INNOVATIVE HARDWARE TROJAN INSERTION USING GNNs

BARNABOSS PULI
CWID: 885167783
Electrical and Computer Science Dept.

ABSTRACT

In today's rapidly evolving technological landscape, the security and reliability of electronic systems have become paramount. As these systems become more integral to our daily lives, the threats they face grow in number and complexity. Among these threats, hardware Trojans represent a particularly insidious challenge. These hidden malicious circuits or mechanisms are designed to disrupt the normal operations of electronic systems, making their detection and mitigation critical. The motivation behind our project stems from the urgent need to stay ahead of these increasingly sophisticated and covert threats. By delving into the study of new types of Trojan hardware, we aim to contribute to the ongoing efforts in securing modern electronic systems.

In this work, our main objective is to discover innovative methods to insert hidden, malicious changes into computer hardware that can evade emerging security measures. To achieve this, we employ powerful neural network architectures in machine learning—Graph Neural Networks (GNNs) [1] and Generative Adversarial Networks (GANs) [6]. These technologies help us make our methods both reliable and flexible. Our research approach involves using GNNs to generate and analyze Data Flow Graphs (DFGs) for in-depth investigation into circuit behavior. GANs are applied to dynamically alter Trojan designs, resulting in modified circuits that are difficult to detect. Reinforcement Learning (RL) is then used to refine the placement of Trojans, ensuring their strategic incorporation within the circuit [7].

Our methodology has been validated through testing with machine learning models designed for asynchronous circuits, demonstrating its effectiveness and flexibility in various hardware scenarios. Expected results indicate that combining GNNs, GANs, and RL significantly enhances the success of hardware Trojan deployment, making it harder for traditional detection methods to identify these malicious entities [4].

This research advances cyber-physical security by developing innovative methods for both inserting and detecting hardware Trojans. Our efforts are not merely about surpassing existing security protocols but also about anticipating future vulnerabilities and safeguarding critical electronic infrastructures that are integral to scientific advancements and daily activities. Our initial tests have shown promising results, with GNNs successfully learning and highlighting irregularities in circuits, thereby contributing to more robust and comprehensive security solutions [5].