國立臺灣大學電機資訊學院電信工程學研究所

碩士論文

Graduate Institute of Communication Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

公平拍賣機制於行動邊緣運算下之資源分配與定價啟發式最佳化

A Fairness-Aware Auction Mechanism with Heuristic-Based Optimization for Resource Allocation and Pricing in Mobile Edge Computing Systems

林家頡

Chia-Chieh Lin

指導教授: 蔡志宏 博士

Advisor: Zsehong Tsai, Ph.D.

中華民國 112 年 7 月

July, 2023

doi:10.6342/NTU202302400

國立臺灣大學碩士學位論文
口試委員會審定書
MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

公平拍賣機制於行動邊緣運算下之資源分配與定價啟發式最佳化

A Fairness-Aware Auction Mechanism with Heuristic-Based Optimization for Resource Allocation and Pricing in Mobile Edge Computing Systems

本論文係林家頡R10942144在國立臺灣大學電信工程學研究所完成之碩士學位論文，於民國112年7月26日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Institute of Communication Engineering on 26 July 2023 have examined a Master's thesis entitled above presented by Chia-Chieh Lin R10942144 candidate and hereby certify that it is worthy of acceptance.

口試委員Oral examination committee:

______ ______ ______
(指導教授Advisor)

林風 ______ 鐘耀梁

______ ______ ______

系主任/所長Director: ______

doi:10.6342/NTU202302400

Acknowledgements

首先我要感謝我的指導教授：蔡志宏博士，感謝教授兩年來的指導，讓這本碩士論文得以順利完成。從教授身上我學習到做事要腳踏實地，要注重實務方面是否能夠達成，而非空談理論。教授學識淵博，對各個方面都有深厚的認知，能夠一眼就看出關鍵問題所在，每次和教授討論都使我獲益良多。

感謝爸媽讓我衣食無憂，你們對我無條件的支持與愛是我能夠勇往直前的原因。感謝爸媽和兩位姊姊以及泰迪在生活上的陪伴，讓我的生活充滿歡笑，使我感受到家庭的溫暖與幸福。

感謝我的女朋友，穎康，你讓我的生活充滿了色彩，讓我明白了如何真正的去愛一個人。你的愛讓我幸福，讓我成為了一個更好的人。

感謝實驗室的同學，翊銘和安浩，你們陪著我一起度過了兩年的研究生生活，激勵了我在研究上的各種想法，使我在做研究方面的能力得以成長，謝謝你們。

這本論文得以完成，要感謝的人太多了，就謝天吧。

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

中文摘要

移動邊緣運算是一種有前途的技術，可提供計算和無線資源，並縮短與移動終端用戶的通訊延遲。移動邊緣運算位於整個網絡架構的邊緣，對於計算資源受限的移動終端用戶來說，移動邊緣運算可以用於卸載應用，而且在距離上比起傳統的雲數據中心更為接近。然而，由於移動邊緣運算的資源有限，因此需要一個好的資源分配方式，以有效地分配資源給終端用戶，並防止資源被低效利用。拍賣機制非常適合用以實現資源的最佳分配，並為資源提供者和資源使用者提供參與資源分配交易的誘因。

目前大部分將拍賣機制應用於移動邊緣運算中的資源分配的研究都沒有綜合考慮計算和無線資源的分配，這些研究都只專注於計算資源的分配，而這會使移動終端用戶面臨暴露問題，最終導致較低的社會福利。此外，大多數拍賣模型的研究都僅關注最大化社會福利，很少考慮拍賣的公平性，但是忽視公平性通常會導致某部分的資源使用者的服務匱乏，造成使用者流失，使用者流失可能會導致寡頭市場的產生，這給剩餘的使用者降低出價的議價權，從而降低社會福利。

為了解決上述問題，我們首先提出了一個單輪拍賣機制，該機制在移動終端用戶卸載任務期限約束下，共同分配資源提供者（即移動運營商）的計算和無線資源給資源使用者（即移動終端用戶），同時最大化社會福利。我們將此資源分配問題建模為一個整數線性規劃問題。為了有效地解決這個整數線性規劃問題，我

doi:10.6342/NTU202302400

們提出了一種貪心近似演算法，它提供了一個與商業求解器的解決方案相比具有快速執行速度的近似最優解。其次，基於一輪拍賣機制，我們提出了一個公平拍賣機制，用於公平分配資源，以解決使用者流失的問題，該機制提供了一種可以有效且公平地分配計算和無線資源給移動用戶的方法。最後，大量模擬實驗的結果證明我們提出的模型確實可以在最大化系統的整體福利之餘，解決暴露問題和使用者流失問題。在這篇研究當中，我們的主要貢獻是提供了一種在資源競爭的移動邊緣運算環境中，使用拍賣機制來有效、高效、且公平地共同分配計算和無線資源給移動用戶的方法。

關鍵字：行動邊緣運算、資源分配、定價機制、拍賣機制、公平拍賣機制

doi:10.6342/NTU202302400

Abstract

Mobile edge computing is a promising technology that provides computing and wireless resources with lower communication latency to the end user. Deployed at the network's edge, mobile edge computing is geographically closer for resource-restricted mobile devices to offload their task than traditional cloud data centers. However, with limited resources, a well-designed resource allocation scheme is needed to efficiently allocate resources to end users and prevent resources from under-utilization. The auction mechanism is well-suited for achieving optimal allocation strategy and incentivizes resource providers and consumers to participate in the resource allocation trade.

Most existing works that apply auction mechanisms failed to jointly allocate computing and wireless resources while maximizing social welfare under task deadline constraint, leading to the exposure problems for mobile users and minor social welfare. Furthermore, most auction-based models focus only on maximizing social welfare. Few have considered fairness among auction participants, but without considering fairness leads to the

doi:10.6342/NTU202302400

bidder dropout problem due to service starvation in a recurrent auction environment, and the bidder dropout problem may lead to an oligopoly market, which gives the remaining bidder bargaining power to lower their bid, thus decreasing social welfare.

To address the problems mentioned above, we propose a one-round auction mechanism, which jointly allocates computing and wireless resources of the resource provider, i.e., the mobile operator, to consumers, i.e., mobile users, while maximizing social welfare under users' offloaded tasks' deadline constraint. We modeled this resource allocation problem as an integer linear programming problem. To solve this integer linear programming problem efficiently, we proposed a greedy approximation algorithm, which provides a near-optimal solution with fast execution speed compared to the commercial solver's solution. Second, based on the one-round auction mechanism, we proposed a fairness-aware auction mechanism for fair resource allocation to address the user's dropout problem. This mechanism provides a way to efficiently and fairly allocate computing and wireless resources to mobile users while preventing the bidder dropout problem from arising. Finally, we performed extensive simulations, proving that our proposed model can solve the exposure and bidder dropout problems while maximizing the social welfare. In brief, our main contribution is providing an effective and efficient way to jointly and fairly allocate computing and wireless resources to mobile users using a fairness-aware auction mechanism

doi:10.6342/NTU202302400

in a resource-competitive MEC environment.

Keywords: mobile edge computing, resource allocation, pricing mechanism, auction mechanism, fairness-aware auction mechanism

doi:10.6342/NTU202302400

Contents

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

List of Figures

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

List of Tables

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

Chapter 1 Introduction

1.1 Background and Motivation

As the fifth-generation (5G) network is being deployed worldwide, more resource-intensive and latency-sensitive applications are running on end users' mobile devices. These applications generate ultrahigh network bandwidth demands, usually hundreds of gigabits per second, which imposes significant pressure on the backbone network. Suppose mobile operators attempt to address this issue by expanding the capacity of aggregation and metropolitan networks. In that case, the cost per unit of media stream transmission will increase substantially, making achieving a return on investment difficult. These applications also require end-to-end ultralow latency, which cannot be satisfied by merely relying on advancements in wireless and network's physical and transmission layer technologies. Last but not least, these applications generate massive amounts of data, leading to significant challenges in operational management. Centralized monitoring from the cloud alone cannot support such complex systems.

Hence, when considering 5G application use cases, technological complexities, and cost considerations, utilizing Mobile Edge Computing (MEC) emerges as the most effective solution for operators to overcome the abovementioned challenges. MEC provides cloud computing functionalities and an IT service environment at the network's periphery.

doi:10.6342/NTU202302400

This system enables the agile and rapid rollout of innovative applications and services tailored to clients. It is designed for implementation at cellular base stations or other edge nodes. Using MEC, it positions content, including but not limited to movies, video games, Virtual Reality (VR) / Augmented Reality (AR) applications, and other multimedia materials, nearer to the user at the fringe of the network. This approach reduces latency and lessens bandwidth demands by preventing the delay associated with application traffic that typically needs to traverse multiple internet hops to reach its endpoint. As a result, customers can fully capitalize on the benefits of 5G networks.

It is recognized that both public clouds and mobile operators bring their unique advantages to the table. Public clouds are exceptionally proficient in delivering comprehensive Platform as a Service (PaaS) capabilities. This encompasses many services, including Internet of Things (IoT), Artificial Intelligence (AI), big data, middleware, video services, and a fully developed suite of technology stacks. Mobile operators excel in their unique network resource advantages. If the two are combined, it will be an exciting development and a complementary, win-win situation. MEC is an emerging and evolving computing paradigm that can combine the strength of public clouds and mobile operators. It is expected to mitigate the challenges mentioned above, including relieving the pressure that ultra-high bandwidth requirements put on the backbone network, satisfying applications' ultra-low latency requirements, and ease of management of such a vast and complex system.

An example of MEC is Nokia's partnership with Amazon Web Services (AWS) to deploy Greengrass on MEC, combining AWS's edge computing with Nokia's proprietary mobile network solutions. By integrating AWS's Greengrass, machine learning, and Nokia's MEC and IoT smart management platform (IMPACT), they aim to seize the

doi:10.6342/NTU202302400

emerging IoT market. Another example of MEC is Verizon's partnership with Amazon, which puts AWS's WaveLength technology at the edge of the 5G network, which is closer to mobile users' devices. According to Amazon, it can deliver single-digit millisecond latency to users. Amazon also works with global partners like Vodafone, KDDI, and SK Telecom.

In a typical MEC system, mobile users can offload their applications to the system, requiring both wireless channel resources and computing resources for the applications to complete their execution. Wireless channel resources are provided by a mobile operator such as Verizon AT&T, and the mobile operator also provides computing resources by establishing a partnership with an edge computing company such as AWS. Since computing resources are limited and wireless channel resources are scarce, there can be frequent competition among mobile users over these resources. Furthermore, the mobile operator and users can behave according to their own needs/benefits only. For example, the mobile operator will provide resources to consumers only when there is potential economic profit, and the mobile users will offload their applications to the mobile operator only when the application's requirements can be satisfied. Thus, an incentive-based resource allocation scheme is needed to allocate both types of resources to mobile users effectively. Finally, it is essential to note that the effectiveness of a MEC system largely relies on the resource distribution strategy. Any shortcomings in the resource allocation process could result in diminished Quality of Service (QoS) for mobile users and escalated operational expenses for the mobile service provider.

Auction theory is outstanding in its allocation efficiency since it provides incentives for both sellers and buyers who participate in the auction, and it also guarantees mutual satisfaction for both parties while achieving an optimal allocation strategy. Auction theory

doi:10.6342/NTU202302400

has long been used for resource allocation in cloud computing. However, it is still at the beginning stage of MEC, and studying the performance issue of using auction theory on resource allocation in a MEC has attracted much attention from researchers in recent years.

Most existing works that apply auction mechanisms for resource allocation in a MEC environment only consider computing resources, where the competition and allocation of wireless channel resources are ignored, which is not practical for the real-world scenario in which wireless bandwidth is limited, and transmission delay affects the overall completion time delay for an offloaded application. Ignoring the allocation of wireless channel resources may lead to the exposure problem [1] for mobile users, which results in an inefficient allocation. The exposure problem is defined as follows: bidders intend to secure multiple items in an auction, yet they risk acquiring fewer than desired due to unexpectedly intense competition for specific items. For example, suppose there are two mobile users, A & B. A may win the auction and get the computing resources he wants. However, B may take the wireless channel resources needed for A's application offloading due to the competitiveness over limited wireless channel resources. Now, A lacks the wireless channel resources to offload his application for further computing. The computing resources he wins become useless to him or even considered a liability since he paid for them.

One salient drawback of the exposure problem lies in the potential risk of overpayment. Bidders could land in situations where they secure, and thus pay for, certain items but not the entire set that they initially desired. If the individual items' value pales compared to the entire set, bidders might end up overpaying for something that offers little value in isolation.

Adding to the complexity is how the exposure problem can convolute a bidder's

doi:10.6342/NTU202302400

strategy formulation. Bidders often have to incorporate a risk assessment mechanism into their bidding strategy to account for the probability of only partially fulfilling their desired item set. To mitigate potential exposure, they might strategically lower their bids, which can unintentionally induce inefficiencies within the auction process.

Potential bidder participation can also take a hit due to the exposure problem. High exposure risk can deter engagement from bidders who assign a higher value to the complete set of items than individual ones. The subsequent reduction in active participation can deflate competition and ultimately pull down the final bid prices.

Lastly, the exposure problem can lead to inefficiency in the allocation of resources. In scenarios where the bidders who assign the highest value to a complete set of items do not secure all the items due to exposure, an inefficient distribution of goods is likely. This inefficiency denotes that the goods do not end up with the bidders with the highest valuation.

Aside from ignoring the allocation of wireless channel resources, most existing works that apply auction mechanisms for resource allocation neglect fairness-related issues, leading to the bidder dropout problem [2], especially in a recurrent auction environment. A recurrent auction is when the same or similar goods or services are auctioned off repeatedly over time. The bidder dropout problem is defined as follows: within a recurring auction framework, applying conventional auction mechanisms, for instance, English or Vickery [3], may inadvertently lead to a persistent resource shortage for specific customers. This could prompt the impacted customers to abstain from participating in future auction rounds, reducing the long-term demand for the resource provider.

The bidder drop problem can have significant implications for auction outcomes and

doi:10.6342/NTU202302400

efficiency. In a typical auction, bidders compete by placing bids to acquire the auction item. However, in some cases, bidders may drop out of the auction for various reasons. For example, bidders drop out if they perceive the auction rules or conditions as unfavorable, believe their chances of winning are low, or face budget limitations. The bidder dropout problem can lead to suboptimal auction outcomes. When bidders drop out, the auction may result in lower competition, reduce the final market price, and result in low social welfare.

In a MEC system, where there are different types of mobile users based on their willingness and capability to pay for the offload services, they have limited patience. They might leave this MEC system either because they do not get the service requested for too long or because the Quality of Experience(QoE) did not meet their expectations. When applying auction mechanisms for resource allocation in a MEC environment, the bidder drop problem will cause mobile users to leave the MEC system. Users leaving the system will decrease the goodwill or, as is commonly known, the business reputation of the mobile operator. Furthermore, once the percentage of users left exceeds a certain threshold, the remaining users are given the power to form an oligopoly market in the auction mechanism used by the mobile operator. They can lower their bid due to the asymmetric balance of negotiation power between the mobile operator and the remaining users.

1.2 Research Objective

This research addresses the abovementioned problems, including the exposure and dropout problems of mobile users. We considered a MEC system with one mobile operator in partnership with an edge computing company, where multiple pairs of base stations and

doi:10.6342/NTU202302400

edge servers are deployed at the edge of the radio access network. Each pair of them jointly provides both types of resources to mobile users. We denote this base station and edge server pair as an Edge Node (EN). We also assumed that many mobile users need to offload their applications to this MEC system. They compete for this system's limited computing and wireless channel resources.

In this research, we use auction theory for resource allocation, where every edge node acted as a seller in the auction, and mobile users acted as buyers. The seller sells computing and wireless channel resources as bundles, and buyers bid for these resource bundles. A trusted third-party acts as the auctioneer to host this centralized auction.

First, to address the exposure problem for mobile users, a one-round auction mechanism is proposed, which jointly allocates computing and wireless channel resources from the edge nodes to mobile users while maximizing social welfare under mobile users' offloaded applications' deadline constraint. In other words, the computing and wireless resources are sold as bundles, similar to Amazon EC2 [4]. This resource allocation problem is modeled as an integer linear programming problem. Since the problem is NP-hard, a greedy approximation algorithm is proposed.

Extensive simulations were conducted, and the results are compared with two other baselines: when computing and wireless channel resources are sold separately, and results obtained by CPLEX [5], a commercial solver. We chose three evaluation metrics: social welfare, served user percentage, and execution time. Simulation results showed that the proposed algorithm significantly improved social welfare and served user percentage over the case where resources are sold separately. While the results were near-optimal compared to that of CPLEX, the execution time of the proposed algorithm was significantly

doi:10.6342/NTU202302400

lower.

Second, to address the bidder dropout problem, a fairness-aware auction mechanism based on the aforementioned one-round auction is proposed. This fair resource allocation mechanism is powerful when resources are insufficient. It can reduce the mobile user's dropout rate, increase overall fairness among all mobile users, prevent low social welfare, and ensure the long-term profitability of the mobile operator.

Extensive simulations were conducted, and we compared the results of the fairness-aware auction mechanism and those of the non-fairness-aware multi-round auction mechanism in a recurrent auction environment. We chose three evaluation metrics: user dropout rate, fairness index value, and social welfare. Simulation results showed that the proposed fairness-aware multi-round auction mechanism could prevent bidders from dropping out of the auction; in doing so, the bidder dropout problem can be averted, as well as the uneven distribution of negotiating power between the mobile operator and the remaining mobile users, which would otherwise reduce social welfare. This approach also prevents the potential reduction of the mobile operator's long-term profits due to an oligopolistic market condition.

1.3 Thesis Organization

The remainder of this study is structured as follows: Section 2 reviews various auction mechanisms and related works in auction-based resource allocation. Section 3 details the proposed auction mechanism and network models and frames the resource allocation problem alongside its resolution techniques. Section 4 presents a fairness-oriented auction mechanism. Section 5 elucidates the dataset utilized for simulations, the execution of

doi:10.6342/NTU202302400

these simulations, and the evaluation of results generated by our proposal. Finally, Section 6 concludes the research and outlines potential areas for future exploration.

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

Chapter 2 Related Work

2.1 Auction Mechanism

Within the framework of MEC, traditional static methods of resource allocation fall short in optimizing the use of edge resources, particularly when edge services require adaptive, on-demand, and real-time provisioning to users. Conversely, strategies rooted in economics and business management are considerably more effective, given their dynamic nature and efficiency in resource allocation [6].

Auction theory [7], a well-established economic approach, has seen widespread application in Cloud Computing and presents itself as a viable solution for resource allocation in MEC. This is primarily because it allows for a reasonable and efficient distribution of limited provider resources to customers through a competitive pricing structure in a trade-like scenario. Furthermore, it can encapsulate the bidirectional interactions between resource providers and customers [8].

A well-designed auction mechanism strikes a good balance among performance in terms of social welfare, dominant-strategy incentive-compatible for bidders, and computational efficiency [9].

Incentive compatibility is a crucial concept in the arena of auction theory, playing a

doi:10.6342/NTU202302400

significant role in shaping bidder behavior and the ultimate outcomes of auctions. An auction is deemed incentive-compatible (IC) when each participant's optimal strategy aligns with bidding their true valuation for the auctioned goods or services. The strength of an IC auction lies in its propensity to promote truthful bidding. This eliminates strategic complexity for bidders and substantially enhances the efficiency of the auction mechanism. Moreover, incentive compatibility can foster a trustworthy and predictable auction environment, boosting credibility and prompting increased bidder participation. Despite these advantages, implementing an incentive-compatible design is challenging. For instance, achieving incentive compatibility might be challenging in complex auction scenarios with multiple heterogeneous items.

As in single-item environments, First-Price Auction [10], also referred to as a First Price Sealed Bid Auction (FPSBA), is a type of auction mechanism where bidders simultaneously submit sealed bids, unaware of the bids offered by other competing bidders. The essential characteristic of this mechanism is that the highest bidder secures the item for sale and pays the exact price of his or her bid. In contrast to other auction formats, such as the Second Price Auction [3] where the highest bidder pays the amount of the second-highest bid, the First Price Auction inherently invites strategic bidding. Bidders are incentivized to shade their bids, bidding less than their actual valuation, to avoid the "winner's curse," in which the winner overpays for the item. The First Price Auction format is widely used in various contexts, from online advertising to government procurement contracts. Its strategic nature prompts bidders to consider their valuation, potential competition, and risk tolerance when determining their bids. Notably, in a First Price Auction, no dominant bidding strategy exists, hence making the process dynamic and context-dependent. Nevertheless, this format may not always lead to efficient outcomes, where the item goes

doi:10.6342/NTU202302400

to the bidder with the highest valuation. This potential inefficiency is a trade-off for the simplicity and straightforwardness of the First Price Auction structure.

In contrast to the First-Price Auction, Vickrey auction [3], a Sealed Second-Price Auction, has been proven that it can maximize social welfare while guaranteeing bidders' dominant strategy is incentive-compatible [11]. The auction procedure runs in polynomial time. In a Vickrey auction, each participant submits a bid without knowledge of the bids proposed by other participants, thereby maintaining the sealed nature of the bids. Once all bids are submitted, they are opened simultaneously. The participant who has made the highest bid wins, yet the price they pay corresponds to the second-highest bid. This unique approach to pricing insulates bidders from the risk of the ”winner's curse,” which is the possibility of overpaying. A crucial feature of the Vickrey auction is that it is incentive compatible, providing a dominant strategy for bidders to bid their true valuation. This mechanism encourages truthful bidding, as the final payment does not depend on the bidder's bid but rather the bid of others. Regardless of the other bids, a bidder can maximize their payoff by bidding their true valuation, leading to an efficient outcome. Despite these advantages, Vickrey auctions also present specific challenges, including vulnerability to collusion and a general lack of familiarity and intuitive understanding among participants compared to other auction formats.

As in multi-item environments, the same goal for maximizing social welfare and guaranteeing bidders' dominant strategy is incentive-compatible can be achieved by Vickrey-Clarke-Groves (VCG) auction [12] [13], a combinatorial auction. The Vickrey-Clarke-Groves (VCG) auction is an influential mechanism within economics and game theory, known for its robustness in achieving efficiency and incentive compatibility in multi-item auctions. It is a generalization of the Vickrey auction to situations involving allocating

doi:10.6342/NTU202302400

multiple heterogeneous items. The VCG auction operates on the principle of social optimality, aiming to allocate goods to maximize total societal value. Each bidder submits a bid for different combinations of goods, elucidating their valuations. The auctioneer then assigns the goods to maximize the sum of these stated valuations. The payment rule in a VCG auction is unique. Although the items are assigned to maximize total value, the price paid by each winning bidder is based on the external cost their winning imposes on other bidders. Specifically, a winning bidder pays an amount equal to the reduction in total value to other bidders caused by their winning the items rather than paying their bid. This mechanism discourages participants from misrepresenting their true valuations, thereby achieving incentive compatibility. The VCG auction's principal advantages are its ability to lead to an efficient outcome (where goods are allocated to those who value them most) and its incentive compatibility. However, while the VCG auction mechanism is excellent in theory and guarantees truthful bidding, it has drawbacks such as low or even zero revenue for sellers, making it impractical in the real world.

On the other hand, Generalized Second Price (GSP) auction mechanism [14] [15] does not guarantee truthful bidding, but it guarantees a locally-envy free equilibrium. It is used mainly in the context of keyword auctions [16], where sponsored search slots are sold on an auction basis. The Generalized Second Price (GSP) auction is an instrumental mechanism that has been extensively adopted in search engine advertising. This auction model, which is non-truthful, underpins the functionality of widely-used advertising platforms, including Google AdWords.In a GSP auction, advertisers submit bids for keywords. When a user executes a search that includes one of these keywords, the bidding advertisers are ranked according to their bid amounts. The highest bidder is awarded the most prominent advertisement position, generally at the top of the search results page; the

doi:10.6342/NTU202302400

second-highest bidder receives the next most visible position, and so on. The number of advertisement slots is typically limited; hence not all bidders may receive an advertisement position. The payment structure in GSP auctions is distinct. Although the advertisement slots are assigned based on the bid amounts, the cost per click each advertiser pays if their ad is clicked is determined by the advertiser's bid ranked directly below them. In other words, the highest-ranking advertiser pays the bid amount of the second-highest bidder per click; the second-highest pays the bid amount of the third-highest, and so forth. GSP auctions effectively strike a balance between the efficiency of allocating advertisement slots to those advertisers who value them the most and revenue generation for the search engine platform. However, the strategic complexity of bidding in a GSP auction, where truthful bidding is not always the best strategy, remains a challenge for advertisers. Despite this, the widespread adoption of GSP auctions in digital advertising signifies their practical utility and effectiveness in the rapidly evolving online marketplace.

2.2 Auction-based Resource Allocation and Pricing Mechanism

So far, the application of auction theory in edge cloud environments has attracted significant attention in recent years.

Gao et al.[17] model the computing resource allocation problem in distributed edge cloud as an n-to-one weighted bipartite graph matching problem, with the computing resource capacity of edge cloud and task deadline as constraints, the utility of provider is maximized. In the proposed system, edge cloud nodes act as resource providers and mobile users as consumers. A one-round Vickrey auction-based mechanism, alongside a

doi:10.6342/NTU202302400

greedy approximation algorithm, is proposed to solve the allocation and payment determination problem between providers and consumers, which obtains near-optimal social welfare while satisfying multiple properties of the auction mechanism, e.g., truthful, individual rationality, and computational efficiency.

Similar to [17], Liu et al. [18] model the computing resource allocation problem in distributed edge cloud as an n-to-one weighted bipartite graph matching problem, with the computing resource capacity of edge cloud, task deadline and energy consumption as constraints, the utility of provider is maximized. A deep learning-based one-round optimal auction is proposed to solve the allocation and payment determination problem, which can effectively improve the utility of the edge provider while preventing mobile users from making unrealistic bids and unfairly influencing the allocation.

Bahreini et al. [19] model the computing resources allocation problem in Mobile Edge Computing (MEC) systems where users have heterogeneous demands as a Mixed-Integer Linear Program (MILP). MILP is solved by a General Second Price (GSP) [14] based combinatorial auction mechanism with a greedy approximation algorithm, while social welfare is maximized. The auction mechanism proposed can obtain near-optimal social welfare compared to that obtained by CPLEX [5]. However, while individual rationality and computational efficiency are ensured, the property of truthfulness cannot be guaranteed.

Wang et al. [20] addressed the incentive profit maximization of MEC service providers in both non-competitive and competitive scenarios using market pricing and auction models. For the non-competitive case, a non-linear optimization problem is formed, where the utility of providers is maximized, with the Quality of Experience (QoE) of mobile users

doi:10.6342/NTU202302400

as constraints. By using the KKT conditions, the optimal price of computing resources is found. The competitive case uses the results of the non-competitive case as a baseline. Since computing resource is limited in this case, a profit maximization problem is formed, where the utility of providers is maximized, with the capacity of computing resource as a constraint. The optimal price of computing resources is solved by a Vickrey auction-based online profit maximization multi-round auction (PMMRA) mechanism alongside a greedy approximation algorithm. While the utility of providers is maximized, properties of auction mechanisms such as truthfulness, individual rationality, and computational efficiency are also satisfied.

Zhang et al. [21] studied the matching problem between the MEC service providers and user equipment (UE) in a multi-MEC and multi-UE scenario. While social welfare is maximized, with both computing and wireless resources considered, a 2-dimensional 0-1 knapsack problem is formed and solved by two separate Vickrey auction-based mechanisms alongside Dynamic Programming, in which the computing and wireless resources are sold jointly. Furthermore, the auction mechanism proposed is extended to a multi-round auction. Although the proposed resource allocation and pricing problem guarantees essential properties such as truthfulness, individual rationality, and computational efficiency, it must be noted that the implementation of Dynamic Programming (DP) as a solution strategy results in pseudo-polynomial time complexity. This temporal complexity, inherent in the application of DP, often renders it impracticable in real-world scenarios where computational resources and time are limited. Consequently, the expeditious allocation and pricing, typically required in dynamic and competitive markets, may be hindered, underscoring a need for the exploration of more efficient computational approaches to effectively manage this problem in practical contexts.

doi:10.6342/NTU202302400

As evidenced in Table 2.1, to the best of our knowledge, no extant study has proposed a fairness-aware auction mechanism that simultaneously sells computing and wireless resources as an integrated package to mobile users in MEC systems. A gap exists in the literature for an auction mechanism that jointly allocates these resources and underpins fairness in recurrent auction environments. A recurrent auction, characterized by repeated bidding rounds over time, aptly mirrors real-world scenarios wherein resources are persistently traded in dynamic markets. Additionally, such a mechanism would need to perform these allocations with an expedient and pragmatic speed. This feature is essential given the dynamic and real-time nature of resource demand in MEC systems. This underscores a critical research gap that warrants rigorous scholarly exploration, aiming to develop an auction framework that encapsulates fairness, joint resource allocation, and practical execution speed in the context of MEC systems.

doi:10.6342/NTU202302400

Ref.	Scenarios[1]	Constraints[2]	Commodity[3]	Solving Technique[4]	Package[5]	Fairness[6]
[17]	EC	CR, D	CR	Greedy	-	N
[18]	EC	CR, EL, D	CR	Deep Learning	-	N
[19]	EC	CRs	CRs	Greedy	Y	N
[20]	MEC	CR	CR	Greedy	-	N
[21]	MEC	CR, WR	CR, WR	DP	N	N
This Research	MEC	CR, WR	CR, WR	Greedy	Y	Y

[1] Scenarios is the context of resource allocation, which includes edge cloud (EC) and Mobile Edge Computing (MEC).

[2] Constraints of the corresponding resource allocation problem, including the capacity of computing resource(s) (CR(s)), energy consumption limit (E), and the deadline of offloaded tasks (D).

[3] Commodity is what buyers bid for in an auction. Multiple types of resources can be jointly considered, including computing resources (CR) and wireless resources (WR). Examples of computing resources include Virtual Machine (VM), CPU cycle, memory, and disk, and examples of wireless resources include resource block (RB) and sub-channels.

[4] Solving Technique is the solution to the proposed problem. Greedy represents the greedy approximation algorithm, while DP represents Dynamic Programming.

[5] Package denotes whether the commodities, if there exists multiple of them, are sold as a bundle (Y) or sold separately (N).

[6] Whether fairness is considered in the auction mechanism, Y means yes, and N means no.

Table 2.1: Comparison of Researches on Auction-based Resource Allocation

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

Chapter 3 Auction-based Resource Allocation and Pricing Model

3.1 System Model

For the following description of the proposed system model, definitions of notations are in Appendix 6.2. The illustration diagram of the proposed mobile edge computing system is in Figure 3.1.

Consider a Mobile Edge Computing (MEC) environment in an urban area, where one mobile operator partners with an edge computing provider to provide computing and wireless resources to mobile users. The mobile operator is assumed to have N base stations deployed in this network, in which N is a fixed positive integer. Each base station is paired with an edge server which forms an edge node, and the edge computing provider provides the edge servers. Each base station is used to serve as a microcell. The circle in Figure 3.1 is the microcell's serving area. Mobile users are assumed to be with low moving speeds. Thus the handover issue is not considered. The edge server provides computing resources in the form of central processing unit (CPU) cycle. The base station provides

doi:10.6342/NTU202302400

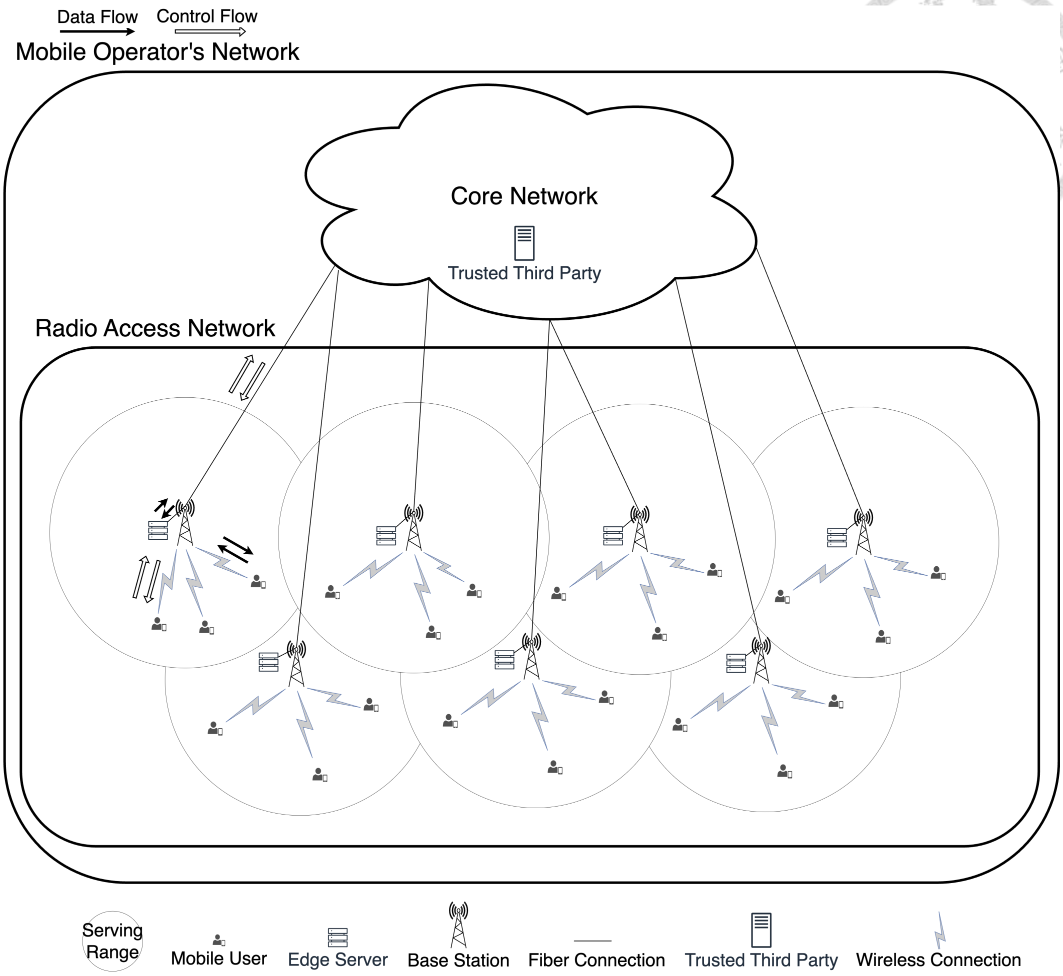

Figure 3.1: A illustration diagram of the proposed mobile edge computing system

wireless resources in the form of resource block (RB) to connected mobile users. We assume that the orthogonal frequency division multiple access (OFDMA) [1] is used for uplink and downlink transmission, and each resource block consists of N_{RB} sub-carriers. Following the 5G standards, N_{RB} is set to 12 in this study.

The system is assumed to have M mobile users, where M is a fixed positive integer. The mobile users connect with the base stations through wireless connections, and the base station connects with the edge server through a high-speed ethernet-like fiber connection.

Since both types of resources (wireless & edge computing) are limited, the mobile

[1] Other than OFDMA, Non-Orthogonal Multiple Access (NOMA) in 5G or Beyond 5G environments can be considered for further research. When designing an auction model for resource allocation in MEC, especially wireless resources, the characteristics of NOMA need to be considered.

doi:10.6342/NTU202302400

operator assigns a trusted third party in the network to serve as an auctioneer, who uses an auction-based mechanism to determine the allocation and pricing for these resources, fulfilling users' demands dynamically. The auctioneer connects with edge nodes through ethernet-like fiber connections. The auctioneer is in charge of the auction and is responsible for processing the control signals flow. Figure 3.1 presents the control flow as a hollow arrow. The solid arrow in Figure 3.1 is the data flow, which represents the flow of application tasks that mobile users offloaded and their execution results.

Inspired by Amazon's Elastic Compute Cloud (EC2) spot pricing [4], computing and wireless resources in this study are sold together as resource packages. Similar to [17] [19] [20] [21], the system structure of our proposed model is described down below.

This system has N edge nodes; each edge node is a tuple of a base station and an edge server. We denoted edge node i as

$$E_i = (S_i, W_i), i = 1, 2, \cdots, N \tag{3.1}$$

where

$$S_i = (s_i^1, s_i^2, \cdots, s_i^k) \tag{3.2}$$

$$s_i^z = (c_z, w_z), \forall z \in (1, 2, \cdots, k) \tag{3.3}$$

$$Q_i = (q_i^1, q_i^2, \cdots, q_i^k) \tag{3.4}$$

$$\tag{3.5}$$

For each edge node i, $E_i = (S_i, W_i)$ in (3.1) is a fixed tuple, where S_i in (3.2) is the set of resource packages, each of which is defined to be a pair of CPU frequency and

doi:10.6342/NTU202302400

the number of resource blocks this resource package provides. We denote W_i as the total number of resource blocks of edge node i.

There is a total of k types of unique resource packages. For a type-z resource package, denoted as $s_i^z = (c_z, w_z)$ in (3.3), c_z is the CPU frequency this resource package offers and w_z is the number of resource blocks it provides. Q_i in (3.4) is the set of resource package's capacity, where q_i^z represents the total number of s_i^z edge node i has.

Since the total amount of wireless resources allocated to mobile users cannot exceed the edge node's total capacity of the wireless resource, denote W_i as the total number of wireless resource blocks of edge node i, we have

$$\sum_{z=1}^{k} w_z = W_i, i = 1, 2, \cdots, N \tag{3.6}$$

Assume that M mobile users compete for computing and wireless resources in this system. Each mobile user act as a buyer in the auction, and each buyer will submit a bid request to the auctioneer, requesting resource packages from edge nodes. Only if a user is within an edge node's serving area can this user submit a bid request for this edge node's resource package.

Different users operate under diverse tariff plans, introducing a level of heterogeneity that accurately reflects the variability found in real-world settings. In this scenario, a user's bid request is not manually generated but is the product of an automated auction agent. This agent removes the burden of active participation from the user, streamlining the auction process and enhancing its speed and efficiency.

The automated auction acts as a compact operating system within a user's mobile

doi:10.6342/NTU202302400

device, and it is activated whenever an application task requires offloading. Under such circumstances, this agent is responsible for generating a bid request. With access to the user's tariff plan details, the automated auction agent generates a quota-like bid under the user's assigned plan. Each tariff plan corresponds to a specific range, within which the automated auction agent randomly generates bid values. This bid generation is governed by a pre-determined probability distribution, introducing a dimension of randomness into the bidding process yet remaining within the constraints defined by the user's specific tariff plan. This stochastic approach cultivates a range of bid values, fostering a dynamic and unpredictable auction environment.

In the context of this study, we assume that the pre-determined probability distribution adheres to a uniform distribution. However, the model's flexibility allows for replacing this distribution in future studies. Other distributions, such as exponential or Gaussian, can broaden the model's applicability and adaptability to different bidding behaviors and auction scenarios.

For buyer j, its bid request r_j is denoted as

$$r_j = (b_j; g_j, f_j, s_j, t_j^d), j = 1, 2, \cdots, M \tag{3.7}$$

where b_j is the value of its bid, g_j is the type number of resource package it is bidding for, f_j is the CPU cycle required for the offloaded task to complete, s_j is the size of the offloaded task, and t_j^d is the deadline of the offloaded task. It is assumed that f_j, s_j, t_j^d are generated following a uniform distribution.

Leveraging the variables f_j, s_j, and t_j^d, the automated auction agent selects a resource

doi:10.6342/NTU202302400

package that appears most conducive to the timely completion of the offloaded applications. In other words, the agent targets those resource packages that ensure the total completion delay remains less than t_j^d. The automated auction agent employs a random selection strategy in scenarios where multiple resource packages meet the requisite criteria. The type of chosen resource package is denoted as g_j.

Once the type of resource package has been selected, the automated auction agent proceeds to generate its bid. This bid generation process is conducted following (3.8).

$$b_j = u_j \cdot \frac{f_j}{c_{g_j}} \cdot w_{g_j}, j = 1, 2, \cdots, M \tag{3.8}$$

where u_j is the unit bid of user j, and its generation follows uniform distribution.

Definition 3.1.1 (Social Welfare)**.** *The social welfare V is the sum of all winning bids, i.e., $V = \sum_{i=1}^{N} \sum_{j=1}^{M} x_{i,j} b_j$.*

Given the definition mentioned above of social welfare, represented as V, we establish our objective function as V.

3.2 Network Model

Total bandwidth allocated from edge node i to mobile user j can be denoted as follows:

$$w_{i,j} = w_{g_j} \cdot N_{RB} \cdot \Delta f \tag{3.9}$$

doi:10.6342/NTU202302400

where N_{RB} is the number of sub-carriers each resource block consists of, and Δf is the sub-carrier spacing.

This research uses the path loss model [22] to describe the wireless channel between mobile users and edge nodes while neglecting fading and shadowing effects. Based on this assumption, we can get the data rate $a_{i,j}$ between edge node i and mobile user j according to Shannon Capacity Formula [23] as follows:

$$a_{i,j} = w_{i,j} \cdot \log_2(1 + \frac{P_j h_{i,j}}{N_0}) \tag{3.10}$$

where P_j is the transmission power of user j's mobile device, and $h_{i,j} = d_{i,j}^{-H}$ [22] represents the channel gain between mobile user j and edge node i, with $d_{i,j}$ denotes the distance between the two, and H denotes the channel gain constant. Finally, N_0 is the background noise power.

The completion delay t_j^T of an offloaded task includes transmission $t_{i,j}^t$, propagation $t_{i,j}^p$, and execution delay $t_{i,j}^e$. Denote o_w as the speed of light, and $d_{i,j}$ as the distance between mobile user j and edge node i, the completion delay t_j^T for task offloaded is

$$t_j^T = t_{i,j}^t + t_{i,j}^p + t_{i,j}^e \tag{3.11}$$

$$t_{i,j}^t = \frac{s_j}{a_{i,j}} \tag{3.12}$$

$$t_{i,j}^p = \frac{d_{i,j}}{o_w} \tag{3.13}$$

$$t_{i,j}^e = \frac{f_j}{c_{g_j}} \tag{3.14}$$

doi:10.6342/NTU202302400

Since the size of the computing result is usually relatively small, we ignore downlink transmission and propagation delay.

3.3 Procedure of a Centralized Single Round Auction

We abstract the system mentioned above (Section 3.1) into an auction model for further study. One base station and edge server form a pair and act as one edge node. Edge nodes provide both types of resources. One mobile user acted as one bidder, and a trusted third party within the mobile operator network was the auctioneer hosting this centralized auction. The auctioneer takes bid requests from mobile users, collects edge nodes' resource availability state information, broadcasts information to mobile users, runs a resource allocation algorithm, and delivers allocation outcomes to mobile users and edge nodes.

As illustrated in Figure 3.2, we first focus on a single-round, centralized auction procedure.

Every auction round allocates available resource packages from every edge node to its bidders. A resource package is defined as a pair of CPU frequency and the number of resource blocks this resource package provides. One mobile user can only bid on one resource package at a time, i.e., one bidder can only get resources from at most one edge node, while an edge node can serve many mobile users simultaneously.

doi:10.6342/NTU202302400

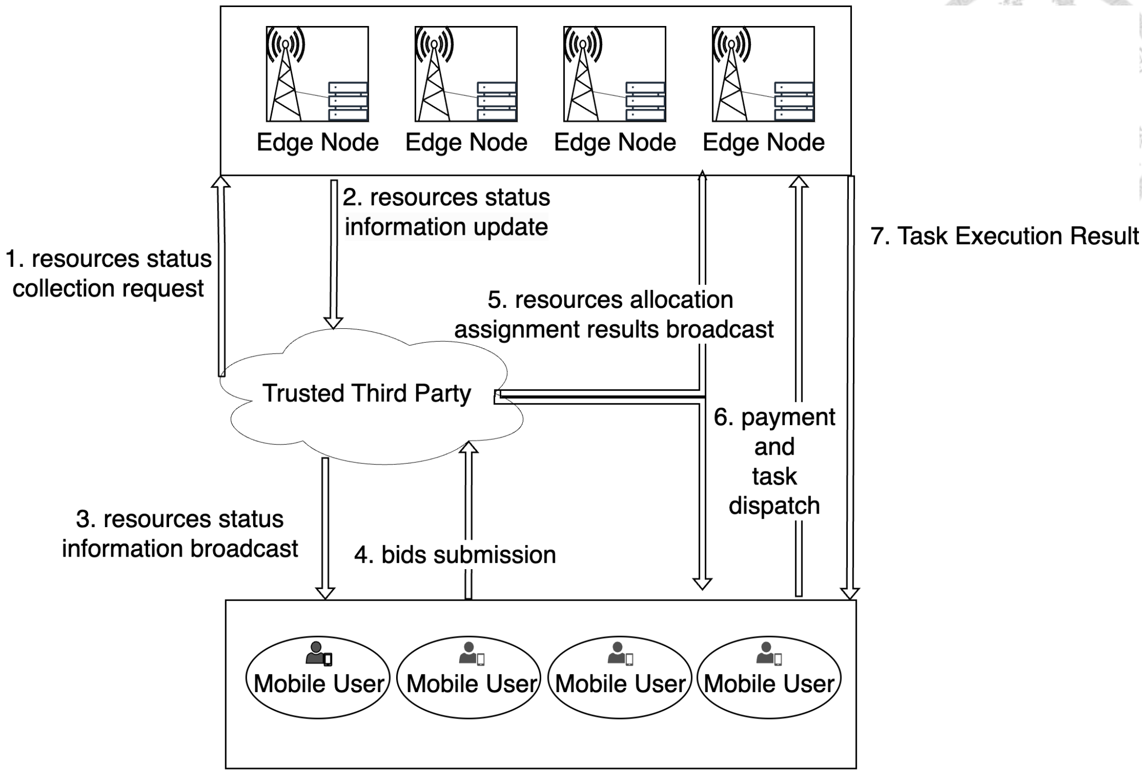

Figure 3.2: Auction Flow chart of a single round auction

In a single-round auction, the edge nodes and mobile users are viewed as sellers and buyers of computing and wireless resource, respectively. The following are more thorough processes for a single round of auction :

Step 1: The auctioneer requests resource status information from edge nodes.

Step 2: Resource availability status state information of edge nodes is sent back to the auctioneer and updated, which includes the number of available resource packages.

Step 3: Resource availability state information of each edge node is broadcasted to their connected mobile users, respectively.

Step 4: Mobile users decide which resource package they want to bid for based on the information they receive. Then they submit their bid requests, which include their bid value, type number of the bidding resource package, and the characteristic of the

doi:10.6342/NTU202302400

offloaded task, to the auctioneer.

Step 5: The auctioneer runs a single round of auction to determine resource allocation and pricing using a pre-determined resource allocation algorithm, and this algorithm is replaceable. Finally, the auctioneer returns the results to the mobile users and edge nodes.

Step 6: Mobile users submit their corresponding payments to the edge nodes and then upload their application task to edge nodes. Finally, edge nodes start executing application tasks.

Step 7: The task execution results are sent back to the mobile users.

From a single mobile user's perspective, the following are detailed processes for a single round of auction:

Step 1: The resource availability state information of edge nodes whose serving range covers mobile user i is sent to mobile user i by the auctioneer.

Step 2: Mobile user i decides which resource package it wants to bid for based on the information it receives. Then it submits its bid requests, which include their bid value, type number of the bidding resource package, and the characteristic of the offloaded task, to the auctioneer.

Step 3: The auction result is returned to mobile user i by the auctioneer.

Step 4: If mobile user i wins the resource package it bids for in the auction, it submits its payment and uploads its application task to the corresponding edge node.

Step 5: The task execution results are returned to mobile user i.

doi:10.6342/NTU202302400

3.4 Mobile Edge Computing Resource Allocation and Pricing Problem

Considering the auction model mentioned above, two problems need to be solved, the first one is the resource allocation problem, and the other is the payment determination problem. This study aims to maximize social welfare, as defined in Definition 3.1.1.

We choose to maximize social welfare for multiple reasons. First, social welfare constitutes a fundamental objective function relevant to numerous real-world scenarios. For example, welfare maximization is positioned as the principal objective in government-held auctions that sell wireless spectrum. While revenue generation carries significance, it is typically subordinate to welfare maximization. In addition, sellers are commonly advised to prioritize welfare maximization within the framework of competitive markets. The reasoning is that neglecting to do so may create an opportunity for competitors to fill that gap, potentially losing customers. Third, economic efficiency stands as a fundamental component in this discourse. Social welfare maximization ensures that the resources are allocated to those who value them the most, and this is an efficient outcome from an economic perspective. Finally, If the auction is designed to maximize social welfare, it can incentivize participants to bid their true valuations, leading to more accurate price discovery. Based on the reasons mentioned above, in this study, we choose to maximize social welfare.

Based on the definition of social welfare V, the Mobile Edge Computing Resource Allocation and Pricing Problem can be formulated as an integer linear program (ILP) called MECRAPP-ILP in the rest of this study.

doi:10.6342/NTU202302400

MECRAPP-ILP:

$$\max_{x_{i,j}} V = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{i,j} b_j \tag{3.15}$$

$$\text{subject to: } \sum_{j=1}^{M} g_{j,z} \cdot x_{i,j} \leq q_i^z, \forall z \in K, \forall i \in N \tag{3.16}$$

$$t_j^T x_{i,j} \leq t_j^d, \forall j \in M \tag{3.17}$$

$$x_{i,j} h_{i,j} \leq D_i, \forall i, \forall j \tag{3.18}$$

$$\sum_{i=1}^{N} x_{i,j} \leq 1, \forall j \in M \tag{3.19}$$

$$x_{i,j} \in \{0, 1\}, \forall i \in N, \forall j \in M \tag{3.20}$$

where $x_{i,j}$ is binary decision variable, it is equal to 1 if mobile user j wins the auction and offload its task to edge node i, and is equal to 0 if he lose. $g_{j,z}$ is a binary variable; it is equal to 1 if user j decides to bid for type z resource package; otherwise, it is equal to zero. (3.16) ensures that a resource package can only be allocated to one mobile user. (3.17) guarantees that the offloaded task can be completed before its deadline. (3.18) make sure that every mobile user can only get resources from its serving base station, where $h_{i,j}$ is the distance between user j and edge node i, D_i is the serving radius of edge node i. (3.19) guarantees that each mobile user can only offload a task to, at most, one edge node. (3.20) guarantees the integrality of the decision variables.

Since the MECRAPP-ILP is a knapsack-liked problem, in the sense that each resource package sold by edge node is a 0-1 knapsack problem. It is well known that the 0-1 knapsack problem is NP-Complete, thus making the MECRAPP-ILP NP-hard.

Greedy approximation algorithms are commonly used for solving NP-hard problems

doi:10.6342/NTU202302400

because they provide a trade-off between efficiency and optimality. It is well known that NP-hard problems are a class of computational problems for which no known efficient algorithm exists to find an optimal solution in polynomial time.

When faced with NP-hard problems, finding an optimal solution within a reasonable time becomes impractical or impossible. In such cases, approximation algorithms often come into play. Approximation algorithms usually aim to find solutions that are close to optimal, although not necessarily the absolute best.

Greedy approximation algorithms are desirable because they are often simple to implement and have low time complexity. They make decisions based on the local optimal choice, allowing for efficient computation. While these algorithms do not guarantee to find the globally optimal solution, they can provide reasonably good solutions that are close to the optimal solution.

By using greedy approximation algorithms, one can strike a balance between computational feasibility and obtaining solutions that are close to optimal. While the solutions may not be perfect, they are often sufficient in practice and can provide valuable insights or approximate solutions to complex problems.

Thus, in this study, to solve MECRAPP-ILP, an NP-hard problem, a greedy heuristic algorithm is proposed in section 3.5.

doi:10.6342/NTU202302400

3.5 A Greedy Heuristic Algorithm for Resource Allocation

Since MECRAPP-ILP is NP-hard, we designed a greedy heuristic algorithm called Algorithm 1 to solve this problem in reasonable polynomial time. The pseudo-code for Algorithm 1 is illustrated in Figure 3.3. First, the auctioneer gets requests sent by mobile users and then deletes requests whose tasks' deadlines cannot be satisfied, i.e., for those users with $t_j^d > t_j^T$. After deletion, the remaining requests are sorted based on bid heuristic B_j in non-increasing order. The bid heuristic is defined as

$$B_j = \frac{b_j}{t_{i,j}^e w_{i,j}}, \forall j \in M \tag{3.21}$$

where $t_{i,j}^c$ represents the execution time needed. B_j represents the bid value per unit resource for mobile user j. Once sorting is done, the auctioneer starts allocating resources of edge nodes to its connected mobile users by processing the sorted requests list from the start until there are no available resources left or all requests are processed. In scenarios where a mobile user is located within the service areas of multiple edge nodes and requests a specific resource package, denoted as k, the selection process becomes pivotal. We assign the user to the edge node that has the most ample supply of the requested resource package k. This ensures that the user is served by an edge node that is not only within their proximity but also capable of meeting their requirements most effectively. Finally, the payment and social welfare are calculated based on the allocation results.

doi:10.6342/NTU202302400

Require: requests R($r_1, r_2, \cdots, r_M$), resources' capacity Q($Q_1, Q_2, \cdots, Q_N$)
Ensure: Allocation Matrix X, Social Welfare V, Payment Matrix P

$X \leftarrow 0$
$V \leftarrow 0$
$P \leftarrow 0$
Delete request r_j in R if $t_j^d > t_j^T, \forall j \in 1, 2, \cdots, M$
$B_j \leftarrow \frac{b_j}{t_{i,j}^e w_{i,j}}, \forall j \in 1, 2, \cdots, M$
Sort R in non-increasing order of B_j
$i \leftarrow 1$
while $Q \neq 0$ and $i \leq M$ **do**
    pop r_j from the head of sorted requests vector
    find the edge node in range for mobile user j that has the most g_j resource packages left, denoted this edge node as i
    if resource package g_j's capacity of edge node i is greater than zero **then**
        $V \leftarrow V + b_i$
        $q_i^{g_j} \leftarrow q_i^{g_j} - 1$
        $X_{i,j} \leftarrow 1$
        Write the winning bid price, denoted as b_j, for the resource package g_j in edge node i, to the corresponding entry in P
    else
        update the $k + 1$ highest price for resource package g_j in edge node i using b_j, write the result to P. k is the current number of resource package g_j that are occupied
    end if
    $i \leftarrow i + 1$
end while

Figure 3.3: Algorithm 1 - The pseudo-code of resource allocation in a single round of auction mechanism

doi:10.6342/NTU202302400

Definition 3.5.1 (Computational Efficiency). *An auction mechanism satisfies computational efficiency if its allocation rule can be computed in polynomial time efficiently.*

Theorem 3.5.1. *Algorithm 1 is computationally efficient*

Proof. In Algorithm 1, the request sorting takes $O(M \log M)$. The while loop, which starts at line 8, will iterate at most M times, and the execution procedure to find the base station with the most desired resource packages left (line 10) is an $O(NK)$ operation. Therefore the time complexity of Algorithm 1 is $O(M \log M + MNK)$. We assume the ratio of the number of serving edge nodes and the number of mobile users as a constant α, i.e., $\alpha = \frac{N}{M}$, and assume that the total number of unique resources packages K is constant. With these assumptions, the time complexity of Algorithm 1 can be simplified as $O(M \log M + MNK) = O(M \log M + M^2K) = O(M^2)$, which is polynomial. □

3.6 Payment Rule

After the resource allocation problem is solved, each winning mobile user's payment in the auction must be determined. We assume that mobile users have quasi-linear utilities (i.e., $u_i = v_i - p_i$), and all of them are rational, where u_i is the utility of user i, v_i is the true valuation user has toward the item he bids, and p_i is the price he has to pay if he wins the item. Here we assume that they are rational, meaning their goal is maximizing their utility.

The payment rule will determine whether the auction mechanism can satisfy some desired properties, such as Truthfulness, Individual Rationality, which are defined in the following:

doi:10.6342/NTU202302400

Definition 3.6.1 (Truthfulness). *An auction mechanism satisfies truthfulness if all agents bid truthfully, i.e.,* $b_i = v_i$ *is a dominant strategy for each bidder:* $x_i v_i - p_i$ *is maximized by bidding their true valuation. Where for bidder* i*,* b_i *is the bid value,* v_i *is the true valuation,* x_i *is* 1 *when bidder* i *wins, and is set to* 0 *vice versa, and* p_i *is the payment that bidder* i *has to pay if he wins.*

Definition 3.6.2 (Individual Rationality). *An auction mechanism satisfies individual rationality if for every agent* i*:* $p_i \leq v_i$*, i.e., nobody is ever asked to pay more than their true valuation for the outcome. For bidder* i*,* b_i *is the bid value,* v_i *is the true valuation, and* p_i *is the payment that bidder* i *has to pay if he wins.*

There are a total of q_i^k of the type-k resource package in edge node i (3.4). It is assumed that a total of $x, x > q_i^k$ mobile users are attending the auction for this type-k resource package in edge node i. After the resource allocation is determined, q_i^k mobile users are allocated with a type-k resource package. The necessary and sufficient condition for a mobile user j to be allocated with type-k resource package is that its bid b_j is higher than the critical bid $b_{i,k}^*$ of this type-k resource package, which is $q_i^k + 1$th highest bid among all x mobile users.

Denote $p_{i,j,k}$ as the payment that mobile user j has to pay for resource package k in edge node i. The payment rule follows Vickrey Philosophy [3], and is given by

$$p_{i,j,k} = b_{i,k}^* \tag{3.22}$$

The proof of this payment rule is given in Appendix 6.2.

Definition 3.6.3 (Monotone Allocation Rule). *An allocation rule x for a single-parameter*

doi:10.6342/NTU202302400

environment is monotone if for every bidder i and bids b_{-i} by the other bidders, the allocation $x_i(z, b_{-i})$ to i is non-decreasing in its bid z.

Theorem 3.6.1. *The proposed auction mechanism is truthful*

Proof. In Algorithm 1, the requests R are sorted in non-increasing order of B_j; thus, for an arbitrary request $r_i \in R$, if the bid b_j wins, then a larger bid $b_j^{'} \geq b_j$ must win, which means that this allocation rule follows Definition 3.6.3. According to Myerson's Lemma [11], if an auction's allocation rule is monotone, then there exists a unique payment rule (3.22), and the auction mechanism that uses this allocation rule and payment rule is truthful. □

Theorem 3.6.2. *The proposed auction mechanism is individual-rational*

Proof. By Theorem 3.6.1, the proposed auction mechanism is truthful, i.e., $v_i = b_i$ for each bidder $i, i \in M$, and second-price payment is adopted, plus the greedy heuristic allocation rule. Thus, for each bidder i, its payment p_i must be smaller than b_i, thus, $p_i \leq b_i \leq v_i \Rightarrow p_i \leq v_i$. □

In summary, the proposed auction mechanism and its payment rule incentivize bidders to bid truthfully and guarantee individual rationality.

3.7 Concluding Remark

This chapter proposes a single-round auction-based resource allocation and pricing mechanism. While Zhang et al. [21] also jointly considered the allocation of computing and wireless resources, the payment rule they proposed forces mobile users to attend two separate auctions, one for computing resources and one for wireless resources. This

doi:10.6342/NTU202302400

can convolute a mobile user's bidding strategy and put them at the potential risk of overpayment. On the contrary, our proposed mechanism can efficiently and jointly allocate computing and wireless resources as resource packages from edge nodes to mobile users; this alleviates mobile users from the need to develop complex bidding strategies and the risk of overpayment. The proposed auction mechanism solves the exposure problem for mobile users, ensuring a user will either get both types of resources or none while maximizing social welfare.

Second, Myerson's Lemma [11] can not be directly applied to an auction where the item being auctioned has a capacity over one, and Zhang et al. [21] did not provide proof to support the payment rule they proposed, which makes the auction model they proposed untruthful. In contrast, we provide direct proof for our proposed payment rule, making the proposed auction model truthful; this gives users incentives to bid truthfully.

Last but not least, Zhang et al. [21] use Dynamic Programming to solve the knapsack-liked resource allocation problem, which has exponential time complexity, which is not practical in real-world MEC scenarios. Conversely, our proposed greedy approximation algorithm (Algorithm 1) has linear time complexity, which is significantly lower than that of the algorithm Zhang et al. [21] proposed, which makes Algorithm 1 the better choice for real-world MEC environments.

In the next chapter, we extend this single-round mechanism with a fairness mechanism since fairness-issue in a recurrent auction environment need to be dealt with to avoid the bidder dropout problem.

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

Chapter 4 Fairness-aware Resource Allocation and Pricing Model

4.1 Fairness Issue in Recurrent Auction Environments

As stated in [24], fairness is a research topic that spans various disciplines and is often associated with allocating resources. The notion of fairness permeates every aspect of our lives. In economics, the equitable distribution of revenues among shareholders, providing economic assistance to those in need, and sharing resources within a society are typical areas that require consideration of fair allocation. In computer architecture, the fair distribution of computing resources among various processes and their threads is essential. In contrast, in computer network research, nodes are expected to receive equitable bandwidth shares and quality of service (QoS).

Most of the works that studied auction-based resource allocation focus on maximizing social welfare, which is the aggregated payoff of all agents involved in the auction. Such an approach works well when the auction is only held for one round. While the auction is being held for multiple rounds, this utilitarian view of social welfare will cause

doi:10.6342/NTU202302400

problems such as bidder drop problem [25] and the asymmetric balance of negotiation power [26].

The bidder dropout problem in auctions refers to the situation where bidders who are interested in an item drop out of the auction before it ends, and this leads to the asymmetric balance of negotiation power between the remaining bidders and the sellers, which can result in the lower final price when compared with the case that all bidders had stayed in the auction. The bidder drop problem can happen for various reasons, such as bidders losing interest or becoming discouraged by high bids from competitors.

In this research, for simplicity, we assume that there are two types of mobile users, type 1 and type 2. They follow different tariff plans. Type 1 users are powerful, resourceful, and wealthy. They can bid for a resource package at a higher price, while type 2 users are less wealthy and resourceful. They cannot afford to bid at such a high price in the long run.

The first problem is the bidder dropout problem, which arises due to the nature of the auction model, where bidders (type 1 users) with a higher bid have more chance to win than bidders (type 2 users) with a lower bid. In the recurrent auction scenario, bidders with less wealth to bid may need help getting the resources they want in a limited auction round, i.e., they will continue losing because of the capability of wealthier bidders submitting bids with higher value. Once the dissatisfaction accumulates and the losers' patience runs out, they may consider dropping out of the auction, which leads to the second problem, i.e., the asymmetric balance of negotiation power. After the losers drop out of the auction, the wealthier bidders remain in the auction and can form an oligopoly due to a lack of competitors. They can control the market price and lower their bidding price, which can

doi:10.6342/NTU202302400

cause bankruptcy for the auctioneer.

Thus, we aim to design a fairness-aware auction model to mitigate the abovementioned problems, prevent type 2 users from dropping out of the auction, and maintain the market's competitiveness. Ultimately, more bidders mean greater competition, leading the market to healthier conditions.

4.2 Fairness Weight and F-index

To build a fairness-aware auction model, inspired by [25] [27], each mobile user is given a fairness weight, for user i, denote its fairness weight w_i as

$$w_i = 1 - \frac{\max(0, L_{max} - L_i)}{L_{max}} \tag{4.1}$$

Where w_i is the fairness weight for user i, w_i ranges from 0 to 1, and the higher the weight of the user, the higher consecutive lost auctions and its probabilities of leaving the market. A higher w_i means user i is given more chance to win the auction, and vice versa. L_{max} is the maximum losing streak a user can take, and it is assumed that if a user keeps losing for a number greater than L_{max} rounds, this user will quit participating in the following auction rounds. L_i is the current losing streak of user i. When user i won in the last round of auction, L_i is reset to 0, which makes w_i become 0. If user i keeps losing, then L_i will be closed to L_{max}, making w_i approach 1.

By the definition of w_i, the greedy resource allocation algorithm (Algorithm 1) mentioned in Chapter 3 needs the following modification: in the original context, at line 6, the users' requests R are sorted in non-increasing order of B_j, $\forall j \in 1, 2, \cdots, M$. Now we

doi:10.6342/NTU202302400

first sort the requests based on the fairness weight in non-increasing order and then sort the requests again based on b_j in non-increasing order. After the sorting part is done, we begin the resource allocation and pricing determination process. Finally, the losing streak for all users must be recorded and updated at the end of each auction round. The complete pseudo-code is called Algorithm 2, and it is illustrated in Figure 4.1.

With Algorithm 2, the auctioneer can improve the system's fairness when resources are insufficient.

Users' fairness can be improved by using the fairness weight, but a fairness index is needed to quantify the level of fairness. Since our goal is to make sure the maltreated users will not leave the system due to a lack of patience, in this research, we consider fairness from a user-oriented perspective. Thus, instead of using QoS as our evaluating metrics, we use F-index (4.4), which is a QoE fairness index proposed in [28]. While Jain's index satisfies many desired properties such as population size independence, scale, and metric independence, it does not satisfy deviation symmetric, i.e., Jain's index is not at its minimum when the users' experience is maximally different. On the other hand, F-index satisfies deviation symmetric, reflecting the system perspective of fairness and quantifying fairness of the entire system across all users.

Inspired by [29], we defined the mapping model from QoS to QoE as follows (4.3). The QoS index is defined as the number of rounds a user has to wait before he gets the resource he requested in the recurrent auction, i.e., the consecutive number of rounds lost for a user. The QoS index and QoE index are evaluated once per auction round. $Q(x)$ is the mapping function from the QoS index to the QoE fairness index, L, H is the lower bound and upper bound of the QoE index, in this search, L equals 0, and H equals to

doi:10.6342/NTU202302400

Require: requests R($r_1, r_2, \cdots, r_M$), resources' capacity Q($Q_1, Q_2, \cdots, Q_N$), Losing Streak Vector L($L_1, L_2, \cdots, L_M$)

Ensure: Allocation Matrix X, Social Welfare V, Payment Matrix P

$X \leftarrow 0$
$V \leftarrow 0$
$P \leftarrow 0$
Delete request r_j in R if $t_j^d > t_j^T, \forall j \in 1, 2, \cdots, M$
$w_j \leftarrow 1 - \frac{\max(0, L_{max} - L_j)}{L_{max}}, \forall j \in 1, 2, \cdots, M$
Sort R in non-increasing order of w_j
Sort R in non-increasing order of b_j
$i \leftarrow 1$
while $Q \neq 0$ and $i \leq M$ **do**
    pop r_j from the head of sorted requests vector
    find the edge node with the most g_j packages left, denoted this edge node as i
    if package g_j's capacity of edge node i is greater than zero **then**
        $V \leftarrow V + b_i$
        $q_i^{g_j} \leftarrow q_i^{g_j} - 1$
        $X_{i,j} \leftarrow 1$
        $L_j \leftarrow 0$
        Write the winning bid price, denoted as b_j, for the resource package g_j in edge node i, to the corresponding entry in P
        Schedule a resource release event of package g_j from edge node i in the corresponding time index in E using t_j^T
    else
        $L_j \leftarrow L_j + 1$
    end if
    $i \leftarrow i + 1$
end while

Figure 4.1: Algorithm 2 - The pseudo-code of resource allocation in a single round of fairness-aware auction model

doi:10.6342/NTU202302400

L_{max}. σ is the standard deviation of all users' QoE index in a given round. At the same time, σ_{max} is the maximum value possible of the standard deviation, which happens when half of the users have the maximum QoE index value (H), and the other half have the minimum (L), so σ_{max}, in this case, is equal to $\frac{H-L}{2}$.

$$Q : x \rightarrow y = Q(x) \in [L, H] \tag{4.2}$$

$$Q(x) = (H - L) \cdot e^{-\beta x} + L \tag{4.3}$$

$$F = 1 - \frac{\sigma}{\sigma_{max}} = 1 - \frac{2\sigma}{H - L} \tag{4.4}$$

Figure 4.2 is an example bar chart of the mapping function (Equation 4.3) when $H = 40, L = 0$, and $\beta = 0.07$. The decline of QoE as the QoS increases characterizes the decrease in patience for mobile users when they experience consecutive losses in recurrent auction scenarios. One can adjust the β parameter to simulate different loss tolerance levels of mobile users.

doi:10.6342/NTU202302400

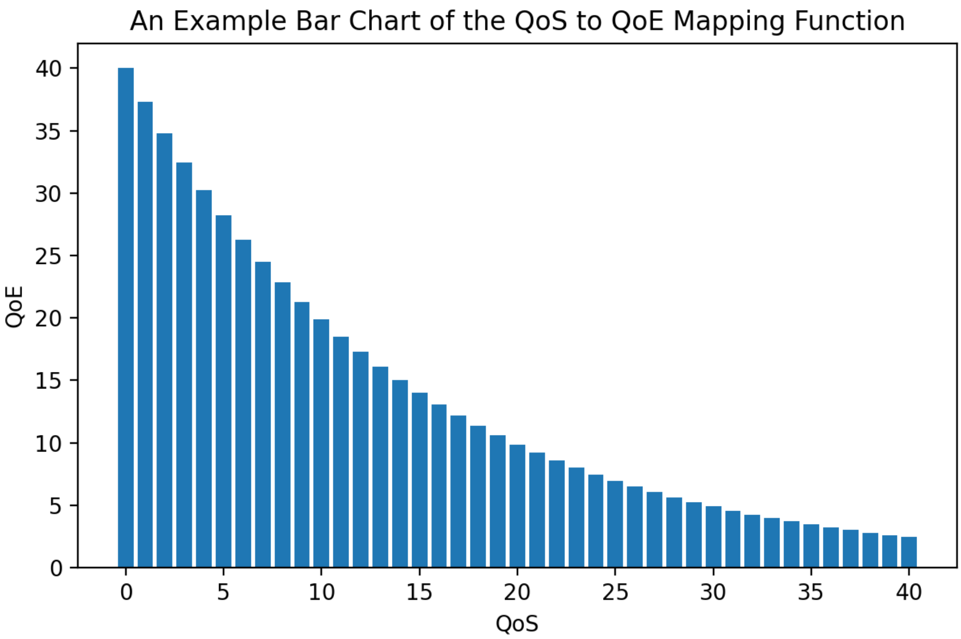

Figure 4.2: An Example Bar Chart of the QoS to QoE Mapping Function, where $H = 40, L = 0, \beta = 0.07$

Using the F-index, the fairness of each single auction round can be quantified and evaluated.

4.3 Payment Rule

Let $p_{i,j,k}$ denote the payment required from user j if said user is victorious in obtaining resource package k of edge node i. The payment rule for the fairness-awarded auction model proposed herein is dictated by the principles of the First-Price Auction [10]. In this model, the winner of resource package k is obligated to remit payment equivalent to its original bid b_j. This can be encapsulated in the following mathematical representation:

$$p_{i,j,k} = b_j \tag{4.5}$$

doi:10.6342/NTU202302400

Despite its apparent simplicity, it is crucial to note that this payment scheme fails to uphold truthfulness. According to Myerson's Lemma [11], truthfulness is preserved only when the allocation rule displays monotonicity within the confines of a single-environment auction. However, in the fairness-aware auction mechanism proposed, our allocation rule noticeably lacks this monotonicity. This deviation primarily stems from the fact that the allocation of resource packages also incorporates the fairness weight. To put it succinctly, this implies a deliberate trade-off between the assurance of incentive compatibility and the pursuit of fairness.

4.4 Procedure of a Centralized recurrent Auction

This section explains the procedure of a centralized recurrent auction in detail. The Auction Round Span is the duration of a single-round auction. In this research, we consider a discrete-time system for simplicity. In each round, as depicted in Figure 4.4, the procedure of each round is the same as the procedures described in Section 3.3, except that now every task that gets dispatched by edge nodes will be released and be free to be used by mobile users in future rounds. For all tasks that finish their execution before Resources Information Update in round $n + 1$, resource packages they occupied will be available for use at round $n + 1$. Its corresponding update event will be processed at the beginning of auction round $n + 1$, and the edge node's resource capacity will be updated correspondingly. Take Figure 4.4 as an example; application task 1 and 2 are not available at round $n + 1$, while task 1 is available.

doi:10.6342/NTU202302400

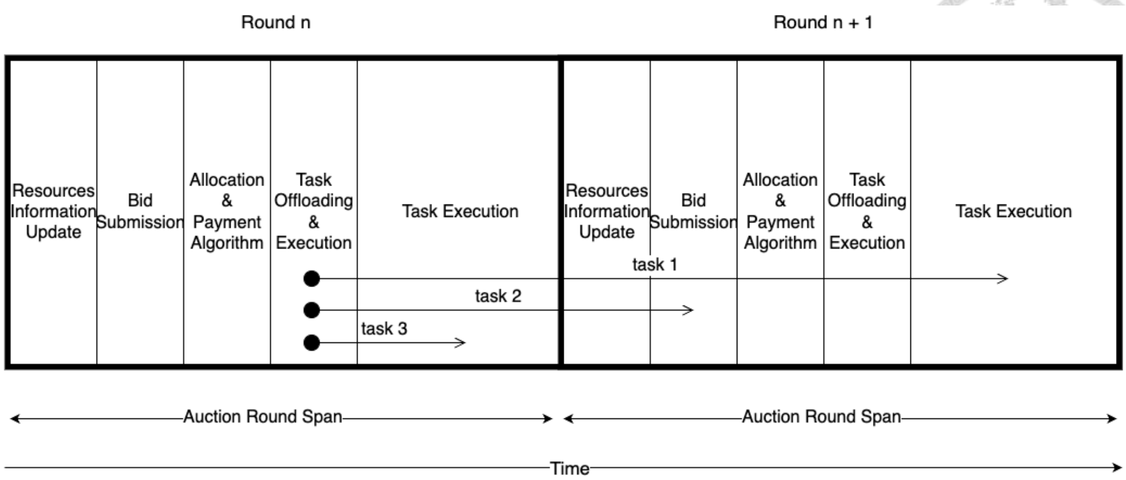

Figure 4.3: The relation between consecutive auction rounds

In this recurrent auction model, the arrival of a user's bid submission is modeled as a Poisson Process. Since a discrete-time system is assumed, it can be proved that the Poisson process is equivalent to a continuous-time version of the Bernoulli trials process. Therefore, the number of arrivals in a single round of auction is the sum of a set of Bernoulli random variables. i.e., whether a user submits a bid request in each auction round is modeled as a single Bernoulli trial. The decision flow chart for a user i in each round of auction is presented in Figure 4.4.

After users submit their bid request to the auctioneer, the auctioneer runs Algorithm 2 to determine the resource allocation results, calculates the fairness index value, and updates the fairness weight for each user for the next auction round.

The detailed procedure of a round in the recurrent auction is as follows.

Step 1: Process all resource update events for this round r, the resource update events consists of resource packages that are being used to execute the offload tasks and the execution finished after the beginning of the last round $r-1$ and before the start of this round r. For each package, the corresponding edge node's capacity is updated.

doi:10.6342/NTU202302400

Step 2: Mobile users employ edge nodes' resource capacity information to generate their bid requests based on the procedure illustrated in Figure 4.4.

Step 3: The auctioneer collects all the requests submitted and runs Algorithm 2 to determine resource allocation and pricing. Finally, the losing record is updated for each user that attends this auction round.

Step 4: The fairness weight for each mobile user is updated based on its losing record.

Step 5: Auction round r ends, and the system enters round $r + 1$.

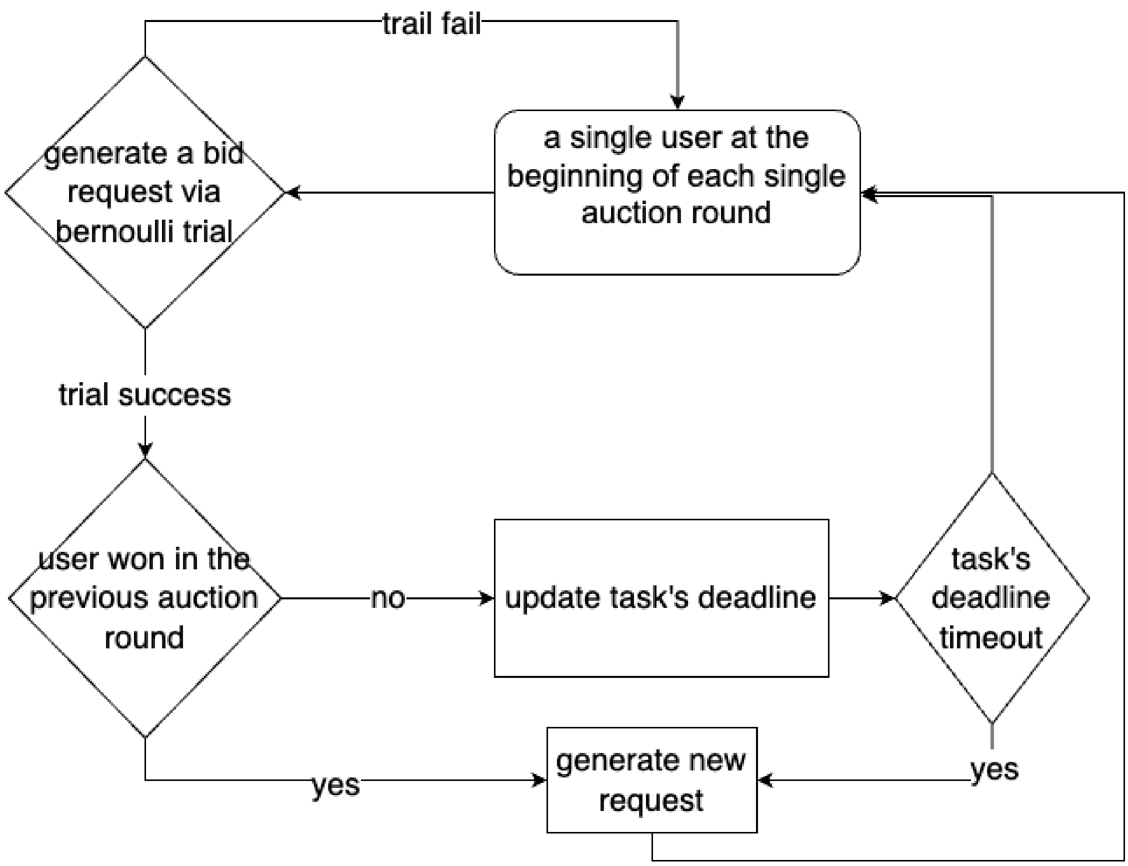

Figure 4.4: Flow Chart of a single mobile user to generate its request in an auction round

doi:10.6342/NTU202302400

4.5 Concluding Remark

In this chapter, a fairness-aware auction model is proposed. It is built on top of the single-round auction model proposed in chapter 3. While Wang et al. [20] and Zhang et al. [21] both proposed the use of the multi-round auction model, they did not consider fairness when it comes to resource allocation, leading to bidder dropout in the long run. In this thesis, our proposed fairness-aware auction model focuses on fair resource allocation, which can effectively solve the bidder dropout problem. For simplicity, we assumed only two types of mobile users: type 1 users are wealthier than type 2 users. The proposed model is expected to prevent type 2 users from dropping out of the auction in a recurrent auction environment, avoiding the bidder drop problem from arising, and further preventing the asymmetric balance of negotiation power between the mobile operator and type 1 users. In other words, our work can prevent the oligopoly market from arising, thus protecting the mobile operator's profit in the long run.

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

Chapter 5 Simulation Analysis

5.1 Dataset Description

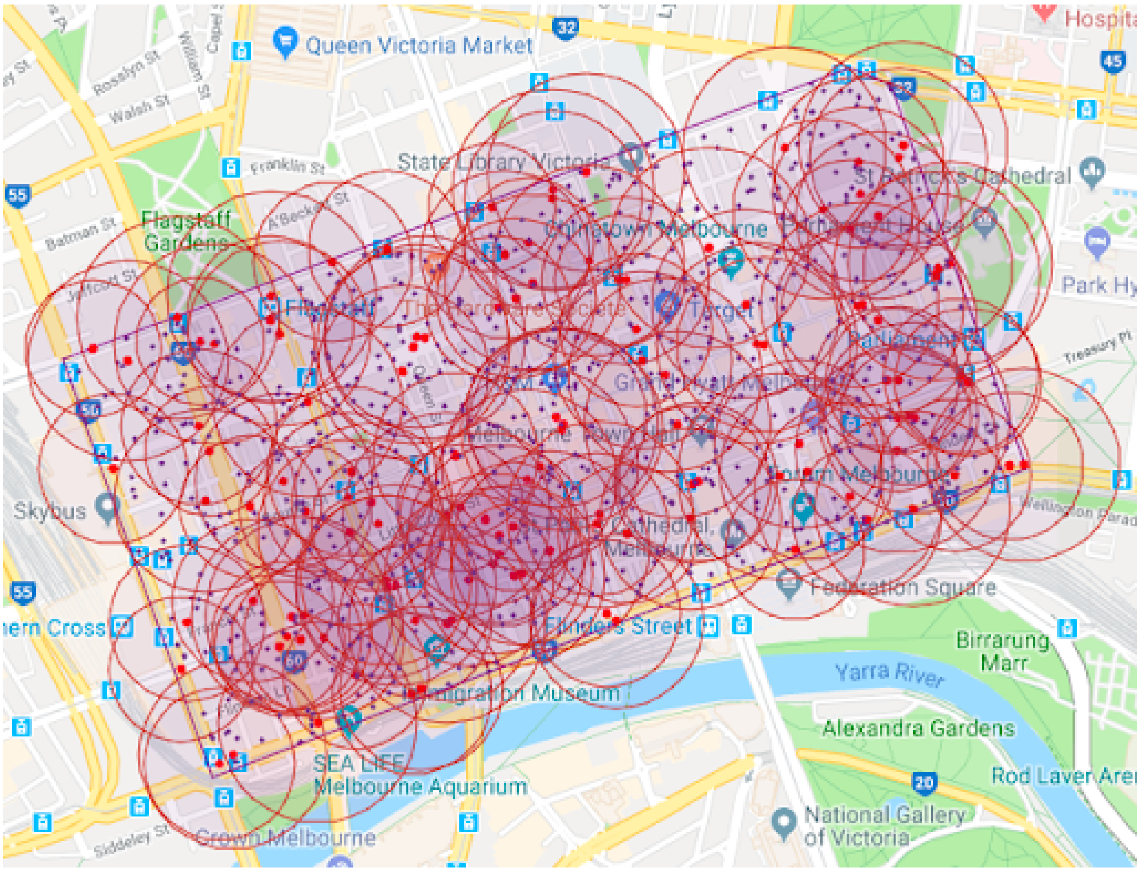

Figure 5.1: Edge nodes with their coverages and end-users positions in Melbourne central business district(CBD)

Because real users request data and their locations have not been publicly released by mobile operators yet, we adopt a widely-used dataset called EUA Datasets [30] in our

doi:10.6342/NTU202302400

simulations, which only contains the geographic location of 125 base stations (Figure 5.1) located at Melbourne Metropolitan area in Australia, where the circle represents each base stations coverage, and the dots are mobile users' geographic locations generated. The total area of this polygon, representing the Melbourne Metropolitan region, is 1.755036 square kilometers.

For mobile users, we uniformly distribute their location in the polygon formed by the frontier of the base stations as mentioned above, as illustrated by Figure 5.2, where the dots are mobile users, the lines are the frontier formed by base stations, and the x and y-axis represent longitude and latitude, respectively.

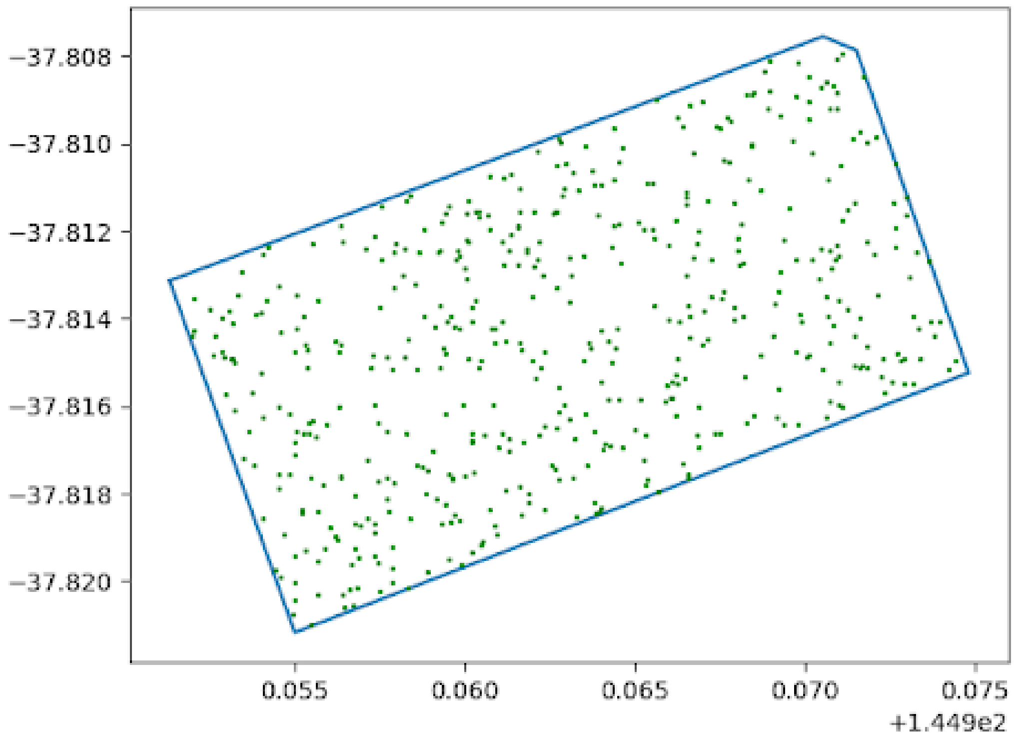

Figure 5.2: Uniformly generated mobile users in the polygon formed by the frontier of base stations

doi:10.6342/NTU202302400

5.2 Simulation Settings

5.2.1 Single Round Auction

The distance between two base stations typically ranges from 0.5 to 1 km. Hence the cell radius for each base station is set to 300 meters in this simulation.

Similar to [21], we set the task deadline to be uniformly distributed from 1 second to 5 seconds, and referring to [20] [21], the task size is uniformly distributed from 10 KB to 1 MB, CPU cycle requested by consumer is uniformly distributed from 200 Mega cycles to 1000 Mega cycles, the transmission power of user's mobile device is uniformly distributed from 257 mW to 326 mW, noise constant is set to -50 dBm. The channel gain constant is set to 4 [22], since the environment we considered is an urban area similar to the Melbourne Metropolitan area in Australia.

Similar to 5G band n78 owned by Chunghwa Telecom [31], assume Time Division Duplex (TDD) and a 90 MHz channel are used, with a total number of resource blocks equal to 246. The resources package are formed into the following format, which is similar to that of Amazon EC2 [4]: for computing resource, the CPU frequency c_z is set to 2.4 and 3.3 GHz, and for wireless resource, the number of resource block is set to 11, 16, 32, 64. Thus each edge node E_i has 8 kinds of unique package, i.e. $S_i = (s_1, s_2, \cdots, s_8)$, with each package's capacity equals 1, i.e. $q_i^1 = q_i^2 = \cdots = q_i^8 = 1$. With this setting, the theoretical data rate ranges from roughly 100 to 750 Mbps, which fits the mobile service data rate in Taiwan, according to a survey conducted by OPENSIGNAL in December of 2022 [32], where the average data rate of downlink transmission for mobile users ranges from 107 to 374 Mbps.

doi:10.6342/NTU202302400

As for the bid value of mobile users, in a single round auction, for user j, its bid generation follows (3.8).

5.2.2 Recurrent Auction Environments

As for the settings of recurrent auction environments, some additional parameters exist to set up.

First, the percentage of type 1 users is set to 20 percent, and 80 percent for type 2 users. Their bid value is generated the same way using (3.8), except that the unit bid for type 1 and type 2 users are denoted as u_j^1 and u_j^2, respectively, in which the lower bound of u_j^1 is always greater than that of u_j^2. The maximum losing streak L_{max} is set to 100 inspired by [27]. Since the average of the user's offloaded job's total execution time is around 0.1 seconds, we set the auction round span to 0.01 seconds, i.e., the auctioneer hosts a single round of auction every 0.01 seconds. Besides, the auction's resource allocation algorithm (Algorithm 2) maximum execution time is around 0.0003 seconds, so this setup is feasible.

We use $X \sim [a, b]$ to denote a uniform distribution for a variable X in Table 5.1.

The following simulations are implemented in Python and conducted on an Apple M1 system with 16 GB RAM.

doi:10.6342/NTU202302400

Parameter	value
Cell radius	300(m)
Task deadline t_j^d	[1s, 5s]
Task size s_j	[10KB, 1MB]
CPU cycle requested f_j	[200M, 1000M]
Transmission power P_j	[257mW, 325mW]
Noise constant N_0	-50dBm
Channel gain constant H	4
Light speed o_w	299792.458 (Km/s)
Sub-carrier spacing	30 KHz
Resource block capacityW_i	246
CPU frequency c_z	2.4, 3.3 GHz
Resource block w_z	11, 16, 32, 64
Unit bid u_j	[1, 10]
Type 1 user's unit bid u_j^1	[2, 4]
Type 2 user's unit bid u_j^2	[1, 3]
Auction round span	1 millisecond
Maximum losing streak L_{max}	100

Table 5.1: Simulation parameter settings

5.3 Simulation Results of the Single Round Auction

5.3.1 Target of Comparison

When a user is limited to offloading its application to a singular edge node, it necessitates the acquisition of both computing and wireless resources from that same edge node. When an edge node sells its computing and wireless resources separately, a user must procure these complementary resources sequentially. Under these conditions, a user

doi:10.6342/NTU202302400

can first attempt to secure and bid for computing resources. Following the successful procurement of computing resources, the user can subsequently strive to acquire and bid for wireless resources or take the inverse approach.

In this research, we introduce the metric of "Exposure Percentage," which represents the fraction of users who can acquire either computing or wireless resources upon the conclusion of the auction.

The targets for comparison in our analysis are as follows:

1. CPLEX: A high-performance mathematical programming solver provided by IBM ILOG CPLEX optimization studio for academics initiative [5] that can solve linear programming. This solver represents a robust, industry-standard, state-of-the-art approach.

2. Computing Resources First (CRF): Under this strategy, users bid for computing resources first; if successful, they bid for wireless resources. This provides a sequential, resource-specific bidding process that can be used as a baseline for comparison.

3. Wireless Resources First (WRF): Conversely, under this approach, users bid for wireless resources first and, if successful, subsequently bid for computing resources. This strategy offers a variant perspective on the bidding sequence for the complementary resources.

For Figure 5.3 and Figure 5.4, we ran simulations for different numbers of mobile users, ranging from 100 to 2000. The simulation process is repeated for 20 times, and then the average results are presented. In assessing the incoming traffic, we implement a three-tier classification system based on the number of users. We designate the incoming

doi:10.6342/NTU202302400

traffic as ”low” when the user count is below 500. As the number of users falls between 500 to 1250, we categorize the incoming traffic as ”high.” Lastly, when the number of users exceeds 1250, we classify the traffic as ”overloaded.”

5.3.2 Comparison of Served User Percentage

First and foremost, we set the percentage of served users obtained by CPLEX to 1, i.e., 100%, as a baseline.

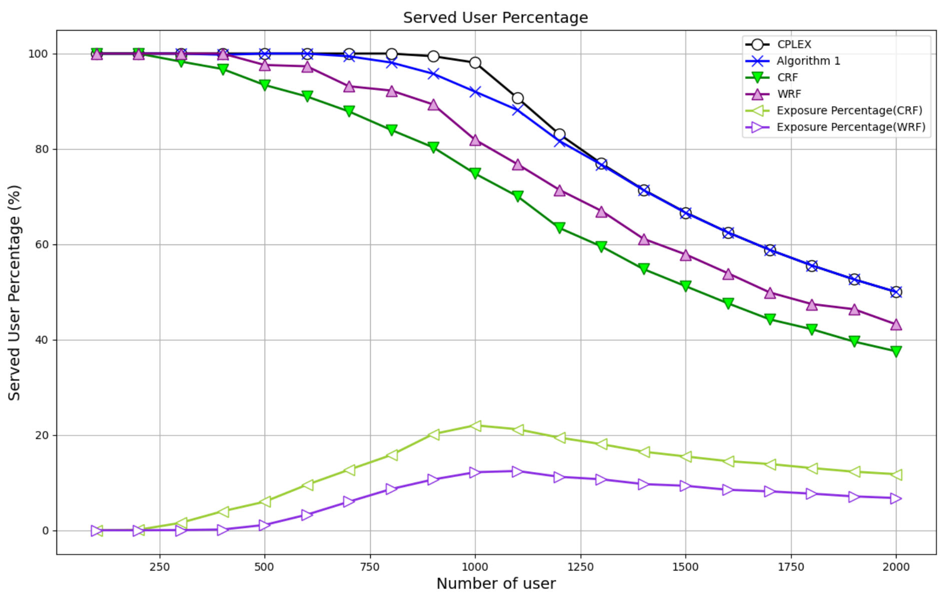

Figure 5.3: Comparison of the percentage of served mobile users.

As illustrated in Figure 5.3, our proposed algorithm (Algorithm 1) achieves near-optimal results across all traffic scenarios. During periods of low incoming traffic, the performance of Algorithm 1 aligns precisely with that of the CPLEX solver. In contrast, the Computing Resources First (CRF) and Wireless Resources First (WRF) strategies fall short of the 100% benchmark.

In the context of high incoming traffic, Algorithm 1 surpasses CRF and WRF, reach-

doi:10.6342/NTU202302400

ing a maximum performance margin of 20%. As the system becomes overloaded with incoming traffic, our proposed mechanism maintains a served user percentage over 50%, mirroring the performance of CPLEX. In contrast, CRF and WRF strategies trail CPLEX by a minimum of 10%.

These outcomes suggest that Algorithm 1 is capable of achieving near-optimal results with significantly less execution time compared to CPLEX. Notably, Algorithm 1 outshines CRF and WRF during periods of low incoming traffic, and this performance gap further widens as the incoming traffic escalates and overloads the system.

For the exposure percentage for both the CRF and WRF, they initially ascend during periods of low incoming traffic, subsequently peaking as the incoming traffic escalates. Ultimately, these percentages stabilize at a steady ratio when the system is inundated with incoming traffic or, in other words, when the traffic load becomes overloaded. This dynamic can be attributed to the intensifying competition for resources. As this competition amplifies, the exposure percentage increases, reducing the served user percentage. However, once the traffic load reaches overload status and competition becomes extreme, the exposure percentage declines to a comparably low ratio.

Regarding scenarios where the exposure percentage for CRF exceeds that of WRF, translating to a lower served user percentage for CRF compared to WRF, a logical explanation can be derived from our simulation parameters. Specifically, the resource packages are constituted as follows: for computing resources, CPU frequency c_z is set at either 2.4 or 3.3 GHz, and for wireless resources, the number of resource blocks is configured at 11, 16, 32, or 64. These parameters yield eight distinct packages and double to sixteen when both types of resources are sold separately. Given that there are only two potential

doi:10.6342/NTU202302400

CPU frequencies but four possible quantities of resource blocks, resource block packages are in scarcer supply than those offering CPU frequency when computing and wireless resources are sold separately. Consequently, a user is statistically more likely to secure both resources when first securing a resource block package. This circumstance elucidates why the served user percentage is higher for WRF.

5.3.3 Comparison of Social Welfare

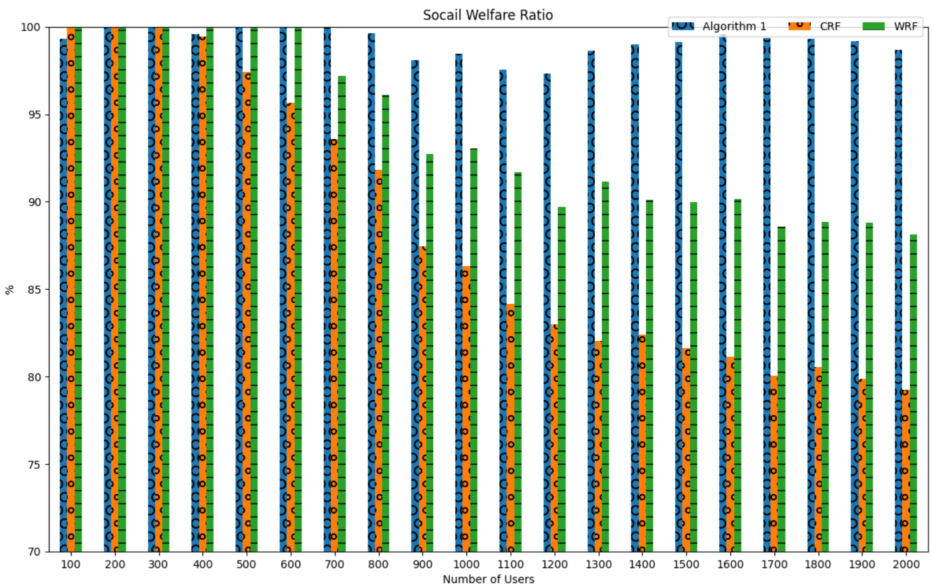

Figure 5.4: Comparison of the Relative Ratio of Social Welfare.

Secondly, establishing the social welfare generated by CPLEX as the baseline at 100%, we observe from Figure 5.4 that our proposed Algorithm 1 consistently maintains a relatively high ratio of 95% across all traffic conditions. During periods of low incoming traffic, both CRF and WRF strategies manage to uphold ratios exceeding 95%. However, with the escalation of incoming traffic, the ratios for CRF and WRF begin to fall, dropping below 90% and 95%, respectively. As the traffic influx reaches the point of system over-

doi:10.6342/NTU202302400

load, CRF dwindles further to less than 80%, while WRF dips below 90%. The generated social welfare is directly linked to the served user percentage, with increased user service leading to greater social welfare.

The reason behind the higher social welfare ratio for WRF compared to CRF is again derived from the specifics of our simulation parameters. Specifically, the structure of resource packages is as follows: for computing resources, the CPU frequency is set at either 2.4 or 3.3 GHz, whereas for wireless resources, the number of resource blocks is set at 11, 16, 32, or 64. These settings result in eight packages and double to sixteen when both types of resources are sold separately.

Considering only two possible CPU frequencies but four potential quantities of resource blocks, resource block packages are inherently scarcer than those offering CPU frequency when the resources are sold separately. Hence, users are statistically more likely to secure both resources when they first secure a resource block package. This scenario provides a clear explanation of why the served user percentage and, consequently, the social welfare ratio tends to be higher for the WRF strategy.

5.3.4 Concluding Remark

The simulation results show that our proposed one-round auction mechanism can obtain near-optimal results, effectively avoid the exposure problem, and allocate both types of resources from edge nodes to mobile users with a low execution time of around 0.0003 seconds, which is practical in real-world scenarios.

One thing to be noted is that our proposed mechanism not only performs well when the incoming traffic is low, but it also performs exceptionally well when the incoming

doi:10.6342/NTU202302400

traffic is high and overloads the system.

5.4 Simulation Results of Fairness-Aware Auction Model in Recurrent Auction Scenario

As stated in section 5.6, the proposed fairness-aware multi-round auction mechanism is expected to prevent type 2 users from dropping out of the auction and avoid the bidder drop problem.

5.4.1 Target of Comparison

The targets for comparison in this analysis are as follows:

1. Fairness-Aware Auction Model (FAAM): Auction Model proposed in chapter 4, using Algorithm 2 to determine resource allocation and pricing in a single auction round.

2. Non-Fairness-Aware Auction Model (NFAAM): Auction Model proposed in chapter 3, using Algorithm 1 to determine resource allocation and pricing in a single auction round.

Firstly, we observe the influence of the proposed fairness mechanism under conditions of a fixed arrival rate. Secondly, we examine the impact of this mechanism under varying arrival rates. As explicated in section 4.4, the process of a user submitting a bid request in each auction round is modeled as a Bernoulli trial.

doi:10.6342/NTU202302400

In the scenario with a fixed rate, the success probability for the Bernoulli trial is established at 0.2. In the multiple rates scenario, the arrival rate is adjusted between 0.01 and 0.26 with a gap of 0.01 for analysis. The total number of mobile users in both scenarios is set to 1,000. The total number of auction rounds conducted is fixed at 1,500. The entire simulation is replicated 20 times to ensure robustness, with the average results subsequently reported.

There are three target metrics we want to observe, the first is the dropout percentage for both type 1 and type 2 users under FAAM and NFAAM, the second is the fairness index value for both types of users under FAAM and NFAAM, and the third is the social welfare ratio of FAAM to NFAAM.

5.4.2 Comparison of Fixed Arrival Rate in Recurrent Auction Scenario

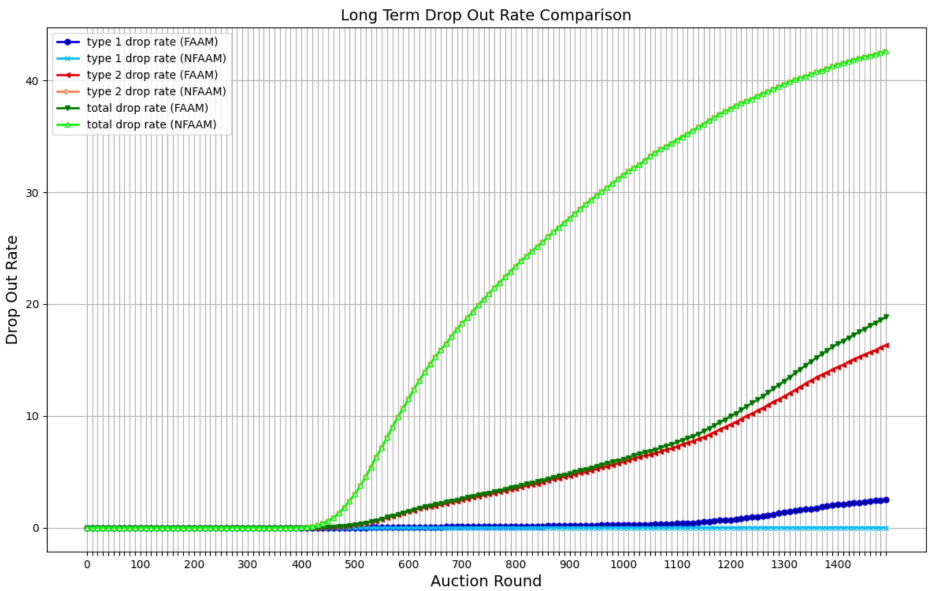

Figure 5.5: The effect of Fairness Mechanism on User's Drop Out Percentage

doi:10.6342/NTU202302400

From Figure 5.5, it is observable that within the Non-Fairness-Aware Auction Model (NFAAM), the dropout percentage for type 2 users sharply escalates as the auction round approaches 400 rounds, eventually surging to 42% by the time the auction round culminates at 1500 rounds. In contrast, the dropout percentage for type 1 users within NFAAM remains nearly stagnant at 0%. This is expected due to the inherent structure of auctions that favor bidders who propose higher bids.

In the case of our proposed Fairness-Aware Auction Model (FAAM), it effectively curtails the dropout percentage for type 2 users to around 16% while concurrently maintaining the dropout percentage for type 1 users at a mere 2.5%. Consequently, the total dropout percentages for NFAAM and FAAM are 42% and 18%, respectively.

These findings substantiate the effectiveness of our proposed FAAM in significantly reducing the dropout percentage for type 2 users and, in turn, the total dropout percentage. This result highlights the importance of incorporating fairness considerations into auction models, particularly in recurrent auction environments with diverse types of participants.

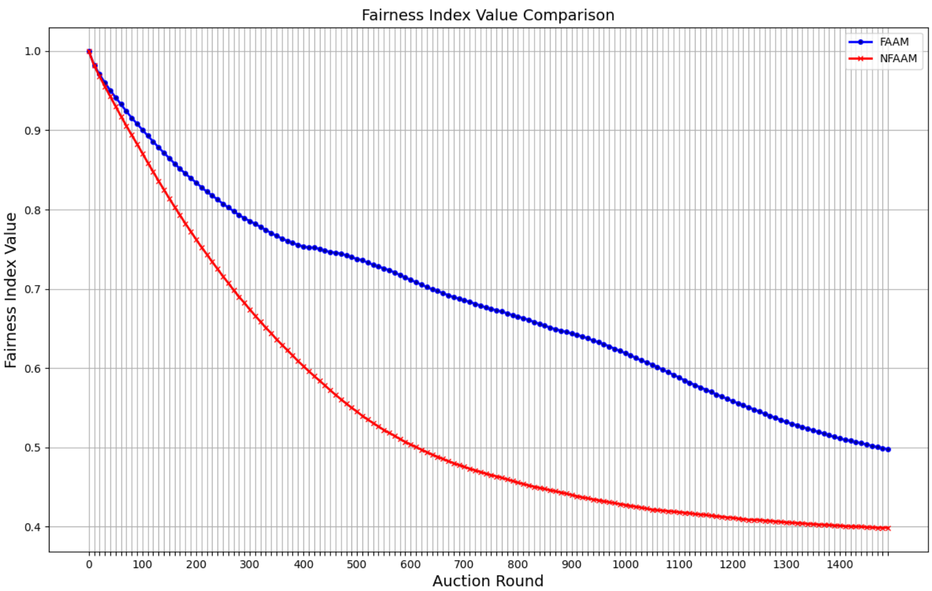

Figure 5.6: The effect of Fairness Mechanism on Fairness Index Value

doi:10.6342/NTU202302400

As depicted in Figure 5.6, the fairness index values for the FAAM and NFAAM decrease as the auction proceeds. However, the decline in the fairness index is notably more rapid and pronounced for NFAAM, with the disparity between NFAAM and FAAM widening as the recurrent auction continues. The maximum margin observed is around 20%. This trend is intuitive, given that the total dropout percentage for NFAAM is considerably higher than FAAM's.

These findings underscore the importance of fairness considerations in auction models. While both models experience a decline in fairness over time due to the inherent dynamics of auctions, the proposed FAAM demonstrates significantly improved sustainability of fairness compared to its non-fairness-aware counterpart. This has crucial implications for long-term user engagement and overall system performance.

5.4.3 Comparison of Multiple Arrival Rates in Recurrent Auction Scenario

In this subsection, the effectiveness of the fairness-aware auction model is evaluated under a range of different arrival rates. The arrival rates assessed span from 0.01 to 0.25 with increments of 0.01, yielding 26 distinct arrival rates. Arrival rates below 0.05 are classified as low incoming traffic, rates between 0.05 and 0.2 are deemed high, and rates exceeding 0.2 are characterized as overloaded.

For each arrival rate, the auction is executed for 1,500 rounds, with the results of the 1,500th round serving as the outcome. The simulation is replicated 20 times to enhance the robustness of the findings, with the average results subsequently reported.

By testing the model across this range of arrival rates, we aim to provide a compre-

doi:10.6342/NTU202302400

hensive understanding of its performance under various levels of incoming traffic. This analysis will shed light on the scalability and adaptability of the fairness-aware auction model in handling diverse operational scenarios.

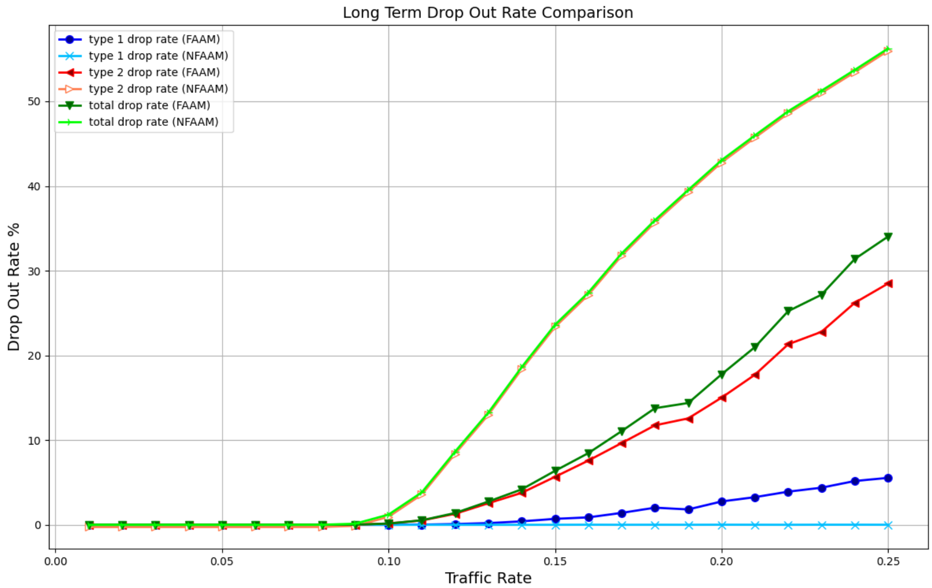

Figure 5.7: Comparison of Users' Drop Out Percentage of Multiple Arrival Rates

As depicted in Figure 5.7, it is evident that our proposed Fairness-Aware Auction Model (FAAM) can notably reduce the overall dropout percentage compared to the Non-Fairness-Aware Auction Model (NFAAM) under all tested conditions.

In situations of low incoming traffic, no users, regardless of type, drop out from the auction under either FAAM or NFAAM.

During periods of high incoming traffic, type 2 users experience a dropout percentage of up to 43% under NFAAM, compared to a substantially lower 15% under FAAM. For type 1 users, the dropout rate under NFAAM is virtually 0%, while under FAAM, it increases slightly to 2.7%. The total dropout rate for NFAAM is 43%, whereas, under FAAM, it decreases to 17.7%.

doi:10.6342/NTU202302400

When incoming traffic becomes overloaded, the dropout percentage for type 2 users can escalate to 56% under NFAAM, yet effectively halved to 28% under FAAM. For type 1 users, the dropout rate under NFAAM remains negligible at nearly 0%, while it rises modestly to 5.5% under FAAM. The overall dropout rate for NFAAM reaches 56%, whereas, under FAAM, it is significantly mitigated to 33.5%.

These results underscore the power of our proposed FAAM. Not only does it substantially reduce the dropout percentage of type 2 users, which represent the group we specifically aim to support, but it also achieves this with only a marginal increase in the dropout percentage of type 1 users. This beneficial mechanism is effective across all incoming traffic levels, demonstrating exceptional resilience, particularly under system-overload conditions.

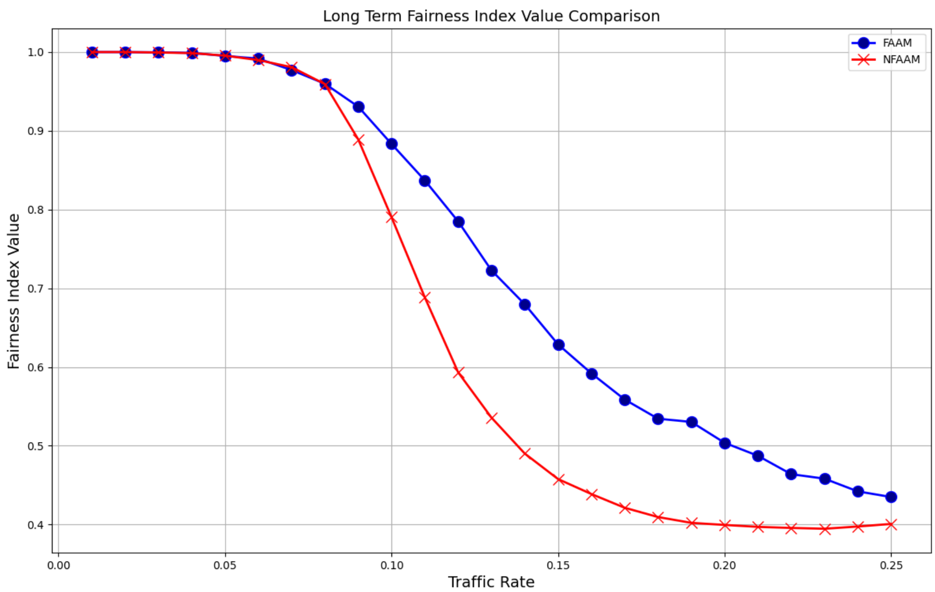

Figure 5.8: Comparison of Fairness Index Value of Multiple Arrival Rates

The fairness index results are presented in Figure 5.8. When incoming traffic is low, there is no discernible difference between the fairness index of FAAM and NFAAM. How-

doi:10.6342/NTU202302400

ever, as incoming traffic increases, FAAM maintains a notable advantage, showing an over 20% margin compared to NFAAM. This indicates greater fairness in resource allocation during high-traffic periods under the FAAM model.

Interestingly, as the system transitions to an overloaded state due to high traffic volumes, the gap between the fairness indices of FAAM and NFAAM narrows to around 10%. Even in these extreme conditions, FAAM maintains a relative advantage, demonstrating its robust capability to preserve fairness in resource allocation.

Despite this diminishing gap under maximum traffic load, it does not diminish the efficacy of FAAM. It underscores FAAM's resilience and adaptability under various degrees of system load. Most notably, FAAM exhibits strong performance under high and overloaded traffic conditions. These findings underscore the strength of the FAAM in promoting equitable resource allocation, particularly under challenging operational scenarios.

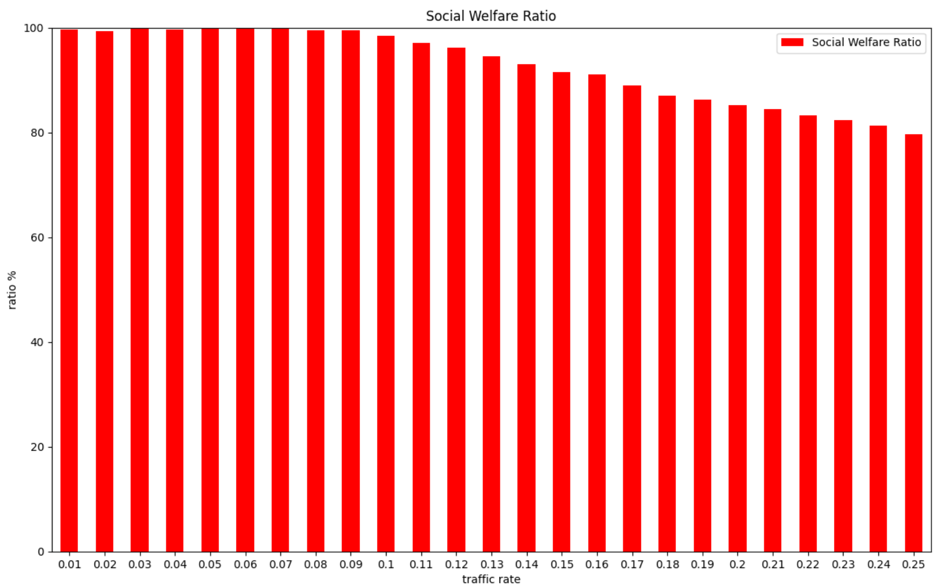

Figure 5.9: Social Welfare Relative Ratio for Fairness-Aware and Non-Fairness-Aware Auction Mechanism

doi:10.6342/NTU202302400

Figure 5.9 reveals the trade-off enacted by FAAM. Although the overall dropout percentage is reduced under FAAM, it is achieved at the expense of a higher dropout percentage among type 1 users. These type 1 users are generally willing to pay more for the packages they bid on, contributing significantly to the overall social welfare. However, this compromise can be seen as a strategic and beneficial trade-off in the long run.

The rationale for this trade-off becomes apparent when considering the potential implications of NFAAM's approach. As shown in Figure 5.7, the dropout percentage for type 2 users under NFAAM can soar to as high as 56 percent. This exceedingly high dropout rate threatens to erode the market diversity, potentially enabling type 1 users to establish an oligopolistic market once type 2 users abandon the auction.

In contrast, by preventing such an extreme scenario, FAAM safeguards the diversity and competitiveness of the market. The slightly reduced contribution to the overall social welfare of type 1 users is a worthwhile sacrifice to preserve a healthy, inclusive, and competitive auction environment. This strategy ensures the longevity and sustainability of the auction mechanism, preventing potential market distortion and the formation of a user oligopoly.

5.5 Concluding Remark

Extensive simulations have indeed verified the effectiveness of the proposed Fairness-Aware Auction Model (FAAM). Results consistently demonstrate that FAAM deters type 2 users from exiting the auction. This approach mitigates the bidder dropout issue and precludes the development of an asymmetrical balance of bargaining power between the mobile operator and type 1 users.

doi:10.6342/NTU202302400

Such an asymmetrical power dynamic would potentially decrease social welfare over time. Moreover, it could lead to an oligopoly market that could ultimately reduce the mobile operator's profit in the long term. Therefore, the FAAM's ability to maintain a diverse and competitive bidder pool is crucial to the sustainability of the mobile resource auction ecosystem.

The effectiveness of our proposed model is not limited to a specific traffic condition. The FAAM has proven to perform efficiently under various levels of incoming traffic. Whether the incoming traffic is classified as low, high, or even overloaded, FAAM consistently maintains its ability to prevent type 2 user dropout, thus fostering an inclusive and balanced auction environment. This highlights the robustness and versatility of the FAAM in managing different traffic loads while upholding the principle of fairness.

doi:10.6342/NTU202302400

doi:10.6342/NTU202302400

Chapter 6 Conclusion and Future Work

6.1 Conclusions

Auction-based models have been recognized as a highly efficient method for resource allocation in Mobile Edge Computing (MEC) systems. These systems often comprise mobile users offloading applications that require wireless channels and computing resources. Regrettably, most extant auction-based models primarily focus on allocating computing resources, often overlooking the crucial task of wireless channel resource allocation. Furthermore, the prevalent characterization of these models as single-round auctions leaves much to be desired in discussing fairness-related concerns within recurring auction environments. This study addresses these gaps, proposing an auction-based model designed to jointly allocate wireless channel and computing resources while maintaining fair resource allocation.

The research contemplates a MEC system with multiple edge nodes to which mobile users offload their applications. For an offloaded application to execute, computing and wireless resources are necessary. Subdivided into fractions, these resources are combined and offered as packages to mobile users via an auction model. For ease of analysis, the

doi:10.6342/NTU202302400

study assumes two types of mobile users, types 1 and 2, each with distinct tariff plans. Mobile users in our system are equipped with automated auction agents responsible for generating bid requests on their behalf. These agents operate according to the user's specific tariff plan, thus accurately reflecting the bidding capacity of each user in the auction process. Type 1 users, powerful and wealthy, can bid at higher prices for resource packages. Conversely, type 2 users, less affluent and resourceful, lack the financial stamina for sustained high bidding. We proposed a fairness-aware auction model to prevent type 2 users from early auction withdrawal. While Wang et al. [20] and Zhang et al. [21] both proposed the use of the multi-round auction model, they did not consider fairness when it comes to resource allocation, leading to bidder dropout in the long run. In this thesis, our proposed fairness-aware auction model focuses on fair resource allocation, which can effectively solve the bidder dropout problem.

Simulation results reveal that our proposed model excels in tackling the exposure problem for mobile users, dramatically enhancing allocation efficiency. Additionally, it significantly reduces the bidder dropout rate, ensuring fair resource allocation within a recurrent auction scenario. Presently, our model supports at least 2000 mobile users and 125 base stations. During a single-round auction, the proposed model attains a maximum performance margin of 20% over comparison targets in terms of served user percentage and social welfare. Furthermore, in recurrent auctions, it slashes the total bidder dropout percentage from 56% to 28%. It records a 20% increase in fairness index value while retaining 80% of the original social welfare compared to a non-fairness-aware multi-round auction model. Remarkably, our model thrives when incoming traffic surges and overloads the system, distinguishing it from other models and making it an optimal choice for resource allocation in resource-deficient scenarios, a common feature of typical real-world

doi:10.6342/NTU202302400

MEC systems.

In summary, the chief contribution of our research is that we proposed an effective and efficient mechanism for jointly and fairly allocating computing and wireless resources to mobile users. This is achieved via a fairness-aware auction model within a resource-competitive MEC environment.

6.2 Future Work

The main goal of selling both types of resources as packages is to solve the exposure problem when computing and wireless resources complement mobile users' offloaded applications. This prevents bidders who want bundles of resources from placing aggressive bids. The predefined package should be intricately designed to meet the actual needs of mobile users while being structured sufficiently to accommodate specific allocation and payment rules equitably. The major problem with predefined package bids is that, when they are poorly aligned with bidders' objectives, they may cause more harm than good. Thus, the optimal way of dividing resources into packages according to real-world mobile users' objectives can be further studied.

Second, other than OFDMA, Non-Orthogonal Multiple Access (NOMA) in 5G or Beyond 5G environments can be considered for further research. When designing an auction model for resource allocation in MEC, especially wireless resources, the characteristics of NOMA need to be considered.

Furthermore, in the proposed fairness-aware auction model, the bidding strategies of participants can significantly influence the auction outcomes. Over time, under the protection of the fairness mechanism, mobile users could potentially manipulate their bidding

doi:10.6342/NTU202302400

strategies by exploiting the auctioneer's resource allocation scheme, specifically its priority setting. For instance, mobile users could artificially boost their resource acquisition priority by submitting low-value false bids, thereby intentionally losing. This introduces a degree of strategic complexity that could potentially undermine the effectiveness of the fairness mechanism. As such, future research should investigate the incentive compatibility of the fairness-aware auction model to address potential strategic bid manipulation and ensure fair and effective resource allocation.

Lastly, future studies could explore the feasibility of dynamically activating the fairness mechanism better to accommodate the variable traffic rates characteristic of real-world conditions. Such adaptability could bolster the system's robustness and resilience to fluctuations, fostering a more stable and efficient resource allocation environment within the MEC system.

doi:10.6342/NTU202302400

References

[1] P. Milgrom, "Putting auction theory to work: The simultaneous ascending auction," *Journal of Political Economy*, vol. 108, no. 2, pp. 245–272, 2000.

[2] J.-S. Lee and B. Szymanski, "A novel auction mechanism for selling time-sensitive e-services," in *Seventh IEEE International Conference on E-Commerce Technology (CEC'05)*, pp. 75–82, 2005.

[3] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.

[4] "Amazon EC2 On-Demand Pricing." https://aws.amazon.com/tw/ec2/pricing/on-demand, 2016.

[5] I. I. Cplex, "V12. 1: User’ s manual for cplex," *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009.

[6] D. Kumar, G. Baranwal, and D. P. Vidyarthi, "A survey on auction based approaches for resource allocation and pricing in emerging edge technologies," *Journal of Grid Computing*, vol. 20, no. 1, pp. 1–52, 2022.

[7] V. Krishna, *Auction theory*. Academic press, 2009.

doi:10.6342/NTU202302400

[8] H. Qiu, K. Zhu, N. C. Luong, C. Yi, D. Niyato, and D. I. Kim, "Applications of auction and mechanism design in edge computing: A survey," *IEEE Transactions on Cognitive Communications and Networking*, 2022.

[9] T. Roughgarden, "Algorithmic game theory," *Communications of the ACM*, vol. 53, no. 7, pp. 78–86, 2010.

[10] J. C. Cox, V. L. Smith, and J. M. Walker, "Theory and individual behavior of first-price auctions," *Journal of Risk and Uncertainty*, vol. 1, pp. 61–99, 1988.

[11] R. B. Myerson, "Optimal auction design," *Mathematics of Operations Research*, vol. 6, no. 1, pp. 58–73, 1981.

[12] E. H. Clarke, "Multipart pricing of public goods," *Public Choice*, pp. 17–33, 1971.

[13] T. Groves, "Incentives in teams," *Econometrica: Journal of the Econometric Society*, pp. 617–631, 1973.

[14] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords," *American Economic Review*, vol. 97, no. 1, pp. 242–259, 2007.

[15] H. R. Varian, "Position auctions," *International Journal of Industrial Organization*, vol. 25, no. 6, pp. 1163–1178, 2007.

[16] A. Leo-Hansen, "How is the vcg mechanism profiting facebook?," 06 2020.

[17] G. Gao, M. Xiao, J. Wu, H. Huang, S. Wang, and G. Chen, "Auction-based vm allocation for deadline-sensitive tasks in distributed edge cloud," *IEEE Transactions on Services Computing*, 2019.

doi:10.6342/NTU202302400

[18] L. Liu, H. Ge, S. Li, X. Chen, H. Gong, and Y. Cui, “Resource allocation strategy based on improved auction algorithm in mobile edge computing environment,” in *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 355–363, IEEE, 2021.

[19] T. Bahreini, H. Badri, and D. Grosu, “Mechanisms for resource allocation and pricing in mobile edge computing systems,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 3, pp. 667–682, 2021.

[20] Q. Wang, S. Guo, J. Liu, C. Pan, and L. Yang, “Profit maximization incentive mechanism for resource providers in mobile edge computing,” *IEEE Transactions on Services Computing*, 2019.

[21] H. Zhang, F. Guo, H. Ji, and C. Zhu, “Combinational auction-based service provider selection in mobile edge computing networks,” *IEEE Access*, vol. 5, pp. 13455–13464, 2017.

[22] A. Goldsmith, *Wireless Communications*. USA: Cambridge University Press, 2005.

[23] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[24] H. Shi, R. V. Prasad, E. Onur, and I. Niemegeers, “Fairness in wireless networks:issues, measures and challenges,” *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 5–24, 2014.

[25] A. Pla, B. Lopez, and J. Murillo, “Multi-dimensional fairness for auction-based resource allocation,” *Knowledge-Based Systems*, vol. 73, pp. 134–148, 2015.

doi:10.6342/NTU202302400

[26] J. Murillo Espinar *et al.*, *Egalitarian behaviour in multi unit combinatorial auctions*. Universitat de Girona, 2010.

[27] R. Hassanzadeh, A. Movaghar, and H. R. Hassanzadeh, "A multi-dimensional fairness combinatorial double-sided auction model in cloud environment," in *2016 8th International Symposium on Telecommunications (IST)*, pp. 672–677, 2016.

[28] T. Hoßfeld, L. Skorin-Kapov, P. E. Heegaard, and M. Varela, "Definition of qoe fairness in shared systems," *IEEE Communications Letters*, vol. 21, no. 1, pp. 184–187, 2017.

[29] T. Hoßfeld, P. E. Heegaard, M. Varela, L. Skorin-Kapov, and M. Fiedler, "From qos distributions to qoe distributions: a system's perspective," in *2020 6th IEEE Conference on Network Softwarization (NetSoft)*, pp. 51–56, 2020.

[30] P. Lai, Q. He, M. Abdelrazek, F. Chen, J. Hosking, J. Grundy, and Y. Yang, "Optimal edge user allocation in edge computing with variable sized vector bin packing," in *International Conference on Service-Oriented Computing*, pp. 230–245, Springer, 2018.

[31] C. Telecom, "Chunghwa telecom 5g." https://www.cht.com.tw/home/campaign/5g/index.html, Dec 2022.

[32] F. Sam, "Taiwan mobile web experience report." https://www.opensignal.com/zh-hant/reports/2022/12/taiwan/mobile-network-experience, Dec 2022.

doi:10.6342/NTU202302400

Appendix A — Notations

N	Number of edge node
M	Number of mobile user
W_i	Capacity of resource block of E_i
N_{RB}	Number of sub-carriers per resource blocks
E_i	Edge Node i
S_i	Set of $s_i^z, \forall z \in (1, 2, \cdots, k)$
s_i^z	Type z package in S_i
Q_i	Set of $q_i^z, \forall z \in (1, 2, \cdots, k)$
q_i^z	Capacity of s_i^z
c^z	CPU frequency provided in type z package
w^z	Number of resource block provided in type z package
k	Number of unique resource packages
r_j	Request of mobile user j
b_j	Bid value of user j
g_j	Package type number user j bid for
f_j	CPU cycle required for user j's offloaded task
s_j	Size of user j's offloaded task
t_j^d	Deadline of user j's offloaded task
$x_{i,j}, g_{j,z}$	Binary decision variable

Table A.1: Notations of System Structure

doi:10.6342/NTU202302400

$w_{i,j}$	Bandwidth between user j and base station i
$a_{i,j}$	Data rate between user j and base station i
P_j	Transmission power of user j's mobile device
$h_{i,j}$	Channel gain between user j and base station i
D_i	Serving range for base station i
$d_{i,j}$	Distance between user j and base station i
N_0	Background noise constant
H	Channel gain constant
t_j^T	Total completion time for user j's offloaded task
$t_{i,j}^c$	Execution time needed for user j's offloaded task on edge node i
o_w	Speed of light

Table A.2: Notations of Network Model

w_i	Fairness weight of user i
L_{max}	Maximum losing streak for a user before his dropout
L_i	Current losing streak for user i

Table A.3: Notations of Fairness-Aware Auction Model

doi:10.6342/NTU202302400

Appendix B — Proof of Equation 3.22 in Section 3.6

We first present the proof of Myerson's Lemma [11], then use Myerson's Lemma to prove Equation 3.22.

Definition B.0.1 (Monotone Allocation Rule). *An allocation rule x for a single-parameter environment is monotone if for every bidder i and bids b_{-i} by the other bidders, the allocation $x_i(z, b_{-i})$ to i is non-decreasing in its bid z.*

Definition B.0.2 (Truthfulness). *An auction mechanism satisfies truthfulness if all agents bid truthfully, i.e., $b_i = v_i$ is a dominant strategy for each bidder: $x_i v_i - p_i$ is maximized by bidding their true valuation. Where for bidder i, b_i is the bid value, v_i is the true valuation, x_i is 1 when bidder i wins, and is set to 0 vice versa, and p_i is the payment that bidder i has to pay if he wins.*

Theorem B.0.1 (Myerson' s Lemma [11]). *For a single-parameter environment with an auction mechanism $A = (x, p)$, where x is the allocation rule, and p is the payment rule. For every bidder i and bids b_{-i} by the other bidders, denote $x_i(z, b_{-i})$ as the allocation and $p_i(z, b_{-i})$ as the payment of bidder i when it bids z. The following three claims hold.*

1. *A allocation rule is implementable if and only if it is monotone.*

doi:10.6342/NTU202302400

2. *If an allocation rule is monotone, then without loss of generality, assume that* $p_i(0, b_{-i}) = 0$*, there exists a unique payment rule p, such that the sealed-bid auction mechanism* $A = (x, p)$ *is truthful.*

3. *The payment rule* p *is given by*

$$p_i(b_i, b_{-i}) = b_i \cdot x_i(b_i, b_{-i}) - \int_0^{b_i} x_i(z, b_{-i})\, dz$$

Proof. Consider an auction with a single item, and with an allocation rule x, assume bidders bid truthfully, and consider y, z, where $0 \leq y < z$, $z = y + h,\ h > 0$. For bidder i, if it has a private valuation of z towards the item, it submits a bid y. The truthfulness property demands that

$$z \cdot x_i(z, b_{-i}) - p_i(z, b_{-i}) \geq z \cdot x_i(y, b_{-i}) - p_i(y, b_{-i}) \tag{B.1}$$

and, when bidder i has a private valuation of y towards the item, and it submits a bid z, the truthfulness property demands that

$$y \cdot x_i(y, b_{-i}) - p_i(y, b_{-i}) \geq z \cdot x_i(z, b_{-i}) - p_i(z, b_{-i}) \tag{B.2}$$

Combine B.1 and B.2, we have

$$z(x_i(z, b_{-i}) - x_i(y, b_{-i})) \geq p_i(z, b_{-i}) - p_i(y, b_{-i}) \geq y(x_i(z, b_{-i}) - x_i(y, b_{-i})) \tag{B.3}$$

The partial result of B.3 is

doi:10.6342/NTU202302400

$$
\begin{aligned}
&z(x_i(z, b_{-i}) - x_i(y, b_{-i})) \geq y(x_i(z, b_{-i}) - x_i(y, b_{-i})) \\
\Rightarrow &(z - y) \cdot (x_i(z, b_{-i}) - x_i(y, b_{-i})) \geq 0
\end{aligned}
\tag{B.4}
$$

, which implies the monotonicity of the allocation rule.

By dividing B.3 by $z - y$, we have

$$
z\frac{(x_i(z, b_{-i}) - x_i(y, b_{-i}))}{z - y} \geq \frac{p_i(z, b_{-i}) - p_i(y, b_{-i})}{z - y} \geq \frac{y(x_i(z, b_{-i}) - x_i(y, b_{-i}))}{z - y} \tag{B.5}
$$

By replacing z with $y + h$

$$
(y+h)\cdot\frac{(x_i(y + h, b_{-i}) - x_i(y, b_{-i}))}{h} \geq \frac{p_i(y + h, b_{-i}) - p_i(y, b_{-i})}{h} \geq y\cdot\frac{(x_i(y + h, b_{-i}) - x_i(y, b_{-i}))}{h} \tag{B.6}
$$

Fix z and take the limit $y \to z$, i.e., taking $h \to 0^+$, and since x is monotone, we have

$$
p_i^{'}(z, b_{-i}) = y \cdot x_i^{'}(z, b_{-i}) \tag{B.7}
$$

By integrating from y = 0, using integration by parts, we have

$$
p_i(b_i, b_{-i}) = b_i \cdot x_i(b_i, b_{-i}) - \int_0^{b_i} x_i(z, b_{-i})\, dz \tag{B.8}
$$

doi:10.6342/NTU202302400

We can prove that B.8 is truthful. Under this payment rule, the utility of a user i is

$$\begin{aligned} u_i(b_i, b_{-i}) &= v_i \cdot b_i - p_i(b_i, b_{-i}) \\ &= (v_i - b_i) \cdot x_i(b_i, b_{-i}) + \int_0^{b_i} x_i(z, b_{-i})\, dz \end{aligned} \tag{B.9}$$

If $b_i \leq v_i$,

$$\begin{aligned} u_i(b_i, b_{-i}) - u_i(v_i, b_{-i}) &= (v_i - b_i)x_i(b_i, b_{-i}) + \int_0^{b_i} x_i(z, b_{-i})\, dz - \int_0^{v_i} x_i(z, b_{-i})\, dz \\ &= (v_i - b_i)x_i(b_i, b_{-i}) - \int_{b_i}^{v_i} x_i(z, b_{-i})\, dz \\ &\leq (v_i - b_i)x_i(b_i, b_{-i}) - \int_{b_i}^{v_i} x_i(b_i, b_{-i})\, dz = 0 \end{aligned} \tag{B.10}$$

If $b_i \geq v_i$

$$\begin{aligned} u_i(b_i, b_{-i}) - u_i(v_i, b_{-i}) &= (v_i - b_i)x_i(b_i, b_{-i}) + \int_0^{b_i} x_i(z, b_{-i})\, dz - \int_0^{v_i} x_i(z, b_{-i})\, dz \\ &= (v_i - b_i)x_i(b_i, b_{-i}) + \int_{v_i}^{b_i} x_i(z, b_{-i})\, dz \\ &\leq (v_i - b_i)x_i(b_i, b_{-i}) + \int_{v_i}^{b_i} x_i(b_i, b_{-i})\, dz = 0 \end{aligned} \tag{B.11}$$

Combine B.10 and B.11, we have

$$u_i(b_i, b_{-i}) \leq u_i(v_i, b_{-i}) \tag{B.12}$$

doi:10.6342/NTU202302400

, which means bidding truthfully is the dominant strategy for every bidder i. Thus, the payment rule is truthful. □

By B.0.1, we know that the payment rule is

$$p_i(b_i, b_{-i}) = b_i \cdot x_i(b_i, b_{-i}) - \int_0^{b_i} x_i(z, b_{-i})\, dz$$

In an auction with a single item, the winner has to pay the second highest price, also known as the critical bid b^*, which can be derived through the payment rule

$$\begin{aligned} p_i(b_i, b_{-i}) &= b_i \cdot x_i(b_i, b_{-i}) - \int_0^{b_i} x_i(z, b_{-i})\, dz \\ &= b_i \cdot 1 - (\int_0^{b^*} 0\, dz + \int_{b^*}^{b_i} 1\, dz) \\ &= b_i - (b_i - b^*) \\ &= b^* \end{aligned} \tag{B.13}$$

In an auction with k identical items, the result of the single-item case can be generalized. The winners in this auction have to pay the k+1-th highest price, denoted as b^*_{k+1}

$$\begin{aligned} p_i(b_i, b_{-i}) &= b_i \cdot x_i(b_i, b_{-i}) - \int_0^{b_i} x_i(z, b_{-i})\, dz \\ &= b_i \cdot 1 - (\int_0^{b^*_{k+1}} 0\, dz + \int_{b^*_{k+1}}^{b_i} 1\, dz) \\ &= b_i - (b_i - b^*_{k+1}) \\ &= b^*_{k+1} \end{aligned} \tag{B.14}$$

Thus, by B.0.1, we can prove that in an auction with k identical items, where the

doi:10.6342/NTU202302400

allocation rule is monotone, and the winners pay the k+1-th highest price, is truthful.

doi:10.6342/NTU202302400