WASP Software Engineering Course Module Report

Maggie Tran

August 16, 2025

1 Introduction of Research Area

We research side-channel attacks on post-quantum cryptosystems. Shor's algorithm and development in quantum computers pose a threat to our current public key cytosystems that rely on two hard problems: factoring large integers and the discrete log problem. In 2016, the National Institute of Standards and Technology (NIST) took the initiative to ask the research community to create new cryptographic systems that can resist both classical and quantum computers. In order to ensure and improve the security of these new cryptographic schemes, there is much effort in designing attacks to break them.

Our research focuses on exploiting side channel information that is leaked through the implementation of these new schemes in order to retrieve the secret key. Examples of side-channel information are electromagnetic radiation, timing, and power consumption that can be measured when a device does operations that involve the secret keys. Of these attacks, the most common and stronger attacks are power-based side channel attacks, which is also the focus of our research. One famous example of a simple power analysis attack is an attack on the implementation of RSA on smartcards. The Kerchoff's principle says that the security of the cryptographic system should rely solely on the secrecy of the key, all the information on the encryption and decryption algorithms, including how to implement them should be publicly known. RSA is a public key cryptographic algorithm that performs a series of modular exponentiation loops through each bit of the secret key (exponent `d`).

As can be seen from Figure 1, multiplications consume more power than the squar-

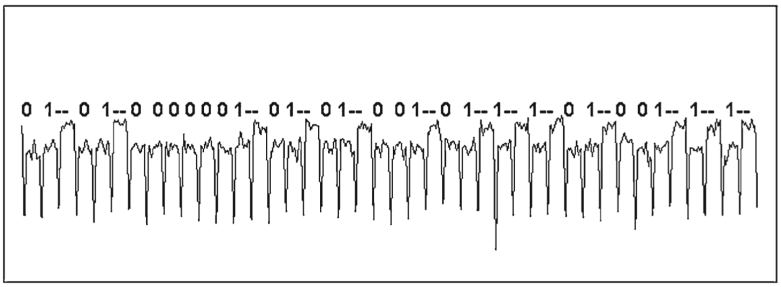

Figure 1: Power leak from an RSA implementation. [3]

ing, resulting in higher peaks in the power trace. While squaring is performed in each iteration of the exponentiation loop, multiplications are only done when a bit of the exponent is 1. Using this fact, we can see the pattern for "1" bit in the exponent being the higher peaks, and "0" being the shorter spikes without a subsequent taller

one. This is how a secret key can be recovered from simple power analysis of the power leakage in the implementation of a cryptographic algorithm.

2 Lecture Principles

1. **Software Testing** is very important in cryptographic systems. Rigorous testing is required to make sure that the implementation performs as intended, is reliable and secure. For my research topic, that includes tests for leakage of side information such as power consumption, whether these signal have structures more than just noise, and can give information on the secret key. Moreover, there are also tests for the effectiveness of proposed countermeasures for known attacks.

2. **Science vs. Engineering**: Scientists ask why the system works while engineers ask how we can make it work reliably, and our research area is very much both science and engineering. Mathematicians and cryptographers design and prove security of cryptographic systems. While engineers are not required to understand all the underlying mathematics for why such systems work or are theoretically secure, there are many details in the implementations that could compromise such systems, case in point, side-channel information leakage.

3 Guest Lecture Principles

1. First lecture: Requirement Engineering,

   (a) Elicitation of Security Requirements
   Identify Security Needs: During the requirements elicitation phase, stakeholders should specify security requirements that address potential vulnerabilities to different side-channel attacks. This includes defining acceptable levels of resistance against such attacks as well as guidelines for code practice during development to protect against attacks such as timing attacks.

   (b) Analysis of Security Requirements
   Feasibility and Impact Assessment: Analyze the feasibility of implementing security measures against side-channel attacks and assess the potential impact of these attacks on the system's integrity and confidentiality.

2. Second lecture: Importance of Human Aspects in Software and AI/ML Engineering

The lecturer raised a point that people could be resistant to change, such as less willing to adapt and use new tools, such as ML, AI. In the area of security, we face similar problems, systems that are known to have been broken in the research community are still in use for many applications. Moreover, despite many known attacks being entirely avoidable, well-studied and effective countermeasures are not adopted as a common practice in industry. One could argue that in this case it was not scepticism that stopped people from adopting new techniques, but lack of knowledge or perhaps industry standards that are difficult to change.

4 Data Scientists versus Software Engineers

I fully agree with the author on the essential differences between data scientists and software engineers in the first two chapters [5]. Firstly, those who work these two roles typically come from different education programs, with software engineering focusing more on software development such as design, testing, and even security,and data science emphasizing more on mathematics, statistics and different machine learning algorithms. This results in two quite different types of expertise. Secondly, the two roles also have very different work experience. While software engineers work more with developing large systems with the intention of deploying and scaling without too much effort, data scientists often work with Jupyter notebooks, in order to quickly develop and test different models. Both types of expertise and work experience are required in developing ML-based products.

In the future, I believe that both roles will not merge but continue to exist. The main reason is the very different educational background and work experience for the two roles. There is much knowledge to obtain in each field alone, hence, it will not be easy to find those who excel at both. However, I also believe that both roles will also learn many of the skills of "the other side". For example, as a data scientist, improvement in programming skills will also result in higher productivity as it will help them set up and test different or newly developed models faster. Better understanding of computer architecture and algorithms will also save computing resources, make it more efficient when developing new models. It is also beneficial for product development if data scientists are aware of good practice in requirement engineering, software deployment and scaling.

5 Paper Analysis

5.1 POLARIS: A framework to guide the development of Trustworthy AI systems [1]

Core ideas and their SE importance

Over the past few years, we have observed the fast development of AI, and how AI has taken part in my critical building blocks of society, such as banking and finance, education, healthcare, drug development, etc.. Along with the benefits also come series of doubts on how trustworthy these systems are. The paper is part of the effort to build guidelines for AI practitioners to apply in the development of their products. The authors review the current research and practice of Trustworthy AI and analyse the results from their surveys and interviews with AI professionals to find out the different challenges and what is lacking. They then propose POLARIS, a framework with different components detailing what actions AI practitioners can take during the software development life cycle (SDLC) of their products to ensure AI trustworthiness.

Relation to my research

As explained in section 1, my research's focus is on side channel attacks on post-quantum cryptographic systems. While we do use deep learning techniques in some attacks, we do not have any focus on developing AI algorithms or AI-enabled systems. Due to the fast development of AI, I find the topic of AI trustworthiness to be very interesting and relevant.

Integration into a larger AI-intensive project

From the survey results, it seemed the AI practitioners questioned were aware of the different solutions to make their AI-enabled systems trustworthy. However, they are concerned that these solutions often lead to system's performance decrease as well as extended time and effort to implement these solutions. One issue that the authors highlight is there is a lack of tools that make AI-enabled systems explainable expressed by the interviewed AI professionals.
The framework, POLARIS, has four knowledge components: Privacy, Security, Fairness and Explainability. For each component, the framework provides identification of different issues related to that component in each SDLC phase, as well as proposed actions to address these issues. For example, for the Security component, they go

through each SDLC phase and identify the threats and subthreats, provide descriptions of such threats, the resulting vulnerabilities, then an actionable plan to address each of those said threats.
POLARIS seems to be a systematic and convenient tool to integrate in the development of a large AI-intensive project. For each phase of the SDLC, professionals can consult the framework and find out what actions to take in order to achieve the four properties: Privacy, Security, Fairness and Explainability in their systems.

Adaptation of my research

In my research, we are the attackers, we try to figure out ways to use side-channel information in order to retrieve the secret keys. We need to assess how realistic and practical the attacks are, as well as find out the countermeasures to these attacks. The framework does not have a direct application to my research. However, it was very interesting to see its Security and Privacy components.

5.2 What About the Data? A Mapping Study on Data Engineering for AI Systems [2]

Core ideas and their SE importance

As discussed in section 4, software engineers and data scientists are two very different roles with different focus in educational background and work experience. In this paper, the author reviews 25 papers on the practice of another also very important role in an AI-enabled system, data engineers, those who are in charge of extracting, moving and preparing data at scale [3]. The contribution of this paper is by analysing and categorizing the data engineering activities, tools, and frameworks, an overview of the current research on data engineering is provided as solutions for AI practitioners. This is also beneficial for researchers to find out what is still lacking in the field.
It is well-known that nowadays, data is considered the most integral part of any AI-based systems, but software engineers are not really equipped with knowledge to handle a large amount of data, nor are data scientists. The data preparation process requires a different type of engineering expertise, although maybe more related to software engineering than data science skills. It is without a doubt a very important part in the development as well as production phase of any AI-enabled systems.

Relation to my research

As mentioned in parts of the analysis of paper [1] in section 5.1, my research is not directly connected to developing new AI and ML algorithms, we do, however, handle a significant amount of time-series data, such as power consumption data. Hence, good practice of data engineering would also be beneficial.

Integration into a larger AI-intensive project

Unlike the first paper [1] introducing a framework, providing guidelines and clear, actionable plans for AI practitioners to make an AI-enabled system trustworthy, this paper is more a structural review, and summary of current research on data engineering. It was mentioned that even just the definition of data engineering was not clearly stated in all 25 selected papers. Moreover, different authors shared slightly different definitions. Reading this study can give AI professionals ideas for good practice in data engineering, as well as direct them to further read other research papers that tackle the problems that they are interested in. Heck [2] pointed out the lack of an overall data architecture since most papers focused mainly on production pipelines for AI systems. The paper also acknowledged the need for guidelines, best practices, and open-source tools to support AI practitioners. Data engineering is clearly an integral part of a large AI-intensive project; however, there is much work to be done to provide data engineers with general frameworks and tools to support them.

Adaptation of my research

Data engineering can be considered a necessary part of my research, as it is common that collecting and processing a significant amount of time-series data is the first step in a study. Storing the data and making it available for other researchers to reproduce our results is also an important part of the research. A general framework or guideline of how to data-engineer will certainly be useful.

6 Research Ethics and Synthesis Reflection

6.1 Search and screening process

Since my research area is somewhat AI-adjacent, we do not study or develop AI algorithms or software engineering frameworks and practices for AI-enabled systems, I could not find any papers on CAIN that are directly connected to my research area. I decided to choose papers that are from more recent years, such as 2024 or 2025,

and topics that I found interesting. However, not all papers that are listed on the conference website are not accessible without a licence. The two papers selected for analysis in section 5 were from arXiv.

6.2 Pitfalls and mitigations

I did not encounter any problems with misleading titles or abstracts. However, as mentioned, I could not find any articles that are strongly related to my research topic on CAIN.

6.3 Ethical considerations

Since the recent research [4] on how using LLM to write your assignments can affect your brains, I have avoided using an LLM service for this report. The two papers were read and analysed, and the report was written by me.

References

[1] Maria-Teresa Baldassarre, Domenico Gigante, Marcos Kalinowski, and Azzurra Ragone. POLARIS: A framework to guide the development of Trustworthy AI systems. pages 200–210, 2024.

[2] Petra Heck. What About the Data? A Mapping Study on Data Engineering for AI Systems. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI*, CAIN 2024, page 43–52. ACM, April 2024.

[3] Paul Kocher, Joshua Jaffe, Benjamin Jun, and Pankaj Rohatgi. Introduction to differential power analysis. *Journal of Cryptographic Engineering*, 1(1):5–27, 2011.

[4] Nataliya Kosmyna, Eugene Hauptmann, Ye Yuan, Jessica Situ, Xian-Hao Liao, Ashly Beresnitzky, Iris Braunstein, and Pattie Maes. Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. 06 2025.

[5] Christian Kästner. *Machine Learning in Production: From Models to Products*. MIT Press, 2025. `https://mlip-cmu.github.io/book/`(visited 2025-08-14).