SECURE MACHINE LEARNING FOR REAL-TIME EDGE

Zarin Mira

College of Engineering
Natural Sciences and Engineering Oral Presentation

Abstract: Machine learning applications are becoming increasingly popular on various computing platforms such as edge computing devices, cloud servers, IoT devices etc. This is because they enable designers to create control systems that are more intelligent, autonomous, and privacy-aware. However, these applications usually come with heavy workloads, which is why edge applications require "real-time" guarantees. This means that application tasks must be completed within a predefined timing bound, often referred to as "deadlines. Unfortunately, current off-the-shelf ML frameworks like Caffe, Darknet, and cONNXr do not consider the strict delay requirements imposed by real-time applications. Moreover, they are often not "security-aware," which can compromise the confidentiality of information, the processing of data, and the integrity of the application tasks. To address these issues, ML applications may require isolated execution in a trusted environment that is protected from other applications. For instance, when critical applications like biometrics use machine learning, they must execute in a trusted environment to prevent them from being compromised by other applications. One way to provide security for ML tasks is to execute some (or all) of the machine learning layers within a "confined environment" known as trusted execution environments (TEEs). TEEs offer tamper-resistant execution and are available in recent trusted execution technologies such as SGX, Arm TrustZone, and RISC-V. In this research project, we focus on the Arm TrustZone TEE and investigate its integration with existing ML frameworks within real-time schedulers and a confined environment (OP-TEE) with limited resources. Our goal is to explore the feasibility of adopting secure, TEE-enabled ML frameworks for real-time applications and to evaluate the implementation of isolated execution of ML workloads. We will also measure the performance overhead (timing, CPU, memory usage) of such integrations. The outcome of this research will enable us to understand the benefits and limitations of using Arm TrustZone for secure machine learning on real-time edge devices.

Faculty Mentor: *Monowar Hasan*