Securing Serverless Edge AI for the Cloud-Edge Continuum

Zhou Zhou, Doctoral Student
Supervisors: Monowar Bhuyan and Erik Elmroth

Department of Computing Science
Umeå University, Umeå SE-90187, Sweden
zhouz@cs.umu.se

May 13, 2024

Abstract

Serverless is an emerging technology that significantly empowers the cloud-edge continuum paradigm since its intrinsic characteristics include user-friendly, pay-as-you-go, strong elasticity, and on-demand resource availability. Edge AI (artificial intelligence) aims to bring intelligence close to the device, where data are generated that includes diverse machine learning applications. The combination of serverless technology and edge AI becomes a modern technosphere termed serverless edge AI. However, ignoring the security of serverless edge AI hinders its widespread applications and the systems as well. Based on our investigation and analysis, the security issue in serverless edge AI is two-fold: secure edge AI and serverless system. For the former, the AI models, in principle, can be employed for diverse tasks, either for optimising resource usage or for orchestration of resources on the serverless platform when deployed in the cloud-edge continuum. The vulnerabilities of AI models could be exploited by attackers and, for example, manipulated data, model parameters or decisions. For the latter, the current serverless system also has some distinctive vulnerabilities, like shared resources and intermittent communication between services, which could also expose the serverless system to attackers. Further, when it combines serveless edge AI and cloud-edge continuum, these problems boost difficulty. Therefore, devising offensive strategies to assess, explore, and provide secure solutions for the serverless edge AI when deployed in the cloud-edge continuum. In this presentation, the primary focus will be on assessing the impact on poisoning attacks in machine learning models and serverless systems.