Machine Learning and Side-channel Attacks a Complicated Relationship

Mateo Estrada
MXE210022
CS 4341.502

December 1st, 2024

Abstract

A side-channel attack is any attack that is based upon extra information gained from a fundamental way a computer protocol or algorithm is implemented. Side-channel attacks have been around for more than 20 years however both the prevalence of side-channel attacks along with the methods used in side-channel attacks have evolved in recent years. This paper aims to detail not just the history of side-channel attacks in their prevalence in attacking cryptographic devices, but also the increasing role that machine learning has had in side-channel attack analysis as well as the negatives that come with this.

1 Introduction

Cryptographic algorithms are mathematical functions used to protect sensitive data and authenticate users, ensuring secure communication and data protection across various businesses and fields [1]. There are two main types of cryptography: symmetric cryptography, which uses a shared key for all communication, and asymmetric cryptography, where each user has a unique key pair. Breaking a cryptographic algorithm involves obtaining the secret key without prior knowledge of it. A cryptographic algorithm is considered secure in practice if it is computationally infeasible to break it within a reasonable time frame using available computing power.

As computational power increased, running cryptographic algorithms on personal computers became more feasible. However, this necessitated the local storage of keys on personal computers, which were more vulnerable to viruses and attacks due to their frequent use and constant internet connectivity. To address this issue, cryptographic devices were developed. These electronic devices implement cryptographic algorithms and securely store cryptographic keys.

The introduction of cryptographic devices added a new layer of security, but also gave rise to attacks specifically targeting these devices to extract the secret keys. One such attack is the side-channel attack, a passive, non-invasive attack that exploits unintended information leakage from the physical implementation of a cryptographic device to determine the secret key by measuring some underlying property of the device.

Side-channel attacks have been known to be tricky to fully understand conceptually, but a good way to understand it is by having roomates. Well actually you don't have to physically have

roomates but if you consider yourself wanting to know some information such as *I wonder when my roomate is cooking* without entering the kitchen then there is a ton of clues that can be gathered by noticing the sounds of the pots, the smells of ingredients, feeling the heat and even the light that comes from the kitchen. All of these clues can help you decipher when your roomate is cooking quite easily. This is fundamentally the same concept that we are dealing with in side-channel attacks. And how sometimes just from the clues surrounding the kitchen not only can you know **when** someone is cooking, with enough familiarity you can even begin to decipher **what** is being cooked, this the power that side-channel attacks can have especially when we mix them with machine learning.

2 Background

The underlying property that is targeted during a side-channel attacks can vary, there exists side-channel attacks that focus on the fundamental hardware properties of cryptographic devices, but as devices have become more complex, there now exists a lot more data that can be analyzed and used to conduct side-channel attacks such as: Timing attacks, which analyze the timing by which data moves in and out of the CPU or memory in a cryptosystem. Electromagnetism attacks which use the electromagnetic leaks that occur within a cryptosystem as a resource to garner sensitive data, to sound side-channel attacks which exploit the sound that is produced during a computation. But one of the most popular is power analysis attacks, because of their powerfulness, they are heavily reliable because the measurements can showcase a clear picture to the attacker of what data is being used and what computation is being done.

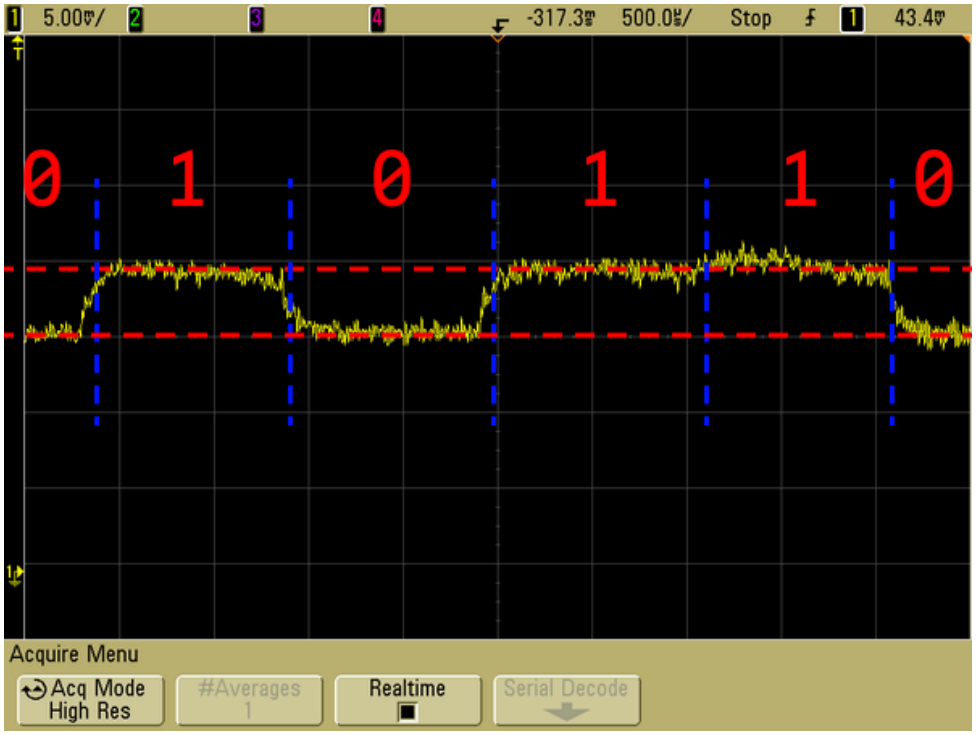

Figure 1: Observing RSA key bits using power analysis [2]

Another attractive feature of power analysis attacks is the fact that they are relatively easy to do compared to other side-channel attacks because they don't require much sophisticated technology, they can generally be done just with a 1 Ω resistor, these reasons has lead to them being one of the most popular studied side-channel attacks. Power analysis side-channel

attacks mainly operate by exploiting the fact that the instantaneous power consumption of a cryptographic device depends on the data it processes and on the operation it performs.

Power analysis side-channel attacks started with **Simple Power Analysis** which was when you would visually analyze graphs of electrical activity of just one power trace. However this would then evolve into **Differential Power Analysis** which is a more advanced form of power analysis which analyzes difference plots to determine the intermediate values, this is possible because the power consumption depends on these intermediate values that are processed during the execution of a cryptographic process. There also exists a third type of power analysis attack which is based on a template attack, template attacks are generally a type of profiling attack in which you attempt to create a template of the architecture used and use that as a starting point to obtain a cryptographic key. In power analysis since you know generally what encryption process is being taken along with what cryptographic device is being used, this means that if before you interact with the power trace of the unknown key, you could have a build up of knowledge that comes from many different plaintexts and cypher key inputs in your template that you use in your analysis to obtain the unknown key.

When it comes to Machine Learning there will a lot of different concepts that will be covered relating to machine learning so this section is just to give a basic overview of the concepts needed to understand the application of machine learning to side-channel Analysis. One advanced implementation of machine learning is a relatively new branch called **Deep Learning** which uses multiple layers to break down the problem by expressing the problem in simpler representations these simpler representations are built upon one another until the entire complex problem is fully tackled.

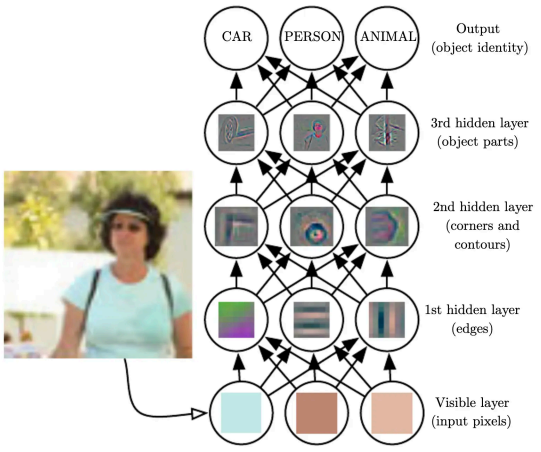

Figure 2: Visualization of a hierarchical neural network that processes image pixels through successive layers (edges, corners/ contours, object parts) to achieve object classification at the output layer. [3]

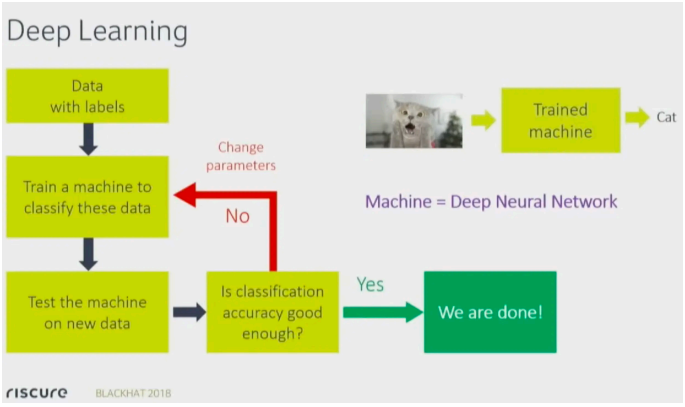

Figure 3: Flowchart illustrating the deep learning training process, from data ingestion through model training and accuracy testing, with an example of cat image classification using a deep neural network. [4]

There are many different architectures of deep learning models, and each one has its different pros and cons and attributes, which is why we will stop to explain each one we encounter as well as any other machine learning concepts needed to understand when we get to it.

3 ML-Based Attack Methodologies

To showcase the massive advantages of deep learning within side-channel attack analysis let us compare the standard way that one would do side-channel analysis for a power analysis attack.

The general flow that an individual or team would take would look something like what Figure 4 details this process is not only a very detail oriented process, with a lot of different aspects of what you can and can't tweak. But it also is somewhat an art, it takes a lot of signal processing experience specifically of power analysis signals to be able to understand what changes need to be made to be able to get a preferential output. One of the biggest challenges being misalignment, algorithms expect a fixed input and if the signals that you have are misaligned, – meaning that the general pattern that you are analyzing isn't Constituently occurring usually because of noise – then that algorithms performance drastically drops. These misalignments can be manually corrected, but this is truly a trial and error process where you try various things and sometimes these attempts can be in the completely wrong direction.

This is where machine learning and more specifically deep learning are able to come in and help alleviate some of these pain points. The biggest advantage of deep learning architecture is that it can be trained with imperfect or misaligned data. A clear example of this is with neural networks and specifically **Convolutional Neural Networks (CNN)**, which mainly are used for image classification. When doing this image classification CNN's are able to detect features independent of their positions, so even if our image in Figure 2 was upside down or sideways as long as all of the original content was in the frame we would have obtained the same features and the same output.

Our CNN when applied to analyze power analysis signals, can now do this feature interpretation even with misaligned or noisy signal data. This was successfully attempted by Perin et al. [5] and the results were a drastic improvement over more standard methods. Figure 6 showcases the results of their CNN compared to correlation power analysis – which is a standard algorithm and process to obtain a key using side-channel attack analysis – and template attack which was previously discussed in this paper. The graph details the number of traces on the x axis and the ranking of the keys in the y axis, meaning that the CNN is the only one which was able to successfully obtain the obfuscated key after around 50 thousand traces.

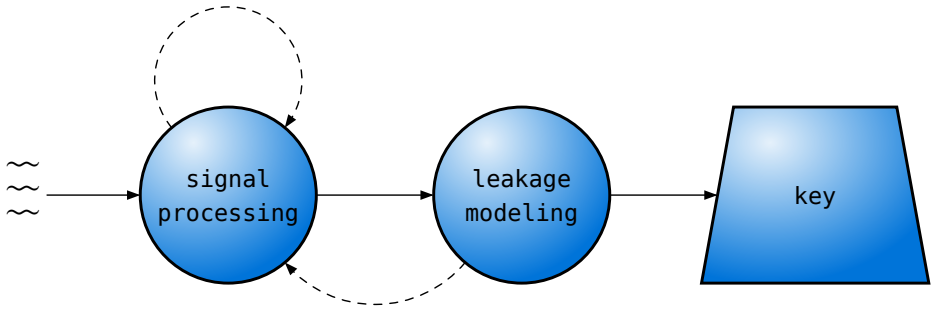

Figure 4: General process when dealing with power analysis side-channel attacks

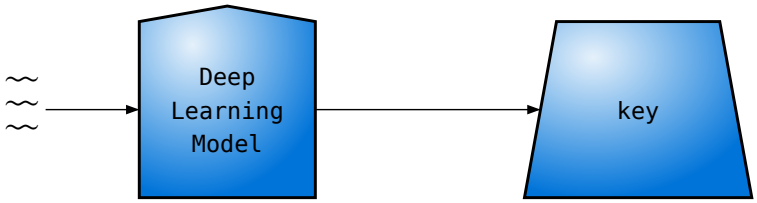

Figure 5: General process when dealing with power analysis side-channel attacks using ML

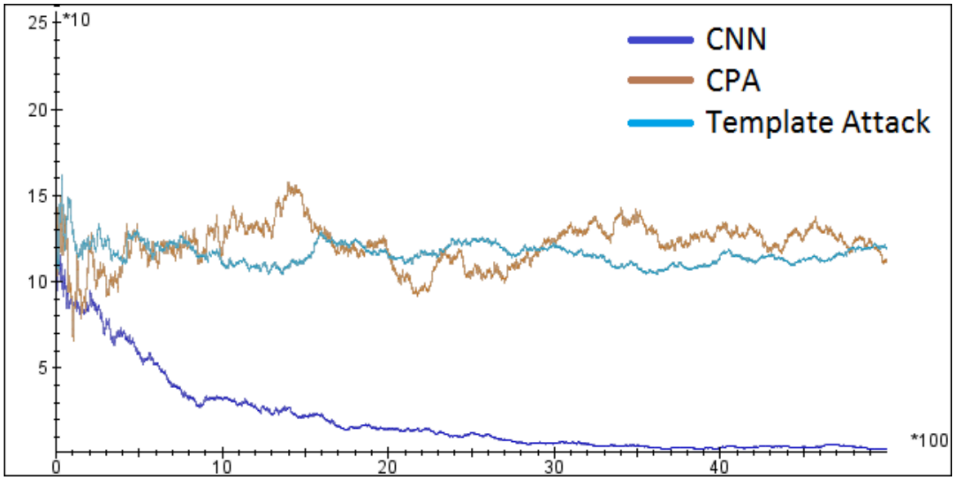

Figure 6: key ranking evolution from Perin et al. [5]
The X axis is the *number of traces* while the Y axis is the *Key Byte Rank*

4 Case Study in Advanced ML Methods

Even though the previous section highlights the power of deep learning to overcome traditional power analysis methods without human intervention or signal pre-processing but this isn't a universal method for all signals. More specifically when it comes to handeling more complex cryptographic computations or purposefully obfuscated encryption processes, deep learning can't always be the best solution. Which is why the research by Lerman et al.[6] is so important. Their research focused on power analysis attacks on both symmetric (3DES) and asymmetric(RSA-512) encryption implementations using machine learning.

The key insight in their paper comes from their novel ML approach which in contrast to our previous section doesn't actually use deep learning. The main process that followed can be separated into three steps

1. Decomposition of the prediction task into separate binary classification problems
2. Dimensionality reduction to handle high-dimensional power traces
3. Model selection find the optimal machine learning technique

The decomposition of the prediction tasks is actually quite advanced so many specifics have been left out, but the main idea is that when you try to decompose a cryptographic key you are trying to predict multiple bits at once, for example one byte (8 bits) you have $256(2^8)$ possible values. Trying to predict all of these at once can lead to a highly complex classification problem – classification problems in ML are one of the two types of ML problems where you want to put a value or target into a specific pre-defined bucket –. So in practice what Lerman et al. did was to decompose the problem into multiple binary classification problems where they created separate models for each bit. This drastically simplified the problem while also

providing interesting results for example some bits where easier to predict compared to others, they found a that MSB (Most Significant Bit) were easier to predict compared to the LSB (Least Significant Bit) which is quite interesting. This is another advantage of this method, it allows for a more granular analysis such as being able to specialize feature selection per bit. Overall this step helped to reduce the complexity which in turn offered higher performance.

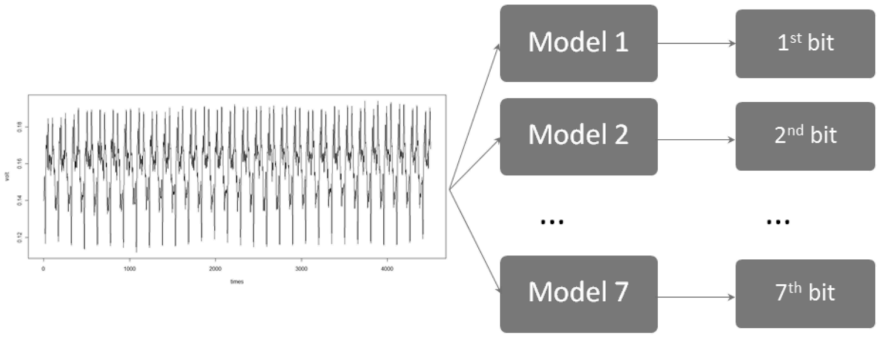

Figure 7: Decomposition of the prediction problem into a set of binary classification tasks from Lerman et al. [6]

Step two of their process was dimensionality reduction, which is needed because of the high dimensionality of signal data and more specifically power traces. Power traces can be represented as multidimensional or multivariate vectors, where each dimension represents the power consumption of a device at a specific time during the execution of a cryptographic algorithm [6]. Since only a subset of these variables can carry useful information about the targeted key then it is necessary to reduce the dimensionality – for example the recorded traces where upwards of 9000 data points per trace with around 50 traces in total – . This again is a preprocessing technique which helps to reduce the complexity of the problem and in turn helps to improve the performance of the model.

The final step was model selection which is where the researchers would test multiple machine learning models, and then would compare their performance to see which one was the best. This is a very common practice in machine learning where you try to find the best model for your specific problem. Some of the models that where tested where:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Self-Organizing maps (SOM)

The researchers found that the best model for power attack analysis was **Random Forest**, which is a type of ensemble learning method that is used for both classification and regression problems and more importantly not a type of deep learning architecture. The results of RF where by far the best with an accuracy of 15.33% for 3DES and 2.79% for RSA-512 with a reduced number of only 50 traces – for context the previous deep learning method used upwards of 50 thousand traces –. Which is an improvement from their template attack results which where 5.80% and 2.14% respectively.

In general this paper showcased that there are multiple approaches to machine learning and that deep learning isn't always the best solution or only solution, but that it is important to test multiple models and techniques to find the best solution. Machine Learning is itself an type of art and trial and error is a big part of the process which is why even if there are noticeable

improvements machine learning isn't just something you can throw at a problem, and in some cases it can be a detriment.

5 Future Directions

This final section will be to talk about the future ML has inside of side-channel attacks as well as how machine learning can sometimes tell the wrong story.

The main reason why I even know what side-channel attacks are are from a blog post [2] for this paper from Cook et al [7] which details their research into a side-channel attack which exists not in any cryptographic machine but rather your own browser. To summarize their research it allowed for websites to abuse a side-channel attack to identify what website you were visiting using machine learning – this is a relatively new idea presented by Shusterman et al [8] dubbed **website fingerprinting** – , but the actual source of the side-channel attack wasn't fully understood. This is because Shusterman et al used machine learning to identify the website being visited. This essentially put a black box in the side-channel analysis process. The paper by Cook et al went about to not only identify a more simple approach to use machine learning to identify the website but they also identified the source of the side-channel attack which was originally thought to be a cache side-channel, but was actually a system-interrupt-based side-channel that was being utilized for website fingerprinting.

Both the paper by Cook et al and the blog post by Cook are excellent introductions to side-channel attacks but they also represent a real problem with machine learning. Cook titled his paper "There always is a bigger fish" because machine learning can end up hiding important underlying properties. As we have seen it is a powerful tool both in side-channel analysis and in general for a multitude of problems but nothing is ever as good as it seems, and there can always be room for the black box to overtake and obfuscate what is actually happening. This is why it is important to always be critical of the results that machine learning gives you, and to always attempt to look for bigger fishes.

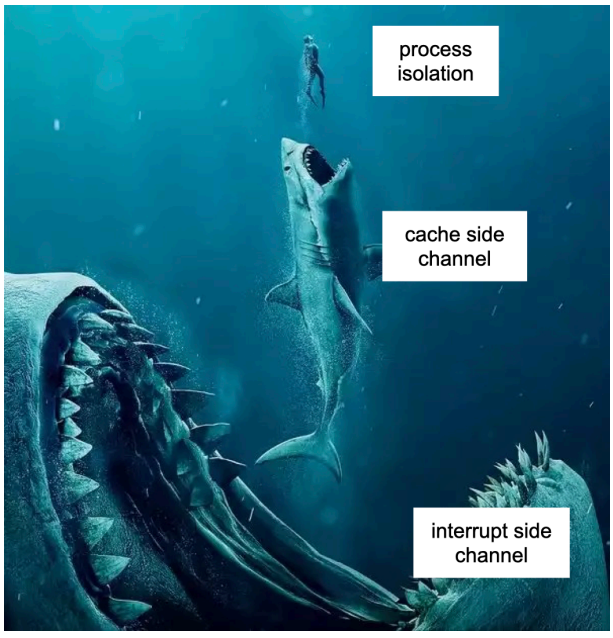

Figure 8: There always is a bigger fish [2]

Bibliography

[1] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards*, 1st ed. New York, NY: Springer, 2007, p. xxiv+338. doi: 10.1007/978-0-387-38162-6.

[2] "When Machine Learning Tells the Wrong Story — jackcook.com." 2024.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[4] J. van Woudenberg, "Lowering the Bar: Deep Learning for Side Channel Analysis." 2018.

[5] G. Perin and B. Ege, "Lowering the Bar : Deep Learning for Side-Channel Analysis (WhitePaper)," 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:199371076

[6] L. Lerman, G. Bontempi, and O. Markowitch, "Power analysis attack: An approach based on machine learning," *Int. J. of Applied Cryptography*, vol. 3, p. ied Cryptography, 2014, doi: 10.1504/IJACT.2014.062722.

[7] J. Cook, J. Drean, J. Behrens, and M. Yan, "There's always a bigger fish: A clarifying analysis of a machine-learning-assisted side-channel attack," *IEEE Micro*, vol. 43, no. 4, pp. 28–36, Jul. 2023, doi: 10.1109/mm.2023.3273457.

[8] A. Shusterman *et al.*, "Robust Website Fingerprinting Through the Cache Occupancy Channel," in *28th USENIX Security Symposium (USENIX Security 19)*, Santa Clara, CA: USENIX Association, Aug. 2019, pp. 639–656. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/shusterman