., joules per second

• The energy consumption of microchip implementations of cryptography depends on the data values ($P_{data}$) and the specific operations (typically mathematical or memory storage) performed ($P_{op}$)

• There is also an electrical noise component ($P_{el.\,noise}$) and constant ($P_{const}$)

• $P_{total} = P_{op} + P_{data} + P_{el.\,noise} + P_{const}$

• This model may be refined for certain situations, but suffices for most

• Note that cryptographic information may be revealed by $P_{op}$ and $P_{data}$
Figure 3.9. Picture of the measurement setup for the attacks on the 8-bit microcontroller.
Example Power Measurement Setup

- The previous slide shows the microcontroller power measurement setup of Mangard et al.
- Microcontroller power supply is 5V and clock frequency 11 MHz
- Voltage drop across a 1 Ω resistor connected to the power supply (Gnd) is measured by an oscilloscope
- Oscilloscope can sample 8 bits of resolution every nanosecond (GHz)
- Typically measure every four ns in experiments (250 Million Samples per second or MS/s)
Consider a Single Point in a Power Trace

• A moment in time
• Aim to determine the probability distributions of $P_{op}$, $P_{data}$ and $P_{el. noise}$
Figure 4.1. Power traces look very similar if the same data is processed. 

Figure 4.2. Histogram of the power consumption at 362 ns of Figure 4.1.
Figures 4.1 and 4.2 from Mangard et al.

• Fig. 4.1 shows five power traces with the same data and instruction (a load of a zero from on-chip memory to a register)

• The differences in the power traces are due to noise

• Fig. 4.2 shows a histogram of 10,000 power traces of the same operation considering the Voltage across the resister (recall P=IV and I is typically a constant)
  • Most of the measurements are near 112 mV
  • Very few measurements are below 109 mV or above 115 mV
Gaussian (a.k.a. Normal) Distribution

• $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

• $\mu = E(X)$

• $\sigma^2 = Var(X) = E((X - E(X))^2)$

• $X \sim N(\mu, \sigma)$ means that $X$ is normally distributed (has a Gaussian distribution) with mean value $\mu$ and standard deviation $\sigma$

• The “standard normal distribution” has $\mu = 0$ and $\sigma = 1$
Back to Our Experiment at 362 ns

• We can estimate $\mu = E(X)$ with the average $\bar{x}$ calculated empirically:
  • $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

• We can estimate $\sigma = \sqrt{Var(X)}$ with the square root of the variance also calculated empirically:
  • $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

• For this experiment, the following is calculated from the 10,000 traces
  • $\mu = \bar{x} = 111.86 \text{ mV}$
  • $\sigma = s = 1.63 \text{ mV}$
  • Thus, $X \sim N(111.86, 1.63)$ where the units are millivolts
Figure 4.3. The normal distribution $\mathcal{N}(111.86, 1.63)$ models the power consumption at 362 ns.
Power Analysis So Far

- From our current experiment consisting of 10,000 executions of the same memory operation with a data value of zero, we find
  - $E(P_{\text{const}}) = 111.86 \text{ mV}$
  - $E(P_{\text{op}}) = E(P_{\text{data}}) = E(P_{\text{el. noise}}) = 0$
  - $\text{Var}(P_{\text{const}}) = 0$
  - Since the operation executed and the data used do not change, $\text{Var}(P_{\text{op}}) = \text{Var}(P_{\text{data}}) = 0$
  - $\Rightarrow P_{\text{el. noise}} \sim N(0,1.63)$
Next: Data Dependence

• Instead of moving a constant zero from memory to a register location as was done for estimating the distribution of noise on the power pin of this microcontroller, we now vary the eight bit memory data value among all 256 possible values

• 200 measurements for each value = 256* 200 = 51,200 total measurements

• Again, let us look at what happens at 362 ns

• (Quick note before we continue: this discussion will assume large amounts of encrypted data values which are evenly distributed among all possibilities, e.g., zero does not occur more often than any other number; another way to say this is that data values are uniformly distributed)
Figure 4.4. Histogram of the power consumption at 362 ns if different data values are transferred from the internal memory to a register.
Histogram Comments on Figure 4.4

• A precise description (close to exact) is not possible with a single Gaussian

• However, upon further inspection, there do appear to be nine Gaussians

• Why could this be?
  • Answer: consider the Hamming Weight (HW) between each possible data value and zero
HW Calculations on 8 bits Values

- HW 0: 00000000 (TOTAL: 1)
- HW 1: 00000001, 00000010, ..., 10000000 (TOTAL: 8)
- HW 2: 00000011, 00000101, ..., 10000001, 100000010, ...
- ...
- HW 7: 11111110, 11111101, ..., 01111111 (Total: 8)
- HW 8: 11111111 (TOTAL 1)

- Result is a binomial distribution
  - Probability of HW 4 is 27.3%, HW 3 or 5 have an equal probability of 21.9%, ...
  - HW 1 or 7 have a probability of 0.031%, HW 0 or 8 have probability 0.004%
Explanation for Figure 4.4

• Function of HW
  • The greater the HW distance from 0b1111111 the greater the energy consumption $P_{data}$
  • Add in $P_{el.\,noise}$

• Load and store instructions in any ISA are known to have a large HW energy consumption component due to the way SRAMs are designed
  • Rows and columns have to be charged and discharged in an array of 6T single bit memory elements
  • Sense amplifiers are used to detect when a 1 is overpowered by a 0 and vice-versa
  • Furthermore, for this microcontroller, bus lines are precharged to 1 each time, so case is 0b11111111 (HW 8 measured from zero) has the least power
Question: How to Remove the Noise?

• Calculate the mean voltage for each HW
• In this example, find 111.9, 117.6, 123.2, 128.7, 134.0, 139.5, 145.1, 151.2 and 159.6 (all in mV)
• Earlier we found that $P_{el.\ noise} \sim N(0, 1.63)$, i.e., we know that $\sigma = s = 1.63 \text{ mV}$
Question: How to Remove the Noise?

• Calculate the mean voltage for each HW

• In this example, find 111.9, 117.6, 123.2, 128.7, 134.0, 139.5, 145.1, 151.2 and 159.6 (all in mV)

• Earlier we found that $P_{el. \text{ noise}} \sim N(0, 1.63)$, i.e., we know that $\sigma = s = 1.63 \text{ mV}$

• Answer: we cannot remove the noise
  • This is due to the fact that the noise has $\mu = 0$, i.e., $E(P_{el. \text{ noise}}) = 0$

• However, we can account for the noise statistically!
  • $\sigma = 1.63 \text{ mV}$
Result

• Can superpose nine Gaussian distributions to accurately model the measurements
Figure 4.5. The distribution of the power consumption when the microcontroller transfers different data from the internal memory to a register.
Nine Gaussian Distributions

• Mangard et al. propose the following on page 68
  • \( E(P_{\text{const}}) = 134 \, mV \) for each of the nine distributions
  • \( E(P_{\text{data}}) = 0 \) overall (for the combination of the nine distributions)
    • Taken individually, -22.67, -16.92, -11.35, -5.86, -0.49, 4.96, 10.53, 16.68 and 25.12 mV
    • Each HW has a Gaussian weighted by the binomial distribution of \( P_{\text{data}} \)
  • Figure 4.4 shows the result where the sum under the curve results in a total of 1 (i.e., the sum of the probabilities sums to 1)
Next: Energy Consumption per Operation

• Similar to previous, but now alter the ISA operation type
  • Some operations, e.g., load and store or add and subtract, can be grouped together
  • Some operations are very specific, e.g., floating point multiply
  • Also may have to account for multicycle operations
  • As stated in the book, the result is that $P_{op}$ can also be approximated reasonably accurately for this microcontroller (and many other instruction-set architectures or ISAs) by a Gaussian distribution
Signal to Noise Ratio (SNR)

- Two questions: what information is the attacker seeking and what do the points of a power trace provide towards revealing this information?

- We begin with a distinction between exploitable power measurement data $P_{\text{exp}}$ which must be due to either $P_{\text{data}}$ or $P_{\text{op}}$
  - The aspects of $P_{\text{data}}$ and/or $P_{\text{op}}$ which may not be exploitable, e.g., due to various bits switching back and forth not under observation are called $P_{\text{switching}}$ or $P_{\text{sw. noise}}$

- Therefore, $P_{\text{exp}} + P_{\text{switching}} = P_{\text{data}} + P_{\text{op}}$

- Furthermore, $P_{\text{total}} = P_{\text{exp}} + P_{\text{switching}} + P_{\text{el. noise}} + P_{\text{const}}$
SNR Continued

• $SNR = \frac{\text{Var}(\text{Signal})}{\text{Var}(\text{Noise})}$

• $SNR = \frac{\text{Var}(P_{\text{exp}})}{\text{Var}(P_{\text{switching}} + P_{\text{el. noise}})}$
Example

• A processor operates on an 8-bit value where each bit is independent and uniformly distributed.

• Assume that the value of the second bit is always the complement of the first bit in the experiments carried out.
  • E.g., 0bX₇X₆X₅X₄X₃X₂X₁₀ and 0bY₇Y₆Y₅Y₄Y₃Y₂Y₁₁ where the first bit considered in our analysis is the case of the LSB = 0 and the second bit considered in our analysis is the case of the LSB = 1.
  • The other 14 bits are independent and uniformly distributed.

• \( P_{exp} \) consists of the energy consumed by the LSB.

• \( P_{switching} \) consists of the energy consumed by the rest of the bits.
Example (continued)

• We have 51,200 power traces as computed already earlier
• Select the 25,600 traces with LSB = 1
• Figure 4.6 shows the resulting histogram at 362 ns
Figure 4.6. Histogram of the total noise ($P_{sw. noise} + P_{el. noise}$) if the exploitable signal is the LSB of the byte that the microcontroller processes. This noise is approximately normally distributed.

NOTE: Recall that we have defined $P_{switching}$ to be identical to (another name for) $P_{sw. noise}$.
Comments

• $P_{\text{switching}}$ in Fig. 4.6 has a binomial distribution

• Fig. 4.6 also includes $P_{\text{el. noise}}$

• However, we can approximate Fig. 4.6 with a single Gaussian as shown, in particular since we assume data is uniformly distributed

• The result is $\sigma = s = 7.54 \text{ mV}$; thus, for the LSB (1-bit) scenario, we find that $\sigma$ of $P_{\text{switching}} + P_{\text{el. noise}} = 7.54 \text{ mV}$

• Hence, $\text{Var}(P_{\text{switching}} + P_{\text{el. noise}}) = (7.54 \text{ mV})^2 = 56.85 \text{ mV}^2$

• Also, earlier we found that $\sigma$ of $P_{\text{el. noise}}$ for one bit $= 1.63 \text{ mV}$

• Hence, $\text{Var}(P_{\text{el. noise}}) = (1.63 \text{ mV})^2 = 2.67 \text{ mV}^2$
Table 4.2. Variance of the components of the power consumption according to the models discussed in (4.1) and (4.8).

<table>
<thead>
<tr>
<th>Component</th>
<th>Variance 8-bit scenario</th>
<th>1-bit scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{data}$</td>
<td>61.12</td>
<td>61.12</td>
</tr>
<tr>
<td>$P_{op}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$P_{el.,noise}$</td>
<td>2.67</td>
<td>2.67</td>
</tr>
<tr>
<td>$P_{exp}$</td>
<td>61.12</td>
<td>6.87</td>
</tr>
<tr>
<td>$P_{sw.,noise} + P_{el.,noise}$</td>
<td>2.67</td>
<td>56.85</td>
</tr>
</tbody>
</table>
Comparison

• SNR is much higher for 8-bits than for 1-bit
Figure 4.7. The signal levels, the standard deviation of the noise, and the SNR when attacking a uniformly distributed 8-bit data value on our microcontroller.
Mean Traces for the 9 Different Hamming Weights

Voltage [mV]

Time [ns]
Standard Deviation of Noise

Time [ns]

Voltage [mV]

0 100 200 300 400 500

0 2 4 6 8 10
Figure 4.7. The signal levels, the standard deviation of the noise, and the SNR when attacking
Figure 4.7. The signal levels, the standard deviation of the noise, and the SNR when attacking a uniformly distributed 8-bit data value on our microcontroller.
Comments
Figure 4.8: The signal levels, the standard deviation of the noise, and the SNR when attacking a uniformly distributed 1-bit data value on our microcontroller.
Mean Traces for LSB=0 and LSB=1
Signal-to-Noise Ratio

Time [ns]

SNR
Figure 4.8. The signal levels, the standard deviation of the noise, and the SNR when attacking a uniformly distributed 1-bit data value on our microcontroller.
Comments
Correlation and Covariance

• Two points are correlated if they vary together in a related way
• Statistical measure: covariance
  \[ \text{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X) - E(Y) \]
• Theoretical and empirical formulas:
  \[ \rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\times\text{Var}(Y)}} \]
  \[ r = \frac{\sum_{i=1}^{n}(x_i-\bar{x}_i)(y_i-\bar{y}_i)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_i)^2\times\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}} \]
• As defined, the correlation coefficient \( \rho \) varies between -1 and 1, i.e., -1 \( \leq \rho \leq 1 \) and also thus -1 \( \leq r \leq 1 \)
**Figure 4.9.** Scatter Plot: The power consumption at 362 ns is correlated to the power consumption at 363 ns. \( r = 0.82 \)

**Figure 4.10.** Scatter Plot: The power consumption at 362 ns is largely uncorrelated to the power consumption at 400 ns. \( r = 0.12 \)