Tolulope A. Odetola received his Bachelor's degree from Obafemi Awolowo University Nigeria in Electronic and Electrical Engineering. He received his Master's degree in Electrical and Computer Engineering from Tennessee Technological University. He is currently a Graduate Research Assistant in the Department of Electrical and Computer Engineering, Tennessee Tech. University, since August 2017 and pursuing his Ph.D. degree in the same department. He works under the supervision of Dr. Syed Rafay Hasan. His research interests include several topics involving security of hardware accelerator-based CNN inference. His research areas include machine learning, FPGA, High level synthesis and so on. He has over five years of vast and robust academic/teaching experience in different

EDUCATION

Doctor of Philosophy, Engineering, Tennessee Technological University, August 2022 (expected), Cookeville, Tennessee

Master of Science, Electrical and Computer Engineering, Tennessee Technological University, May 7, 2021 Cookeville, Tennessee

Bachelor of Science, Electronic and Electrical Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria

College of Engineering

TENNESSEE TECH

The Department of

Electrical and Computer Engineering

Announces the Dissertation Defense of

Tolulope A. Odetola

In Partial

Fulfillment of the Requirements

For the degree of

Doctor of Philosophy in Engineering

July 21, 2022

10:00 a.m.

208 Brown Hall

Tennessee Tech University

FIELD OF STUDY

Deep Learning, Convolutional Neural Network, Hardware Security, Edge Intelligence, FPGA, Hardware Trojan, Hardware Intrinsic Attack, Hardware-Software Co-Verification, High Level Synthesis

DISSERTATION TOPIC

HARDWARE VERIFICATION AND SECURITY CHALLENGES IN DISTRIBUTED CONVOLUTIONAL NEURAL NETWORK INFERENCE

EXAMINING COMMITTEE

Syed Rafay Hasan, Committee Chair
Associate Professor, Electrical and Computer Engineering

J. W. Bruce
Associate Professor, Electrical and Computer Engineering

Mohamed Mahmoud
Associate Professor, Electrical and Computer Engineering

Ghadir Radman
Professor, Electrical and Computer Engineering

Douglas Talbert
Professor, Computer Science

AN ABSTRACT OF A DISSERTATION

Convolutional Neural Networks (CNN) have shown impressive performance in computer vision, natural language processing, and many other applications, but they exhibit high computations and substantial memory requirements. To address these limitations, the use of cloud computing for CNNs is becoming more popular. This comes with privacy and latency concerns that have motivated the designers to develop embedded hardware accelerators for CNNs. Hardware accelerators like FPGAs have become a popular choice for deploying Convolutional Neural Network (CNN). In literature, researchers have explored the deployment and mapping of CNN on FPGA, but there has been a growing need to do design-time hardware-software co-verification
of these deployments. One of the first goals of this dissertation is to investigate a 2-Level 3-Way (2L-3W) hardware-software co-verification methodology and provide a step-by-step guide for the successful mapping, deployment, and verification of CNN on FPGA boards. The 2-Level verification serves the purpose of ensuring the implementation in each stage (software and hardware) is following the desired behavior. The 3-Way co-verification provides a cross-paradigm (software, design, and hardware) layer-by-layer parameter checks to assure the correct implementation and mapping of the CNNs onto FPGA boards. However, designing a specialized accelerator increases the time-to-market and cost of production. Therefore, to reduce the time-to-market and access to state-of-the-art techniques, CNN hardware deployment on embedded accelerators is often outsourced to untrusted third parties (commonly known as 3P intellectual property designers or simply 3PIPs) which raises security concerns. Hence, the security of inference phase deployment of CNNs in hardware accelerators is another research problem that this dissertation addresses. To secure the CNN hardware, part of the CNN computation unit can be outsourced to third party FPGA designers, without providing the sensitive information like initial and final classification layers to these untrusted 3PIP designers. Next, this dissertation demonstrates, for the first time to the best of author's knowledge, that it is possible to insert, we demonstrate that hardware intrinsic attack (HIA) in such a "secure" design. Three different HIAs are proposed, namely: Shuffling of weights and feature maps, Feature map based Stealthy Hardware Intrinsic (FeSHI) attack and Layer Based Noise Injection (LaBaNI) attack. Novel triggering mechanism, exploiting the probability distribution of each individual CNN layer, is investigated. Moreover, unique and stealthy payload techniques are developed that leads to misclassification. These attacks are non-periodic and completely random; hence it becomes difficult to detect and the require minimal additional hardware resources (about ≤ 20%) for their implementation. With the susceptibility CNNs hardware designs to hardware intrinsic attacks and adversarial examples. We propose three defense approaches for the detection and mitigation of adversarial and hardware intrinsic attacks. These defense mechanisms are targeted to defend not only the proposed attacks but also traditional adversarial attacks, with success rate ranging from 45-99%.