Abstract

Next-generation mobile networks (5G/B5G) are tailored to enable vertical industries to provide a diverse range of virtualized services to their users. However, the growing complexity of mobile services poses challenges in meeting demanding performance requirements. Specifically, this becomes significant with varying service and traffic demands over time without resorting to costly overprovisioning. Besides, the trade-off between spectral efficiency and the consumption of data processing resources presents a complex scheduling dilemma. To elaborate, the scarcity of radio spectrum necessitates efficient spectrum sharing to meet various service level agreements (SLAs). Simultaneously, the limited computing resources at the network edge underscore the importance of allocating virtualized resources in a computationally aware manner. The thesis elaborately discusses the above-mentioned issues and simultaneously proposes solutions that mitigate the problem.

Initially, to gain practical insights and key inputs to design efficient resource management in virtualized Radio Access Networks (vRANs), we investigate and characterize the computational and memory requirements of vRANs by developing a srsRAN-based test-bed. Through experiments, we profile the consumption of computing and memory resources, and we assess the system performance. Additionally, we construct regression models to predict system behavior with varying numbers of connected users and diverse radio transmission settings. This involves developing a methodology and prediction models to aid in the design and optimization of virtual RANs.

A step further, leveraging our experimental findings, we formulate a cost-efficient radio resource management in 5G featuring network slicing. We maximize the profit of all slices simultaneously guaranteeing the target data rate and delay specified in the SLAs for the different traffic flows. The numerical results substantiate the effectiveness of the solution, demonstrating a 10-15% reduction in radio resource

consumption for cost-efficient resource slicing while also accomplishing performance isolation and meeting the data rate and delay specified in the SLAs of, respectively, eMBB and uRLLC slices

Furthermore, this work contributes to the study of distributed RAN orchestration and edge resource management using reinforcement learning. The approach enables decision-making logic to be co-located with the services it controls, facilitating local fine-grained and low-latency actions. A key development is the VERA framework, designed for managing resources at the edge. VERA specifically addresses the concurrent execution of user applications and network functions, recognizing the interconnected resource requirements of these services. Taking into account LTE vRAN and a video transcoder as services, our proposed framework, VERA, is demonstrated to consistently meet service Key Performance Indicators (KPIs) over 96% of the observation periods.