AI DETECTION ON HARDWARE TROJAN

Abstract

In today's digital landscape, safeguarding data integrity is paramount, yet the stealthy threat of hardware trojans persists, adept at circumventing conventional detection methods by being implemented during circuit fabrication. This study delves into the transformative potential of artificial intelligence (AI) in thwarting hardware trojans, particularly through the analysis of voltage consumption patterns. By meticulously collecting and extracting features from data, an AI model undergoes rigorous training to distinguish subtle anomalies indicative of trojan activities, offering a proactive defense mechanism against potential infiltrations. The hypothesis posits that AI, armed with its adaptive learning capabilities, can effectively discern these patterns, thereby fortifying digital defenses and serving as an early detection system for trojan threats. Despite the inherent complexities of real-world scenarios, this research lays a robust foundation for AI-driven trojan detection, underscoring its promise in bolstering cybersecurity measures and safeguarding critical infrastructures against evolving threats.

Motivation/Problem

When it comes to modern-day technology, there are many ways for our data to be leaked. Despite advancement in security, many forms of security don't cover all the bases.

In this case, Hardware trojans, present a formidable challenge, of being able to take and leak data, because the problem was implemented during the very fabrication of the circuits, traditional detection methods often fall short in identifying these stealthy threats, since trojans can remain dormant during testing phases while the product was being manufactured.

Moreover, the devices that these products can connect to leave the user to critical infrastructures that amplify the potential impact for cyberattacks, especially as more and more of our personal data end up online. So as our reliance on digital technology grows, so must the need for effective detection and strategies against it.

Because of that, Artificial intelligence may hold the key to being able to detect hardware trojans due to its ability to analyze patterns and anomalies, which can be used in this case to detect an issue that isn't present on the time but may show up only occasionally due to it being dormant.

Therefore, addressing and experimenting to see how strong AI can be against the hardware trojans is an important front against the dangers of hardware trojans.

Hypothesis & Methods

Hypothesis:
It should be possible to teach an Artificial Intelligence to effectively learn the patterns of a voltage consumptions pattern and determine the presence of a Hardware Trojan.

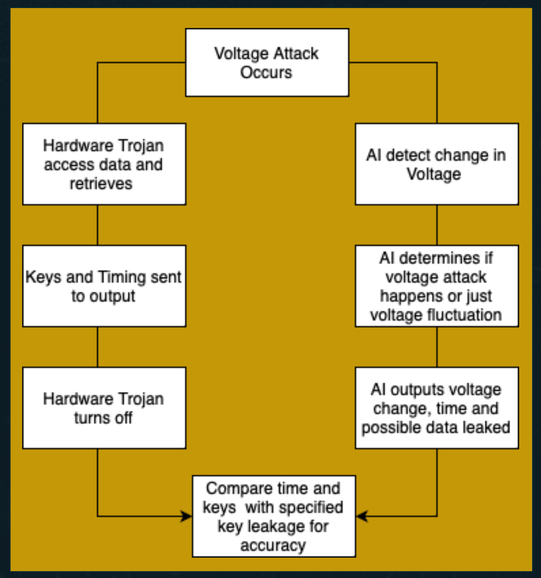
Methods:
Data collection:
Determine proper levels of expected voltage and model train a machine learning algorithm to know when voltage attack occurs.

Features Extraction:
Extract Relevant information like voltage consumption, reasonable admissible fluctuation between power, and times of data keys leaked.

Current Strategy of Work

With the current process, when voltage attacks occur, the hardware trojan will access the leaked data and save the key and time of access as it's output. In that moment, the voltage resets and the Trojan deactivate.

During this time the AI should be able to analyze and determine that the voltage has changed. At this point, giving us the time of the attack and we can compare with the accessed that was leaked.



Discussion

While real-world trojans, provide more complexity and avenues of attack, this is a good base for figuring out if AI could detect as an proof-of-concept.

Although plans include FPGA's, other means of attack is also possible, meaning one form of AI can't detect all.

Future Work

- Software Trojan created to behave similarly to Hardware Trojan
- Training AI Model to determine and recognize patterns for the voltage attacks
- Determine if Side-channel analysis could also be detected by an AI Algorithm
- Determine if using Power Analysis, can determine if a trojan is exfiltrating data

Acknowledgements

Thank you to the Draco Lab and the University of Central Florida for providing the space and equipment for this research.

Yvan Pierre Jr.
yvan.pierre@UCF.EDU
Franco Mezzarapa
franco.mezzarapa@UCF.EDU

DRACO LAB | www.ece.ucf.edu/DRACO
Dr. Mike Borowczak, Lab Director

Dept. Of Electrical & Computer Engineering
College of Engineering and Computer Science
University of Central Florida

DRACO