Synthesis-Level Characterization of Trust-Hub AES Hardware Trojans

Yoonjeong Kim
Dept. of Electronics Engineering
Chungnam National University
Daejeon, South Korea
yjkim.cas@gmail.com

Yujin Eom
Dept. of Electronics Engineering
Chungnam National University
Daejeon, South Korea
yjeom.cas@gmail.com

Soyeon Choi
Dept. of Semiconductor System Engineering
Hanbat National University
Daejeon, South Korea
sychoi@hanbat.ac.kr

Hoyoung Yoo
Dept. of Electronics Engineering
Chungnam National University
Daejeon, South Korea
hyyoo@cnu.ac.kr

Abstract— **The globalization of the semiconductor supply chain increases the risk of hardware Trojan (HT) insertion during design and manufacturing. This paper analyzes pre-silicon structural indicators using Trust-Hub AES and BasicRSA benchmarks. All designs were synthesized under identical Vivado settings, and Look-Up Table (LUT), Flip-Flop (FF), and Configurable Logic Block (CLB) utilization were compared. Results show that internal-triggered Trojans cause notable resource increases, while others remain minimal, demonstrating the potential of synthesis data for early HT detection.**

Keywords; Hardware Trojan;FPGA; Trust-Hubs; AES;

I. Introduction

The globalization of the semiconductor supply chain has increased the risk of hardware Trojan (HT) insertion by untrusted entities, threatening the security and reliability of integrated circuits (ICs). An HT can be introduced during design, IP integration, or fabrication, leading to data leakage, cryptographic failure, or side-channel vulnerabilities. Cryptographic IPs such as AES are especially vulnerable since they directly access secret keys. To support reproducible research on Trojan detection, the Trust-Hub project provides standardized benchmark circuits containing known Trojans with defined insertion methods and trigger conditions. While most existing studies rely on side-channel or functional-testing approaches that require post-silicon measurements[1], pre-silicon synthesis reports offer structural metrics such as LUT, FF, and CLB utilization that may reveal abnormal modifications. This study performs a quantitative analysis of structural resource overhead using six AES and four BasicRSA benchmarks from Trust-Hub. By comparing resource utilization across different trigger types (always-on, internally triggered, externally triggered) and insertion locations, the work identifies structural variation patterns and provides baseline data for developing future resource-based detection methods.

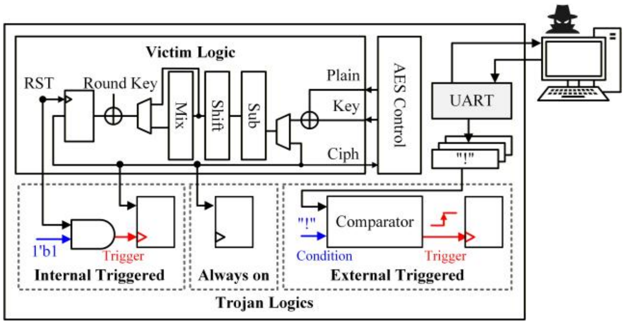

Figure. 1. Activation behavior of always-on and triggered hardware

II. Background

Hardware Trojans are generally classified as always-on or triggered. Always-on Trojans remain active continuously, while triggered Trojans activate only under specific conditions to evade detection. Triggered Trojans can be internally or externally triggered. Internal triggers depend on circuit conditions such as logic states, counters, or environmental parameters like voltage and temperature. External triggers respond to user inputs or signals from external interfaces such as GPIO or serial ports. Detection methods for hardware Trojans fall into three main categories: functional-based, side-channel-based, and structural-based. Functional-based methods rely on test coverage and ATPG analysis, while side-channel-based methods detect anomalies in power, delay, or electromagnetic emissions. In contrast, structural-based analysis examines resource variations such as LUT, FF, and cell utilization to infer Trojan presence without requiring post-silicon measurements. However, studies on how different Trojan types and trigger mechanisms affect hardware resource usage remain limited. The Trust-Hub benchmark suite provides standardized designs with known Trojans inserted into modules such as AES and RS232[2]. Trust-Hub serves as a public benchmark repository for hardware Trojan research, providing both Trojan-inserted and Trojan-free versions of various

This research was supported by the National Research Foundation of Korea (NRF) (No. 2022R1A5A8026986) and the Ministry of Science and ICT (MSIT) under the ITRC program (IITP-2025-RS-2024-00436406 (50%)), and by the IC Design Education Center (IDEC), Korea.

TABLE I RESOURCE UTILIZATION INCREASE AFTER SYNTHESIS

Activation Type		*Trojan IDs*	*LUT*	*FF*	*CLB*
Always-on		AES-T100	2569(100%)	4051(102%)	833(107%)
		AES-T200	2575(101%)	4052(102%)	840(108%)
		AES-T300	2600(102%)	4096(103%)	817(105%)
Trigger	Internal	AES-T500	2738(107%)	3968(100%)	805(103%)
		AES-T1600	2947(115%)	4250(107%)	1011(129%)
		AES-T1900	2731(107%)	3969(100%)	804(103%)
		RSA-T300	599(108%)	491(107%)	125(116%)
		RSA-T400	609(109%)	491(107%)	118(110%)
	External	RSA-T100	597(107%)	459(100%)	113(105%)
		RSA-T200	487(87%)	366(80%)	91(84%)

modules[3]. This enables quantitative comparison of circuit metrics such as area, power, and timing overheads, and allows detection methods and structural impact analyses to be validated under standardized conditions. In this study, the AES and BasicRSA benchmarks from Trust-Hub are utilized to analyze the structural impact of Trojan insertion on FPGA resources, providing foundational data for the development of future resource-based detection techniques. Specifically, this work quantitatively compares the resource variations (LUT, FF, and CLB) between Trojan-inserted and golden designs using six AES and four BasicRSA benchmarks provided by Trust-Hub. All designs were synthesized in Vivado under identical settings, and post-implementation reports were analyzed to evaluate the percentage resource overhead. The primary objective of this study is to identify structural variation patterns associated with different Trojan trigger types and insertion locations, thereby revealing how Trojan characteristics influence FPGA resource utilization.

III. EXPERIMENTAL RESULTS

The experiments were conducted using a Xilinx KCU105 board equipped with a Kintex UltraScale XCKU040 device and Vivado 2021.1 design suite. Table 1 summarizes the synthesis results, comparing the resource utilization (LUTs, FFs, and CLBs) between Trojan-inserted and golden designs for the AES and BasicRSA benchmarks from Trust-Hub. For the AES benchmarks, always-on Trojans exhibited modest overhead, increasing LUT utilization by approximately 1~2%, FFs by 2~3%, and CLBs by 0~8%. In contrast, internally triggered Trojans showed a significant increase, with LUT utilization rising by 7~15%, FFs by 0~7%, and CLBs by 3~29%. Among them, AES-T1600 displayed notably higher resource usage due to its complex internal counter and asynchronous control logic. For the RSA benchmarks, internally triggered Trojans resulted in LUT overheads of 8~9%, FFs of about 7%, and CLBs of 10~16%, whereas externally triggered Trojans induced smaller variations approximately 7% in LUTs, 5% in CLBs, and negligible change in FFs. Notably, RSA-T200 exhibited lower resource utilization than its golden version, as its embedded Trojan logic effectively disables part of the original design.

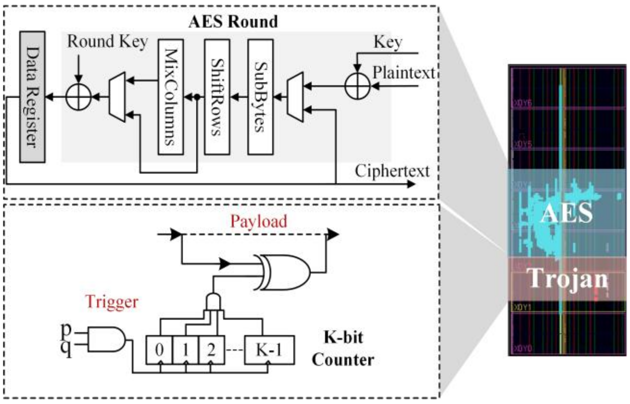

Figure. 2. Structure and synthesis results of the AES encryption module with inserted hardware Trojan

IV. CONCLUSION

This study analyzed FPGA resource overheads caused by hardware Trojan insertion using the Trust-Hub AES and BasicRSA benchmarks. Under uniform synthesis conditions, internally triggered Trojans showed the most significant increases in LUT, FF, and CLB usage, while always-on and externally triggered types caused minimal change. These results demonstrate that pre-silicon synthesis data can effectively reflect structural anomalies, providing a foundation for developing lightweight, resource-based Trojan detection methods.

REFERENCES

[1] A. Jain, Z. Zhou, and U. Guin, "Survey of Recent Developments for Hardware Trojan Detection," *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5.

[2] A. F. M. Bomfim and J. A. M. Nacif, "Systematically Classifying TrustHub Hardware Trojan Benchmarks," Proc. SBMicro SForum: Integrated Circuits and Systems Design Symposium, pp. 1–6, 2023

[3] S. K. Kumar, R. Chanamala, S. R. Sahoo, and K. K. Mahapatra, "*An improved AES hardware Trojan benchmark to validate Trojan detection schemes in an ASIC design flow*," in *Proc. IEEE Int. Conf. VLSI Design and Embedded Systems*, pp. 1–6,