Optimizing Task Offloading Energy in Multi-User Multi-UAV-Enabled Mobile Edge-Cloud Computing Systems

Soha Alhelaly 1, Ammar Muthanna 2 and Ibrahim A. Elgendy 3,*

1 College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia; s.alhelaly@seu.edu.sa
2 Department of Telecommunication Networks and Data Transmission, The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, 193232 Saint Petersburg, Russia; ammarenexpress@gmail.com
3 Department of Computer Science, Faculty of Computers and Information, Menoufia University, Shbin El Kom 32511, Egypt
* Correspondence: ibrahim.elgendy@ci.menofia.edu.eg

Abstract: With the emergence of various new Internet of Things (IoT) devices and the rapid increase in the number of users, enormous services and complex applications are growing rapidly. However, these services and applications are resource-intensive and data-hungry, requiring satisfactory quality-of-service (QoS) and network coverage density guarantees in sparsely populated areas, whereas the limited battery life and computing resources of IoT devices will inevitably become insufficient. Unmanned aerial vehicle (UAV)-enabled mobile edge computing (MEC) is one of the most promising solutions that ensures the stability and expansion of the network coverage area for these applications and provides them with computational capabilities. In this paper, computation offloading and resource allocation are jointly considered for multi-user multi-UAV-enabled mobile edge-cloud computing systems. First, we propose an efficient resource allocation and computation offloading model for a multi-user multi-UAV-enabled mobile edge-cloud computing system. Our proposed system is scalable and can support increases in network traffic without performance degradation. In addition, the network deploys multi-level mobile edge computing (MEC) technology to provide the computational capabilities at the edge of the radio access network (RAN). The core network is based on software-defined networking (SDN) technology to manage network traffic. Experimental results demonstrate that the proposed model can dramatically boost the system performance of the system in terms of time and energy.

Keywords: computation offloading; resource allocation; energy efficient; unmanned aerial vehicle (UAV); mobile edge computing; optimization

1. Introduction

By 2025, market revenue from unmanned aerial vehicles (UAVs) is expected to reach 52 billion dollars, a dramatic increase compared with today’s UAV market [1,2]. By then, UAVs will improve commercial, civil, and government sectors by introducing a wide range of heterogeneous applications [3,4]. These applications will vary from simple environmental monitoring to complex high-security military applications. Based on release 14 of the 3GPP and the technical reports of the ITU-R, UAVs are considered to be among the most important facets of the fifth-generation cellular system (5G/IMT2020), as well as beyond 5G and 6G networks [5–8]. UAVs will be actively used for various purposes, such as military tasks, remote sensing, remote localization, target detection, transportation monitoring, and route assessment [9–11].

Several related system architectures based on UAV-mobile edge computing (MEC) have recently been proposed, in which intensive computation tasks are offloaded from the mobile user to a richer resource located at the radio access network (RAN) edge [12–16]. Most of these architectures address only two layers of computation offloading in UAV-MEC systems [17,18].
although a few studies address multi-layer computation offloading [19]. Most of these architectures share the goal of minimizing energy consumption, allocating radio and computation resources efficiently, and/or meeting mobile user delay requirements.

However, in order to support the network in which UAVs operate, several features such as low power consumption and information processing must be considered. In addition, in a large-scale environment, the connected servers at the edge network are limited by computational capacity, which leads to high delays and makes these devices unsuitable for real-time tasks. Moreover, deriving the best offloading decision in a multi-user multi-UAV wireless environment is considered a non-trivial problem.

Confronting these challenges, we believe that constructing an energy-efficient model for a multi-user multi-UAV-enabled mobile edge-cloud computing system would be useful. Therefore, we proposed an efficient task offloading and allocation of resources model for a multi-user with multi-UAV-enabled mobile edge-cloud computing system. In addition, a combination model of resource allocation and offloading is formulated as a problem which aims to minimize the energy consumption of the whole system with a latency constraint. Moreover, a multi-UAV energy-efficient task offloading algorithm is designed to derive the task offloading decision for all device users. The main contributions of this paper include:

- An energy-efficient system model for multi-user multi-UAV-enabled mobile edge-cloud computing. This system is a scalable system and can support increased network traffic without performance degradation;
- A network of UAVs is deployed to cover the dead and high-density network areas. In addition, UAVs can provide other applications over the smart city network in which the multilevel MEC technology is deployed to provide both the computational capabilities at the edge network as well as a core network based on SDN technology to manage the network traffic and provide an innovative interface to the network operators;
- A combination model of resource allocation and offloading is formulated as a problem which aims to minimize the energy consumption of the whole system with a latency constraint;
- Simulations validate our proposed model and demonstrate that the system’s energy consumption can be significantly decreased.

The remainder of this paper is organized as follows. Related work on computation offloading approaches is presented in Section 2. In Section 3, the system model is presented, followed by the optimization problem formulation. Then, experimental simulations are evaluated in Section 4 to demonstrate the performance improvement of our system model. Finally, the paper’s conclusion is deduced in Section 5.

2. Related Work

In recent years, several system architectures and optimization models based on UAV-MEC have been developed to address the mobile users’ main challenges where the computation offloading concept is applied. Most of these approaches have only considered MEC networks with single [20,21] or multiple edge servers [22,23], although a few approaches have considered MEC networks with multi-edge servers and a centralized cloud [16,19,24]. This section provided a brief overview of popular approaches based on MEC networks with single and multi-edge servers with and without a centralized cloud.

2.1. Single Edge Server

You et al. [20] proposed an energy-efficient framework for wirelessly powered mobile cloud computing that integrates microwave power transfer (MPT) with mobile cloud computing. This framework considers only a single-user, single-edge system in which the mobile user can harvest energy from the edge or offload computation to it. Additionally, an optimization problem is formulated as a non-convex formula that aims to minimize the mobile user’s energy consumption for local computation harvesting while maximizing the energy savings from mobile users for computation offloading. Meanwhile, an efficient and online deep reinforcement learning-based offloading framework is proposed in [21].
Specifically, a deep neural network is used to learn and determine the offloading decision for mobile users such that they can either execute the computation tasks locally using mobile device resources or offload them to the edge for remote execution. In addition, an adaptive procedure is designed to reduce the complexity by automatically adjusting the hyper-parameters of the deep model. Finally, the numerical and experimental results show that this framework can provide a near-optimal solution.

Recently, an efficient algorithm based on deep meta reinforcement learning was proposed in [25] for IoT-edge-cloud computing systems with the goal of reducing the burden of computing and improving the processing of tasks. Specifically, a different number of DNNs are combined with Q-learning to derive an efficient offloading decision and increase the learning ability of the system. Furthermore, the results proved that the proposed algorithm outperforms binary and partial offloading schemes. Meanwhile, in [26], quality of experience (QoE) is optimized in an edge-enabled Internet of Things environment in which transmission and computing delay, the success rate of each task, and energy consumption are considered. In addition, deep reinforcement learning is utilized to solve the proposed model, where double Q-learning and dueling networks-based algorithms are designed to accelerate the convergence and enhance the stability. Similarly, the computation capability limitations of IoT devices and energy consumption are addressed in [27], where an integrated framework is proposed for task offloading and allocating resources with the goal of maximizing energy efficiency while reducing the quality of service (QoS) constraints. However, one of the common drawbacks of [27] is that the transferred data is susceptible to different attacks.

2.2. Multiple Edge Servers

Minimizing the mobile devices’ energy consumption and reducing the mobile applications’ execution latency in a multi-user system are the main goals of [22], in which the computational speed and the power of transmission for mobile devices are jointly optimized with the offloading ratio. Specifically, the authors formulated energy consumption and latency execution as non-convex problems. Then, they used variable substitution and univariate search techniques to obtain the optimal solution. Their results demonstrate that the proposed model can significantly decrease energy consumption and shorten latency with respect to the existing offloading schemes. A multi-user with a multi-edge offloading model for static and dynamic channels is proposed in [23]. First, on the static channel, the computation offloading decision is formulated as a non-cooperative potential game in which each mobile device can selfishly maximize CPU cycles and minimize consumed energy. On the dynamic channel (i.e., under time-varying and unknown channel state information), a Q-learning-based distributed computation offloading algorithm is formulated to determine the optimal solution, where the Nash equilibrium can be achieved. Finally, numerical results indicated that not only can this model outperform the energy consumption performance over local processing and random assignment, but it can also improve average long-term payoffs by up to 87.87% in comparison with the perfect channel state information case.

Yang et al. [28] proposed an energy-efficient offloading technique for multitask unloading of data based on UAV clusters for virtual and augmented reality games on vehicle networks. The VR/AR hybrid gaming scene should allow drivers and passengers to enjoy various visual effects in real time. However, due to dynamic changes in the real-time distribution of users, fixed edge nodes cannot cope with the changing load. Therefore, the use of UAVs together with fixed nodes as edge nodes helps to cope with the computing requirements of automotive networks. The deployment of a large number of UAVs is not practical due to mutual interference and the high cost of UAVs. To solve these problems, the paper presents a system with the following characteristics: multitask unloading (UAVs can work on several tasks and the results of one task can be used to perform other tasks in the future), UAV clusters working on one task, computing optimization, caching and communication resources, and AI-based decision making. In addition, an architecture for UAV clusters with various task implementations is proposed in which all the large-
scale applications for VR/AR games of vehicle networks are considered. Furthermore, an AI-based decision-making system is proposed to facilitate collaboration between UAVs and joint computing optimization, caching, and communication resources. The proposed architecture, compared with traditional MEC structures (both with and without UAVs), has advantages in the amount of information collected, real time performance, decision-making ability, and efficiency. Two algorithms were considered: preliminary deployment based on the analysis of historical data, and real-time deployment based on real-time perception. In addition, in this work, an experiment was conducted to predict the load of several UAVs based on long short-term memory (LSTM).

Elsewhere, ref. [29] presented a multi-server task offloading model in a decentralized game manner with the goal of maximizing the system utility; a decentralized and efficient algorithm is proposed to derive the optimal offloading where the differential neural computing approaches and deep reinforcement learning models are combined without the need of prior knowledge of the user preference and network bandwidth. Meanwhile, in [30], unmanned ground vehicles (UGVs) are introduced as an effective solution to address the computing resources and battery limitations of UAVs, in which the intensive tasks can be offloaded from UAVs to be executed with UGVs. Moreover, a new stable matching algorithm is designed to transform the offloading problem into an equivalent two-sided matching form. Further, an iterative algorithm is introduced to match UAVs with the appropriate UGVs. Nevertheless, despite improvements recorded in [29,30], data transferred is not adequately protected from different types of attacks during transmission.

2.3. Multiple Edge Servers and Remote Cloud

Beijing et al. [19] proposed an energy-efficient edge-cloud framework for computation offloading of a multi-UAV system. They formulated an optimization problem as a three-layer game in which the offloading decision could be determined in a decentralized way. In addition, efficient decentralized algorithms are developed in which the performance of the game is analyzed through its efficiency ratio and the Nash equilibrium can be achieved. Finally, experimental results demonstrate that these algorithms can dramatically decrease the consumption of energy by 30% with less than 10% performance loss compared with other benchmark solutions. Liu et al. [24] presented an energy-efficient cloud-assisted edge computing framework for collaborative task computation offloading in IoT sensors. This framework was composed of a three-tier network with IoT sensors in the first tier and the edge server and the cloud server in the other two tiers. In addition, an offloading strategy was formulated as a mixed-integer problem in which minimizing the energy consumption of IoT sensors was the main goal. Furthermore, a semi-definite relaxation-based energy-efficient collaborative task offloading algorithm was developed to solve this problem and to obtain the optimum decision of offloading. Finally, the simulation results showed that this algorithm can outperform the existing algorithm in terms of energy consumption. In addition, it can efficiently adapt to different system parameters.

The standard MEC structure with fixed nodes is not able to satisfy the needs of users for computing resources and connections with rapidly changing characteristics of time, space, and spectrum. To solve this problem, ref. [31] discusses the use of an air-to-ground MEC-integrated environment, consisting of the deployment of UAVs and ground vehicles as additional edge nodes. The UAVs make it possible to ensure coverage of the network in complex geographical areas, and provide a wider selection of services. The proposed network consists of three levels: the level of cloud computing, the level of edge computing, and the level of devices. The integrated structure basically assumes the use of the main advantages provided by software-defined networking: centralization of control, separation of user data from control data, and a large abstraction of basic structures. The hypervisor in this network is responsible for the virtualization of the resources of user devices, as well as for the computing and storage resources of the upper level (relative to the device level). Variants of the proposed network structure correspond to the main scenarios in networks: support for ultra-low latency—URLLC, distributed processing and analysis of big data along with the caching of content and delivery to mobile devices—eMBB, and scalable
arrays and mobile communications management—mMTC. The benefits provided by the structure include adaptive resource allocation, differentiated QoS guarantees, enhanced connectivity between network nodes, and efficient mobility management.

The delay in transferring data from mobile devices to a remote computing cloud has a large load value in existing traffic offloading systems. To solve this problem, ref. [32] developed code for a traffic offloading system oriented to the network edge called Echo. The network structure for the Echo code consists of three levels: mobile devices, border, and cloud. Echo uses a centralized mechanism in which the decision to offload traffic is carried out at the edge nodes. The code for unloading traffic is implemented on Android devices. Echo solves the problem of which method and which platform are used—the cloud platform or the border. For more efficient use of the resources of the edge nodes, the Preemption-Constrained Shortest-Remaining Time-First (PC-SRTF) scheduling algorithm is used to minimize the average time for solving the problem. The proposed code provides a guarantee of the quality of service when unloading traffic at border nodes. To optimize the data transfer, lazy object transfer algorithms and differential object update algorithms are used. In the first algorithm, when traffic is unloaded, a proxy is created to transfer related objects. In the second method, only the parts of objects that differ when loading several methods on a mobile device are updated. The proposed Echo code was found to have lower average task execution time and power consumption for mobile devices compared to existing traffic offloading systems.

Recently, a multi-user with multi-UAV-enabled and cloud servers’ model is presented in [33,34]. Specifically, in [33], a new weighted cost model is proposed with the goal of minimizing the energy consumption and execution time. Additionally, an efficient placement algorithm is developed to assign the tasks to the most suitable server in an efficient manner. Moreover, a lightweight failure recovery algorithm is designed to optimize the parallel execution for tasks. Xia et al. [34], however, proposed an intelligent task offloading model for a multi-tier MEC network, in which minimizing the execution delay of tasks and meeting the UAVs’ computing capacity are the main goals. In addition, an equivalent and convex form of the formulated problem is given using two-layer optimization technique. In the first layer, differential evolution learning approach is utilized, while a distributed deep neural network is used in the other layer. However, measuring the execution time is not considered in [34], which is one of the major downsides.

Due to the high dynamics in urban networks, which lead to an overload of the connection between the UAV and the boundary server (which also leads to an increase in the time spent on tasks), and due to the large queues that occur on the boundary servers when processing several tasks, the total time of n-th task transfer to the boundary server and the task processing time may exceed the time spent analyzing the task locally on the UAV. Here, we present an optimization process that allows the user to choose where the task will be processed on a UAV. We consider the communication, computational, and energy aspects of the proposed process.

Table 1 introduces a summary of the related works reviewed in terms of number of users, edge, and remote cloud servers, and highlights the main weaknesses of the studies.
Table 1. Overview of studies on MEC.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Objective</th>
<th>Proposed Solution</th>
<th>User</th>
<th>Remote Cloud</th>
<th>Evaluation Methods</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>[20]</td>
<td>Minimize energy consumption</td>
<td>An energy-efficient framework for wireless-powered mobile cloud computing</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Simulation</td>
</tr>
<tr>
<td>[21]</td>
<td>Maximize weighted sum of computation rate</td>
<td>An online deep reinforcement learning framework for mobile edge computing</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Execution time is not considered.</td>
</tr>
<tr>
<td>[25]</td>
<td>Minimize time and energy</td>
<td>An efficient algorithm based on deep meta reinforcement learning</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Deadline for computation tasks is not considered.</td>
</tr>
<tr>
<td>[26]</td>
<td>Optimize quality of experience</td>
<td>A new quality of experience model</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Resource cooperation is not considered.</td>
</tr>
<tr>
<td>[27]</td>
<td>Maximize energy efficiency under minimum quality of service constraint</td>
<td>An integrated model is proposed for task offloading and allocating resources</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Private data protection and the resource cooperation are not addressed.</td>
</tr>
<tr>
<td>[22]</td>
<td>Minimize energy and execution latency</td>
<td>A partial task offloading model</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Deadline for computation tasks is not considered and considering only one IoT user in the model is unrealistic.</td>
</tr>
<tr>
<td>[23]</td>
<td>Reduce energy consumption</td>
<td>A multi-user with a multi-edge task offloading model</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Considering only one IoT user in the model is unrealistic.</td>
</tr>
<tr>
<td>[28]</td>
<td>Maximize system utility</td>
<td>An energy-efficient task offloading technique</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Deadline for computation tasks and data privacy are not considered.</td>
</tr>
<tr>
<td>[29]</td>
<td>Maximize system utility</td>
<td>A multi-server task offloading model</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Data privacy is not addressed.</td>
</tr>
<tr>
<td>[19]</td>
<td>Minimize the overall energy consumption</td>
<td>An energy-efficient edge-cloud architecture for intelligent multi-UAV</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Resource cooperation is not considered.</td>
</tr>
<tr>
<td>[24]</td>
<td>Minimize energy consumption</td>
<td>An energy-efficient collaborative task offloading model for cloud-enabled edge computing network</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Data privacy is not addressed.</td>
</tr>
<tr>
<td>[31]</td>
<td>Minimize energy consumption</td>
<td>An air-ground integration technique for MEC</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Deadline for computation tasks and data privacy are not considered.</td>
</tr>
<tr>
<td>[32]</td>
<td>Minimize the energy consumption and execution time</td>
<td>An edge-centric code offloading model</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Data privacy is not addressed.</td>
</tr>
<tr>
<td>[33]</td>
<td>Minimize the processing delay of tasks</td>
<td>A new weighted cost model for multi-UAV environment</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Task execution’s deadline is not considered.</td>
</tr>
<tr>
<td>[34]</td>
<td>Minimize energy consumption with delay satisfaction</td>
<td>An efficient resource allocation and task offloading model</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Mobility and security are not addressed.</td>
</tr>
</tbody>
</table>
3. System Model

As shown in Figure 1, we present a multi-user with multi-UAV-enabled MEC system model. In this system, we consider $M$ as a set of small UAVs that are equipped with an edge computing server and connected to centralized cloud computing through a backbone router. In addition, we consider $N$ as a set of mobile devices that are associated with UAVs via a wireless channel. Furthermore, each mobile device is involved in highly intensive computation tasks that need to be processed locally or be transmitted and processed by the nearby UAV or cloud server via the wireless channel. In our simulation settings, we assumed a quasi-static scenario where the UAV does not move during the computation offloading period, though it can be moved to another location across different periods [35–37].

![System model](image)

**Figure 1.** System model.

In the next subsections, the communication and computation models are firstly discussed in more detail. Afterward, our problem formulation is presented.

Table 2 summarizes the main notations of this study.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>The set of small UAVs.</td>
</tr>
<tr>
<td>$N$</td>
<td>The set of device users.</td>
</tr>
<tr>
<td>$i$</td>
<td>The $i$th UAV.</td>
</tr>
<tr>
<td>$j$</td>
<td>The $j$th device user.</td>
</tr>
<tr>
<td>$b_{ij}$</td>
<td>Size of data for task of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$c_{ij}$</td>
<td>CPU cycles to accomplish the task of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$\tau_{ij}$</td>
<td>Deadline requirement for task of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$\phi_{ij,a}$</td>
<td>Offloading decision of mobile device.</td>
</tr>
<tr>
<td>$r_{ij}$</td>
<td>Uplink data rate for mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$B_i$</td>
<td>System bandwidth of UAV $i$.</td>
</tr>
<tr>
<td>$p_{ij}$</td>
<td>Transmission power of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$G_i$</td>
<td>Channel gain of UAV $i$.</td>
</tr>
</tbody>
</table>
where

The link data rate for downloading result from server is larger than the uploading [40–42].

TDMA technology is utilized to mitigate the co-channel interference between device users.

Table 2. Cont.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_i$</td>
<td>Density of noise power of channel which is connected with UAV $i$</td>
</tr>
<tr>
<td>$\zeta_{ij}$</td>
<td>Energy consumed per CPU cycle of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Propagation delay between UAV and cloud</td>
</tr>
<tr>
<td>$f_{ij}$</td>
<td>Computational capability of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$f'_{ij}$</td>
<td>CPU resource of UAV $i$ assigned to mobile device $j$</td>
</tr>
<tr>
<td>$f''_{ij}$</td>
<td>CPU resource of cloud assigned to mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$T_{ij}$</td>
<td>Time for local execution for the task of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$E_{ij}$</td>
<td>Energy for local executing task of mobile device $j$ which is connected with UAV $i$.</td>
</tr>
<tr>
<td>$T_{ij}^{off}$</td>
<td>Time for transmitting task of mobile device $j$ to UAV $i$.</td>
</tr>
<tr>
<td>$T_{ij}^{exec}$</td>
<td>Time for executing task of mobile device $j$ at UAV $i$.</td>
</tr>
<tr>
<td>$T_{ij}^{exec,off}$</td>
<td>Time for executing task of mobile device $j$ at Cloud server.</td>
</tr>
</tbody>
</table>

3.1. Communication Model

We first present the model of communication for edge-cloud systems, where the environment has a number $\mathcal{M}$ of small UAVs that are connected to the centralized cloud computing via a backbone router. In addition, each UAV is connected with a number $\mathcal{N}$ of mobile devices that have highly intensive computation tasks that should be completed. We refer to the small UAVs’ set and the mobile devices’ set as $\mathcal{M} = \{1, 2, \ldots, M\}$, and $\mathcal{N} = \{1, 2, \ldots, N\}$, respectively.

We denote $\phi_{ij,a}$ as the decision offloading where the task of mobile device $j$, which is connected with UAV $i$, is assigned to be executed at the server $a$, which can be a local execution (i.e., $a = 0$), cloud execution (i.e., $a = M + 1$), or UAVs execution (i.e., $a \in [1..M]$). In particular, $(\phi_{ij,0} = 1$, where $a \in [1..M]$) denotes that the computation task of device $j$, which is connected to UAV $i$, will be transmitted and executed remotely at one of the available UAVs, while $(\phi_{ij,M+1} = 1$) and $(\phi_{ij,M+1} = 1$) denote that the task of mobile device $j$ will be processed locally by itself or on the cloud server, respectively. Overall, each mobile device $j$ must execute its computation task only one time, whether locally $(a = 0)$ or remotely $(a \in \{1, 2, \ldots, M + 1\})$, i.e., $\sum_{a=0}^{M+1} \phi_{ij,a} = 1$.

Following Shannon’s law, the maximum uplink data rate between each mobile device $j$ and the connected UAV to transmit its task’s data through the communications channel can be expressed as [38]:

$$r_{ij} = B_i \log_2 (1 + \frac{p_{ij} g_i^2}{\omega_i B_i})$$

where $B_i$ denotes the channel bandwidth between UAV $i$ and its connected mobile device, $p_{ij}$ and $g_i$ denote the transmission power of mobile device $j$, the power density, and the gain of the channel between the mobile device and UAV. Guided by the intuition in [39], TDMA technology is utilized to mitigate the co-channel interference between device users.

In this paper, the total overhead consumption in terms of energy and time for returning the output data size is neglected due to its size being smaller than the input data size. The link data rate for downloading result from server is larger than the uploading [40–42].

3.2. Computation Model

This subsection will introduce the offloading model where the simulation has a number $\mathcal{M}$ of small UAVs that are connected to the centralized cloud computing via a backbone router. In addition, each UAV is connected to a number $\mathcal{N}$ of mobile devices that have highly intensive computation tasks that should be finished. For each mobile device $j$, the tuple $\{b_{ij}, c_{ij}, \tau_{ij}\}$ represents the task requirement, where $b_{ij}$ indicates the size of the dataset that needs to be transmitted from mobile device $j$ to UAV $i$, whereas $c_{ij}$ and $\tau_{ij}$ denote the CPU cycles’ number and the required deadline for the task of mobile device $j$. The values of $b_{ij}$ and $c_{ij}$ can be acquired via task execution profiling [40,43].
3.2.1. Local Execution

For the local execution method, the computation task of mobile device \( j \) will be executed locally and the total time execution and consumption of energy can be expressed as, respectively:

\[
T_{lj} = \frac{c_{lj}}{f_{lj}} \\
E_{lj} = \zeta_{lj} c_{lj},
\]

where \( f_{lj} \) indicates the capability of computation (CPU/cycles per second) of the mobile device, and \( \zeta_{lj} \) is a coefficient that indicates the CPU cycle’s energy consumed for mobile device \( j \), which is connected with UAV \( i \). We set \( \zeta_{lj} = 10^{-11}(f_{lj})^2 \), in which the consumption of energy is a superlinear function of the mobile device’s frequency [40,44].

3.2.2. Remote Execution

For the remote execution method, the computation task of mobile device \( j \), which is connected with UAV \( i \), will be transmitted and executed on one of the UAVs or the cloud server, and the total execution time can be expressed as:

\[
T_{c} = T_{off} + T_{exec} \\
T_{e} = T_{j} + T_{exec},
\]

where \( \delta \) is a constant referring to the delay of propagation for transmitting the task between the UAV and the cloud. \( T_{off}, T_{exec} \), and \( T_{exec} \) indicate the transmitting and processing time for executing the task of mobile device \( j \) at the UAV and cloud server, respectively, which can be calculated as follows:

\[
T_{off} = \frac{b_{ij}}{r_{ij}} \\
T_{exec} = \frac{c_{ij}}{f_{ij}} \\
T_{exec} = \frac{c_{ij}}{f_{ij}},
\]

where \( f_{ij} \) and \( f_{ij} \) denote the capability of computation for the UAV and the cloud server that is allocated to mobile device \( j \). As a consequence, the consumption of energy for transmitting and processing the task of mobile device \( j \) remotely at the UAV and cloud server can be calculated as follows:

\[
E_{r} = p_{ij} T_{off},
\]

In view of the above models, the energy and time overhead for processing the computation task of mobile device \( j \) can be expressed as, respectively:

\[
E_{i,j} = \left[ \varphi_{i,j,0} E_{ij} + \sum_{a=1}^{M+1} \varphi_{i,j,a} E_{ij} \right], \\
T_{i,j} = \left[ \varphi_{i,j,0} T_{ij} + \varphi_{i,j,M+1} T_{ij} + \sum_{a=1}^{M} \varphi_{i,j,a} T_{ij} \right]
\]
3.3. Problem Formulation

This subsection will present the formulation of our optimization problem for which computation offloading and allocation of resources for a multi-user and multi-UAV-enabled mobile edge-cloud computing system are considered. The main issue of this problem is to decrease the consumption of energy for the system with a latency constraint. This formulation is as follows:

\[
\begin{align*}
\min_{\psi} & \quad \left[ \sum_{i=1}^{M} \sum_{j=1}^{N} E_{ij} \right] \\
\text{s.t} & \quad \left[ E_{ij} - E_{ij}^l \right] \leq 0, \quad \forall i, j \quad \text{C1} \\
& \quad T_{ij} \leq \tau_{ij}, \quad \forall i, j \quad \text{C2} \\
& \quad \sum_{j=1}^{N} \sum_{a=1}^{M+1} \varphi_{ij,a} r_{ij} \leq B_i, \quad \forall i \in \{1..M\} \quad \text{C3} \\
& \quad \sum_{j=1}^{N} \sum_{a=1}^{M} \varphi_{ij,a} f_{ij} \leq F_i, \quad \forall i \in \{1..M\} \quad \text{C4} \\
& \quad \sum_{a=0}^{M+1} \varphi_{ij,a} = 1, \quad \forall i \quad \text{C5} \\
& \quad \varphi_{ij,a} \in \{0, 1\}, \quad \forall i, j, a \quad \text{C6}
\end{align*}
\]

Constraint C1 is the upper bound of energy consumption. Constraint C2 controls the requirement deadline for each task. Constraint C3 manages the capacity of bandwidth. Constraint C4 describes the edge server’s capacity upper limit. Constraint C5 ensures that each task must be processed only once. Lastly, constraint C6 ensures that the variable of decision offloading is binary.

The offloading decision can be derived by solving this problem. However, this problem is classified as NP-hard, in which the offloading decision variables are binary [45]. Therefore, to solve this problem in an efficient manner, a binary relaxation approach is utilized. Specifically, the binary variables are relaxed to new real variables, such as \( \geq \varphi_{ij,a} \leq 1 \) [46]. Therefore, the formulation of the problem after relaxing the binary variables is shown in Equation (13).

\[
\begin{align*}
\min_{\psi} & \quad \left[ \sum_{i=1}^{M} \sum_{j=1}^{N} E_{ij} \right] \\
\text{s.t} & \quad \left[ E_{ij} - E_{ij}^l \right] \leq 0, \quad \forall i, j \quad \text{C1} \\
& \quad T_{ij} \leq \tau_{ij}, \quad \forall i, j \quad \text{C2} \\
& \quad \sum_{j=1}^{N} \sum_{a=1}^{M+1} \varphi_{ij,a} r_{ij} \leq B_i, \quad \forall i \in \{1..M\} \quad \text{C3} \\
& \quad \sum_{j=1}^{N} \sum_{a=1}^{M} \varphi_{ij,a} f_{ij} \leq F_i, \quad \forall i \in \{1..M\} \quad \text{C4} \\
& \quad \sum_{a=0}^{M+1} \varphi_{ij,a} = 1, \quad \forall i \quad \text{C5} \\
& \quad \varphi_{ij,a} \in [0, 1], \quad \forall i, j, a \quad \text{C6}
\end{align*}
\]

Since the problem is considered as a linear problem, where the objective function and the constraints are linear, the near-optimal solution can be derived using well-studied linear approaches.

3.4. A Multi-Tier Energy-Efficient Task Offloading Algorithm

In this section, a multi-tier energy-efficient task offloading algorithm is presented in which the derivation of the near-optimum task offloading decision for multi-tier environments is outlined as follows (Algorithm 1).
Algorithm 1 Multi-tier energy-efficient computation offloading algorithm.

1: **Initialization**: Each device user $j$ initializes the offloading decision of its task with $\phi_{i,j,0} = 1, \forall i, j$
2: **for each** UAV $i$ and at given time slot $t$ **do**
3: **Upload** all the available computational capabilities and bandwidth to the backbone router.
4: **for each** device user $j$ **do**
5: **Upload** the required information for their task, $\{b_{i,j}, c_{i,j}, \tau_{i,j}, \zeta_{i,j}, p_{i,j}\}$, as well as their local capabilities $f_{i,j}$ to the backbone router.
6: **end for**
7: **end for**
8: **Compute** the available data rate $r_{i,j}$ for each device user regarding Equation (1).
9: **Solve** the formulated problem in Equation (12) and derive the offloading decision values $\phi_{i,j,a}$ for each task.
10: **Send** the offloading decision values $\alpha_{i,j,k}$ to each device user.

First, the offloading decision for all computation tasks of mobile devices are initialized with $\phi_{i,j,0} = 1$, which indicates the local execution. Then, the backbone router iterates over UAVs and gathers their available capabilities including storage and computational as well as the available bandwidth. Next, it receives the required information for the user’s task, including $\{b_{i,j}, c_{i,j}, \tau_{i,j}, \zeta_{i,j}, p_{i,j}\}$, as well as the local capabilities for each device $f_{i,j}$ through the connected UAV. Further, the maximum allowable data rate for each mobile device is calculated according to Equation (1). Afterwards, the task offloading decision values for each computation task are calculated and derived by solving the formulated problem in Equation (12), and then said values for each device user are calculated, in which decreasing the consumption of energy for the system with a latency constraint is the main goal.

Algorithm 1 describes the comprehensive steps of task offloading decision making, in which the computational complexity is $O(MN)$, where $M$ denotes the number of UAVs and $N$ denotes the total number mobile devices.

4. Simulation Results and Discussion

Firstly, the experimental setup will be introduced. Then, the simulation results will be discussed to assess the accomplishment of our model. Finally, the traffic management algorithm is presented.

4.1. Experimental Setup

Our simulation was conducted using a personal computer equipped with an Intel® Core(TM) i7-10750H CPU with a 2.6 GHz frequency and 16 GB RAM capacity running on the Windows 10 Home 64-bit platform and preinstalled with Python for development. A multi-user with a multi-UAV environment is considered in which an area of $100 \times 100$ m$^2$ with five UAVs cooperating and associated with a single cloud server through a core network that is also equipped with an edge computing server is assumed. The height for the UAVs is set at 20 m. In addition, we have 30 mobile device users distributed across UAVs, and each user device has an intensive task that needs to be completed. The size of data is distributed uniformly within a range of $[0, 10]$ MB, whereas the CPU cycles required to process each data bit are set to 1000 cycles/bit. The cloud server’s CPU computational capabilities are set to 20 GHz and for UAV servers is distributed uniformly within a range of $[3.5, 5]$ GHz, whereas the CPU capability of computation for each device is distributed
uniformly within a range of \( \{0.5, 0.6, \ldots, 1.0\} \). The energy consumption for each mobile device is uniformly distributed within a range of \( (0.20 \times 10^{-11}) \) J/cycle [47]. The remaining settings are summarized in Table 3. The optimization is solved using the GEKKO Python-based package [48].

Table 3. Simulation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of UAVs</td>
<td>5</td>
</tr>
<tr>
<td>Number of mobile devices</td>
<td>30</td>
</tr>
<tr>
<td>Height of UAVs</td>
<td>20 m</td>
</tr>
<tr>
<td>System bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Transmission power</td>
<td>0.1 Watt</td>
</tr>
<tr>
<td>Input data size</td>
<td>([0, 10]) MB Unif-Dist</td>
</tr>
<tr>
<td>CPU cycles to accomplish task</td>
<td>1000 cycles/bit</td>
</tr>
<tr>
<td>Computation capability of mobile device</td>
<td>({0.5, 0.6, \ldots, 1.0})</td>
</tr>
<tr>
<td>Energy consumption</td>
<td>((0, 20 \times 10^{-11})) J/cycle</td>
</tr>
<tr>
<td>UAVs server capability</td>
<td>([3.5, 5]) Unif-Dist GHz</td>
</tr>
<tr>
<td>Cloud server capability</td>
<td>20 GHz</td>
</tr>
<tr>
<td>Cloud propagation delay</td>
<td>15 ms</td>
</tr>
</tbody>
</table>

4.2. Experimental Results & Discussion

This subsection evaluates the performance result of our model in comparison with four scenarios:

- **Local Execution**: No offloading exists. The computation tasks of all mobile devices will be processed locally \( \varphi_{i,j,0} = 1 \);
- **UAVs Server Execution**: The computation tasks of all mobile devices will be processed remotely at UAVs servers \( \sum_{a=1}^{M} \varphi_{i,j,a} = 1 \);
- **Cloud Execution**: The computation tasks of all mobile devices will be processed remotely at the cloud server \( \varphi_{i,j,M+1} = 1 \);
- **Computation Offloading Execution in [39]**: The computation tasks of all mobile devices will be processed at one of the available servers, including the local execution based on the model proposed in [39].

First, the consumption of energy for executing the tasks for the mentioned scenarios in relation to different users is shown in Figure 2. As shown, our proposed model achieves the best result compared to the other scenarios. Additionally, the edge and cloud execution scenarios exceed the scenario of local execution as the number of users increases. Moreover, the edge execution policy also exceeds the cloud execution with more than 140 device users. This result is due to increases in communication time when the communication channels are shared between users as well as the higher energy consumption during the offloading process. In addition, the UAV’s server resources are not enough to handle all the device users where the cloud server is more resourceful, thereby becoming better than UAV’s execution for a large number of users. Similarly, Figure 3 shows the response time for executing the computation tasks for different numbers of users. This figure shows that the response time for our proposed model is close to or equal to the response time for UAVs, cloud, and model executions in [39], while being less than the local execution policy. However, as the number of mobile users grows, the performance of UAVs and cloud execution policies improves in comparison to local execution, while our model still outperforms the local policy and outperforms the model in [39]. Further, the cloud execution policy outperforms the UAV’s server execution. This is traced to the communication channels shared among users being overloaded, thereby increasing the time, whereas the proposed model can be intelligently adapted to environmental change.
Second, the consumption of energy for executing the tasks for the mentioned scenarios in relation to different data sizes is shown in Figure 4. From this plot, we can deduce that the energy consumption of our proposed model, for data sizes of less than 20 MB, is close to the energy consumption of the UAVs, cloud, and model in [39] scenarios, whereas it is less than the energy consumption of the local execution scenario. Nevertheless, with the increase in data size, the UAVs server and cloud scenarios’ performance declines compared with the scenario of local execution, whereas the performance of the proposed model remains lower than that of local policies and the policy of the model in [39]. Moreover, the cloud execution policy outperforms UAV’s execution policy with large data sizes (i.e., more than 45 MB). Furthermore, the response time for processing the computation tasks with respect to different data sizes is shown in Figure 5. This figure shows that the time gradually increases as the data size increases. In addition, there is a significant gap between
our model and the other policies, which increases as the data size increases. In addition, the UAV’s server executions policy nearly reaches the cloud execution at a data size of 45 MB, and will then exceed this for larger data sizes. This variation is explained as follows. Increasing data sizes cause an increase in communication time, which then results in increased energy consumption. Nevertheless, our model can be adapted to process the tasks at best locations, UAVs, cloud, or locally, thereby optimizing the energy and time.

![Figure 4. Energy consumption in relation to different values of input data size.](image)

![Figure 5. Response time in relation to different values of input data size.](image)
Finally, the consumption of energy and response time for processing the tasks in relation to different numbers of UAVs is presented in Figures 6 and 7, respectively. It is shown from these two figures that the local and cloud execution scenarios are not affected by the number of UAVs, whereas the energy consumption and response time of the other three scenarios (i.e., UAVs server, model in [39], and proposed model) gradually decrease as the number of UAVs increases. In addition, our model can achieve a lower energy consumption and response relative to the other two scenarios. This is because the costs in terms of energy consumption and response time are decreasing as the mobile users are allocated more resources (i.e., by increasing the number of UAVs), whereas the local execution and cloud execution scenarios do not depend on UAV resources.

![Figure 6. Energy consumption in relation to different number of UAVs.](image1)

![Figure 7. Response time in relation to different number of UAVs.](image2)
5. Conclusions

Our study proposed an energy-efficient system model for a multi-user with a multi-UAV-enabled ME-cloud computing system. This system is scalable and can support increases in network traffic without performance decreases. In addition, a network of UAVs is deployed to cover the dead and high-density network areas. Furthermore, UAVs can provide other applications in a smart city network in which the MEC technology is deployed in multiple levels to provide the capabilities of computation at the edge of the network. The core network is based on SDN technology to manage network traffic and to provide an innovative interface for network operators. Finally, simulation experiments demonstrate that our proposed model can significantly decrease the energy consumption of the entire system.

One promising future direction is to utilize blockchain technology for securing and protecting the UAV network from different types of attacks. Moreover, a more general case where UAVs and device users can dynamically move within the period of offloading will be considered. In such cases, deep learning techniques can be utilized and exploited to address the modeling and formulation of the problem as well as the solution in an efficient manner.

Author Contributions: Conceptualization, I.A.E. and A.M.; methodology, I.A.E.; software, A.M. and S.A.; validation, I.A.E and S.A.; formal analysis, A.M. and I.A.E; investigation, A.M.; resources, S.A.; writing—original draft preparation, I.A.E.; writing—review and editing, A.M. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at Saudi Electronic University for funding this research work through the project number (8117).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Saudi Electronic University for funding this research work through the project number (8117).

Conflicts of Interest: The authors declare no conflict of interest.

References


34. Xia, J.; Wang, P.; Li, B.; Fei, Z. Intelligent task offloading and collaborative computation in multi-UAV-enabled mobile edge computing. *China Commun.* 2022, 19, 244–256. [CrossRef]


