Preface

Since the past decade, the world has seen exponential growth of the Internet of Things (IoT). These things integrate the cyber world with the physical world by sensing and collecting data from the surrounding environment and transmitting it to other devices over the Internet. The rapid growth of IoT fueled the development of advanced apps. Due to limited resources, many resource-hungry IoT apps cannot process sensory data on IoT devices. To address this issue, IoT app users can offload complex computational tasks to the cloud. In recent years, apps that require low latency have emerged, e.g., interactive gaming, natural language processing, face recognition, etc. However, due to unpredictable core network latency and expensive bandwidth, the cloud often fails to meet the stringent requirements of such latency-sensitive apps. Edge computing technology is proposed to address these challenges.

Edge computing reduces latency and saves bandwidth by deploying edge servers in different geographic locations. This technology plays a crucial role in the rapidly growing app market for IoT devices. In edge computing-enabled IoT, the app vendors hire resources from edge servers and allocate them to app users to overcome the challenge of the limited computing capacities of their IoT devices. An effective allocation of edge computing and networking resources to different apps or app users is needed because the resource allocation affects the system's performance in many aspects, as follows: 1) It affects the app users' Quality of Service (QoS) measured in different parameters such as packet loss, bit rate, throughput, transmission delay, availability, jitter, etc. 2) The

resource allocation affects the benefits and the cost of the app vendors. 3) The resource allocation also impacts the resource utilization of the edge servers and the benefits of the edge infrastructure providers. This thesis investigates the edge resource allocation problem from the perspectives of all the actors.

In this thesis, we divide the edge resource allocation problem into two sub-problems: the computing resource allocation problem and the networking resource allocation problem. The problem of edge computing resource allocation is viewed in terms of offloading and executing computational tasks. We allocate networking resources for group communication between the edge server and various locations in the problem of edge networking resource allocation. The contributions of this thesis to solving the problem of resource allocation to app users are as follows:

- In the first work, we investigate the problem of allocating the app user to the hired resources of multi-tenant edge servers by the app vendor while establishing a trade-off between the users' QoS and cost. This problem is studied from the perspective of individual app vendors and their app users and is referred to as the App User Allocation (AUA) problem. We proposed a game-theoretic approach to solve the AUA problem that formulates this problem as a potential User Allocation Game (AUGame).

- In the subsequent work of allocating the edge computing resources, we consider how the different app vendors (services) compete for the same edge computing resources. Here, each app vendor intends to hire optimal bundle resources and serve the most app users at the lowest cost. This problem is referred to as the Edge Resource Allocation (ERA) problem. We propose a game-theoretic approach to solve this problem, which optimizes the cost incurred by the app vendors while maximizing resource utilization. In this approach, the ERA problem is formulated as a potential Edge Resource Allocation Game (ERAGame). This game employs an ERA algorithm to reach the solution faster.

- In the following work, we investigate resource allocation from the perspective of edge infrastructure and app users to maximize resource utilization while improving the users' QoS. We propose a distributed user allocation approach that finds the bottleneck resource on each multi-tenant edge server and balances the load on them. This approach accommodates the geographically dynamic density of app users by moving the app users from the overloaded edge servers to the underutilized edge servers, which efficiently utilizes the computing resources of the multi-tenant edge servers.

- The final work allocates the edge networking resources for group communication from an edge server to app users. A cost-effective multicast tree for group communication is required to use the edge networking resources efficiently and effectively. We consider multiple objectives for building a cost-effective multicast tree, e.g., high throughput, low delay, low energy consumption, etc. These multiple objectives can be efficiently combined in the form of edge costs. In this work, a weighted cost-sharing scheme is proposed to divide the cost of network edges among their users. We then proposed a game-theoretic approach that constructs a cost-effective multicast tree using the proposed cost-sharing method. While building the optimal tree, this approach aims to maximize the throughput and reduce the delay and energy consumption for data transmission.

According to the literature, the problems mentioned above are NP-Complete. Therefore, we find the approximate solution in polynomial time and examine its convergence and optimality. Theoretically, we demonstrate that the suggested solutions are much nearer to the centralized optimum. The numerical results substantiate the theoretical analysis.