PROJECT SUMMARY

Overview:

Mobile Edge Computing (MEC) has been regarded as a pillar technology that enables various latency-sensitive applications in the next generation wireless networks (such as Augmented/Virtual Reality, autonomous driving as well as massive Internet of Things (IoT) video analytics) by deploying cloud computing capabilities at the edge of wireless networks. Recently, serverless computing has emerged as a new event-driven cloud computing paradigm that enjoys many advantages over traditional cloud computing including automatic servers management, pay-as-you-use pricing, massive parallelism, all of which makes it an attractive approach for MEC to overcome the resource underutilization and overpay issues along with traditional cloud computing models. However, both the resource-constrained nature of edge nodes and the latency-sensitive nature of (bursty) IoT workloads result in fundamental challenges to fulfill the promise of serverless computing in MEC networks from the perspectives of both service providers and application developers. To this end, the overarching goal of this project is to develop efficient serverless mobile edge computing networks by designing latency-driven requests dispatching and resource management schemes for service providers as well as investigating the potential of massive parallelism offered by serverless computing to application developers, relying on tools from both queueing theory and online learning, i.e., Bayesian optimization (BO).

Keywords: Mobile edge computing; Serverless computing; Wireless networks; Load balancing; Online learning; Bayesian optimization;

Intellectual Merit:

The proposed research addresses both theoretical and application challenges in developing efficient serverless mobile edge computing networks. Our research will offer versatile analytical techniques and tools as well as algorithms, software programs, and testbed prototypes, which will help achieve cost-effective latency-guaranteed serverless mobile edge computing applications. Due to the unique scientific and engineering challenges in this area, the success of this project will integrate expertise from data networking, machine learning, stochastic control and optimization, algorithmic design, wireless systems, and cloud/edge computing. To unleash the full potential of serverless mobile edge computing, our research is organized around three inter-related thrusts: (i) develop locality-aware load balancing schemes for serverless requests that guarantee optimal response time, (ii) develop cost-effective latency-driven resource configuration schemes via BO, and (iii) develop principle techniques that allow application developers to utilize the great potential of parallelism offered by serverless computing. We will address the unique challenges and develop the analytical foundations for efficient serverless mobile edge computing netowrks, resulting in provably efficient frameworks and algorithms integrated with open-source implementations, which will be extensively evaluated using local testbeds and the NSF-funded COSMOS platform.

Broader Impacts:

From a societal impact perspective, mobile edge computing has enabled various IoT applications across different aspects of the society including security and surveillance, healthcare, industry manufacturing, agriculture, etc. Thus, an efficient serverless mobile edge computing network will directly impact the society at-large. Moreover, our developed algorithms for resource configuration via BO can be a direct add-on for existing serverless computing platforms and our proposed parallelism-assisted BO will be a benchmark for the community at-large when dealing with parallel-computing involved machine learning edge applications. From an education perspective, our work will develop curriculum that will be broadly shared, as well as train undergraduate and graduate students. Special attention will be devoted to recruiting a diverse group of students and provide them opportunities for career development. Finally, we will reach out to middle and high school students to increase awareness in STEM through targeted programs at WSU.