Ph.D. in Information Technology
Thesis Defense

May 22nd, 2025
at 11:00 am
"Emilio Gatti" Conference Room – building 20

Inacio Gaspar TICONGOLO – XXXVII Cycle

Dependency-Aware Placement and Resource Allocation in Serverless Edge Computing

Supervisor: Prof. Giovanni Ennio Quattrocchi

Abstract:

Serverless computing has become a promising model for building scalable and flexible applications, particularly in the context of Multi-access Edge Computing (MEC). MEC involves bringing computation closer to the end users by leveraging edge nodes, such as 5G base stations, to meet the stringent latency requirements of modern applications like autonomous driving, mobile gaming, and augmented reality. By decentralizing computation, MEC significantly reduces latency and enhances performance for latency-sensitive applications. However, deploying serverless functions on MEC environments presents considerable challenges due to the inherent resource constraints of edge nodes. These nodes often have limited CPU, memory, and storage, which complicates efficient resource allocation and function placement. Moreover, the complexity of managing execution inter-dependencies among functions adds an additional layer of difficulty. Furthermore, the highly fluctuating and unpredictable nature of workloads, driven by users' geographic movements, exacerbates the issues of resource allocation and function placement. Existing solutions often fail to adequately address these challenges, leading to inefficient resource usage, increased latency, and suboptimal performance. This thesis proposes two dependency-aware approaches, NEPTUNE+ and PLUTO, which aim to optimize resource allocation and function placement in MEC environments, respectively. Both approaches are built upon the NEPTUNE framework, which was originally designed to manage serverless functions in edge environments, but with limited consideration for function dependencies. NEPTUNE+ introduces improvements to the resource allocation process by explicitly considering function dependencies, leading to up to a 67% reduction in CPU core usage when compared to the original NEPTUNE framework, as well as baseline approaches from the state-of-the-art while maintaining comparable response times and complying with service-level objectives (SLOs). In addition, PLUTO places functions with complex dependencies in MEC environments. In a comparison with five state-of-the-art placement strategies, PLUTO demonstrated clear outperformance, reducing latency by up to 90% and memory usage by up to 87% compared to the other approaches and showcasing its ability to optimize the deployment of serverless functions in resource-constrained edge environments. Through extensive simulations using both synthetic and real-world workloads, the thesis demonstrates that NEPTUNE+ and PLUTO effectively optimize resource usage and improve the performance of serverless applications in MEC environments. By addressing the critical challenges associated with function

dependencies, resource allocation, and placement, these approaches enable serverless applications to be more efficient, scalable, and performant in edge computing scenarios, especially in the face of fluctuating and unpredictable workloads.

Keywords: serverless, edge computing, function dependencies, resource allocation, placement.

PhD Committee

Prof. **Pierluigi Plebani, Politecnico di Milano**

Prof. **Catia Trubiani, GSSI**

Prof. **Patrizia Scandurra, Università degli Studi di Bergamo**