

Article

Joint Resource Allocation Optimization in Space–Air–Ground Integrated Networks

Zhan Xu ^{1,2}, Qiangwei Yu ^{1,2} and Xiaolong Yang ^{1,2,*}

¹ Key Laboratory of Information and Communication Systems, Ministry of Information Industry, Beijing 100101, China; xuzhan@bistu.edu.cn (Z.X.); 2021020486@bistu.edu.cn (Q.Y.)

² Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing Information Science and Technology University, Beijing 100101, China

* Correspondence: xiaolongyang@bistu.edu.cn

Abstract: A UAV-assisted space–air–ground integrated network (SAGIN) can provide communication services for remote areas and disaster-stricken regions. However, the increasing types and numbers of ground terminals (GTs) have led to the explosive growth of communication data volume, which is far from meeting the communication needs of ground users. We propose a mobile edge network model that consists of three tiers: satellites, UAVs, and GTs. In this model, UAVs and satellites deploy edge servers to deliver services to GTs. GTs with limited computing capabilities can upload computation tasks to UAVs or satellites for processing. Specifically, we optimize association control, bandwidth allocation, computation task allocation, caching decisions, and the UAV's position to minimize task latency. However, the proposed joint optimization problem is complex, and it is difficult to solve. Hence, we utilize Block Coordinate Descent (BCD) and introduce auxiliary variables to decompose the original problem into different subproblems. These subproblems are then solved using the McCormick envelope theory, the Successive Convex Approximation (SCA) method, and convex optimization techniques. The simulation results extensively illustrate that the proposed solution dramatically decreases the overall latency when compared with alternative benchmark schemes.

Keywords: edge cache and computing; low latency; convex optimization; resource allocation; space–air–ground integrated networks



Citation: Xu, Z.; Yu, Q.; Yang, X. Joint Resource Allocation Optimization in Space–Air–Ground Integrated Networks. *Drones* **2024**, *8*, 157. <https://doi.org/10.3390/drones8040157>

Academic Editor: Carlos Tavares Calafate

Received: 26 February 2024

Revised: 13 April 2024

Accepted: 15 April 2024

Published: 18 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid advancement of mobile communication has led to various emerging applications that have increased convenience in social life and production [1] and resulted in substantial changes [2]. However, these novel services impose more stringent requirements on communication networks regarding latency, reliability, and speed [3]. Simultaneously, the astronomical growth in data communication volume and energy consumption for computation and transmission put a heavy load on terminal devices with limited computing capabilities [4]. Therefore, the explosive growth of data traffic and network applications presented significant challenges for the future development of mobile communication technology [5]. Simultaneously, ground-based networks suffer challenges in providing coverage to remote areas [6]; in many prospective wireless communication situations, UAVs play a crucial role in meeting various communication needs. Moreover, UAVs can be integrated with low-earth orbit satellites for diverse domains [7], which can provide information services for different network applications in different spatial domains [8]. Therefore, building a SAGIN to enable information from different spatial domains to serve various industries will be a future trend [9].

Recently, UAV-assisted MEC has been a focal point of research and is widely applied in various scenarios [10]. In [11], the authors optimized key variables jointly between multiple drones and ground-based stations to decrease the cost of energy consumption and latency, which improves communication quality. Equipping UAVs with MEC servers enables the

migration of ground-based computing tasks to drones. The work in [12] formulated a joint optimization problem with the limit of energy and latency constraints to minimize the energy consumption of ground user terminals. The authors in [13] utilized drones as relays, presented a dynamic NOMA/OMA strategy, and presented optimization problems to improve the minimum speed between vehicles and the total transmission rate. The security of the network is affected by various factors; the authors in [14] optimized the flight trajectory and power of the UAV and proposed an inequality iteration algorithm to improve the average secrecy rate of the network. Within the UAV-based MEC system network, UAVs can function as small-cell base stations to achieve scalable and secure video transmission through caching, allowing for the provision of services to mobile users [15]. To balance the crucial metrics and the electrical energy consumption of the MEC system, the authors in [16] proposed a problem with the objective of maximizing weighted computational effectiveness. To address the heavy computational workload of UAVs, in [17], the authors employed game theory principles to enhance the computational offloading efficiency of UAVs by balancing energy consumption, latency, and cost. In [18], a data offloading framework is proposed wherein users can offload partial computational tasks to the MEC servers on UAVs, and the distinctiveness of the Pure Nash Equilibrium (PNE) is demonstrated through simulations. The authors in [19] studied a collaborative system with multiple UAVs, and it can utilize a multi-agent deep learning framework to tackle decreasing system latency and energy consumption.

UAV-assisted MEC systems can be integrated with satellites to become SAGIN, providing edge computing services to GTs in remote regions [20]. The work in [21] presented a SAG hybrid cloud-edge computing framework. By jointly optimizing task assignment, transmission power, bandwidth allocation, and UAV computing resources, it achieves the minimization of the maximum computation latency among GTs. MEC services are frequently deployed in SAGIN networks; the authors in [22] presented an iterative optimization approach and utilized a greedy algorithm and the SCA method to minimize the total computing cost of all GTs. Due to the limited battery capacity, the authors in [23] decreased energy consumption by enhancing the offloading ratios and allocation of processing resources for terrestrial users, UAVs, and satellites. While ensuring energy constraints, Ref. [24] optimized offloading decisions to minimize network latency and proposed the BSUM algorithm for simulation verification. The network topology exhibits dynamics, and the authors in [25] presented an algorithm based on ADMM to get the optimal solution for network slicing problems. In [26], the authors proposed an I-SAT network (integrated satellite-aerial-terrestrial network), aiming to enhance the average throughput among users by addressing the joint optimization problem of association variables, power allocation, and UAV trajectories. In [27], the authors presented an integrated SAG network scheme incorporating terahertz; the scheme leveraged the abundant bandwidth of the terahertz spectrum and optimized variables such as terahertz frequency allocation to minimize the energy consumption of the network system. The work in [28] proposed a resource management scheme integrating Software-Defined Networking, which enhanced economic efficiency by analyzing variables, including priority, latency, energy consumption, and service level agreements. Additionally, due to the limitations of spectrum resources and atmospheric effects on free-space optical communication, in [29], the authors investigated a hybrid radio frequency and FSO scheme in a SAGIN and verified its performance through simulations.

Although there have been many studies on resource allocation in MEC systems, optimizing caching decisions is usually merely considered in the two-tier networks of MEC, as it is difficult to optimize caching decisions and other variables in a three-tier network. So, there is a limited amount of work considering the optimization of caching decision variables within the framework of a SAGIN. The interaction between computation and communication in mobile edge networks has a substantial influence. The problem of resource allocation for improving network performance continues to be a subject that merits extensive investigation. Therefore, this paper proposes a three-tier mobile edge network model and, through the proposed algorithm, jointly optimizes association control,

bandwidth allocation, computation task allocation, caching decisions, and the UAV's position to minimize total task latency. The major contributions of this paper are as follows:

- (1) Considering the limited computing capabilities of ground users and aiming to improve the rationality of network resource allocation, we deploy edge servers on UAVs to provide caching content services. We formulate an original problem with the purpose of minimizing the total computation task delay.
- (2) We propose an iterative optimization algorithm based on BCD, decomposing the joint optimization problem into different subproblems. We consider optimizing cache decision variables in SAGIN and use the McCormick envelope theory to address the coupling of association variables and cache decision variables. Additionally, we optimize UAV positions using the SCA method and convex optimization techniques.
- (3) The experimental results demonstrate that the proposed method outperforms others, effectively enhancing the total computation task delay of the network system.

2. System Model

Figure 1 illustrates the SAGIN system, which comprises N GTs, M UAVs, and one low-Earth-orbit (LEO) satellite, where UAVs and satellite are equipped with edge servers. The UAVs can provide services to ground user terminals and satellite while also handling computation tasks from both ground and satellite sources. The sets of GTs and UAVs can be denoted by $\mathcal{N} = \{1, 2, \dots, N\}$ and $\mathcal{M} = \{1, 2, \dots, M\}$, respectively. The computation task of the i -th GT in the current time slot is denoted by I_i . When users generate demands, computation tasks will be generated, such as weather data processing, decoding data packets, handling network protocols, etc. In general, computation tasks typically involve data transmission. The computation tasks generated by ground terminals and satellite will be transmitted, and the total computation task delay consists of transmission and computation delays. The triplet is marked by $\langle F_i, D_i, T_i \rangle$, where

- (1) F_i is defined as the computational complexity, representing the CPU resources required to compute one bit of data.
- (2) D_i denotes the size of the computation task. Note that the computation task can be partitioned.
- (3) T_i is the maximum tolerable delay of the task.

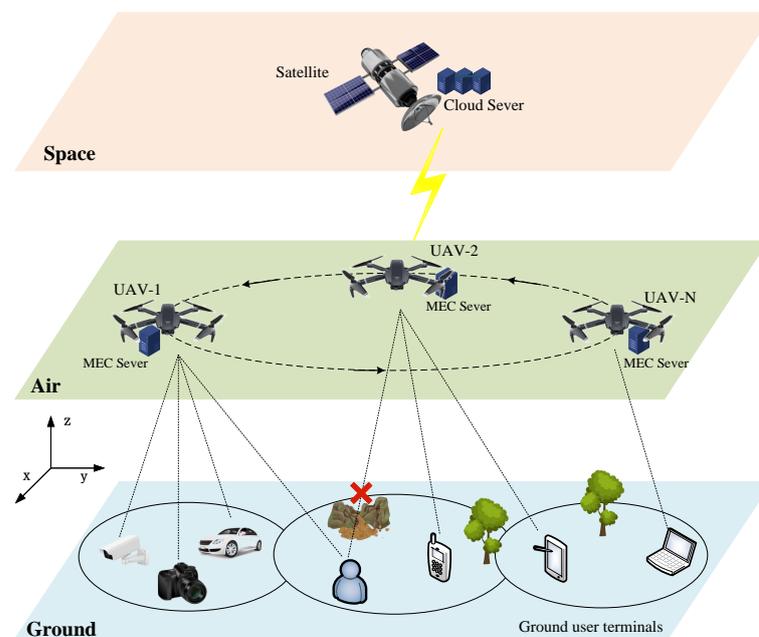


Figure 1. Space–air–ground integrated network architecture.

The overall data volume of computational task D_i can be separated into two distinct sections, denoted by D_i^l and D_i^s ; i.e.,

$$D_i = D_i^l + D_i^s \quad (1)$$

where D_i^l denotes the data produced by ground user terminals, and D_i^s represents the data generated by satellites, which can be transmitted to UAVs and GTs for processing.

Additionally, the data generated by ground user terminals can be divided into three parts, which can be processed on the GTs, UAVs, and satellites; i.e.,

$$D_i^l = D_i^1 + D_i^2 + D_i^3 \quad (2)$$

Each UAV is outfitted with a server, allowing each UAV to offer services for GTs within its service area. The ground terminals can either process tasks locally or offload computation tasks to UAV-based MEC servers or satellite cloud servers. The essential notations are shown in Table 1.

Table 1. Description of important symbols.

Symbols	Description
N	Number of GTs
M	Number of UAVs
I_i	Computation task of the i -th GT device
D_i^l, D_i^s	Computation task I_i generated by local GT and satellite
T_i^l, T_i^u, T_i^s	Computation delay of I_i at local GT device, UAV, and satellite
$h_{i,j}$	Channel coefficient from i -th device to j -th UAV
\mathbf{u}_j^U	Horizontal coordinates of the j -th UAV
$a_{i,j}$	Association variable from i -th device to j -th UAV
E_i^l, E_i^u, E_i^s	Energy consumption of local GTs, UAV, and satellite in processing computational task I_i
f_i^l, f_i^u, f_i^s	Computation capacity of local GTs, UAV, and satellite in processing computational task I_i
$R_{l,u}^U, R_{u,s}^U$	Uplink transmission rate from i -th device to j -th UAV, and from j -th UAV to satellite
$R_{u,l}^D, R_{s,u}^D$	Downlink transmission rate from j -th UAV to i -th device, and from satellite to j -th UAV

2.1. Channel Model

Consider representing the coordinates of ground terminals and UAVs in a three-dimensional Cartesian coordinate system, denoted by $\mathbf{u}_i^{GT} = (x_i^{GT}, y_i^{GT}, 0)$ and $\mathbf{u}_j^{UAV} = (x_j^{UAV}, y_j^{UAV}, H)$, respectively. Here, $\|\cdot\|^2$ represents the square distance between the position vectors of ground user terminals and UAVs' positions. Thus, the distance between the two can be expressed in the form that follows.

$$\begin{aligned} d_{i,j} &= \sqrt{H^2 + (x_i^{GT} - x_j^{UAV})^2 + (y_i^{GT} - y_j^{UAV})^2} \\ &= \sqrt{H^2 + \|\mathbf{u}_i^{GT} - \mathbf{u}_j^{UAV}\|^2} \end{aligned} \quad (3)$$

According to [18], the channel coefficient can be represented as follows.

$$h_{i,j} = h_0 d_{i,j}^{-\alpha} \hat{h}_{i,j} \quad (4)$$

where $\hat{h}_{i,j}$ denotes the small-scale fading, h_0 is the reference channel gain at the distance $d_0 = 1$, and α represents the path-loss exponent. According to [30], assuming that $\alpha = 2$, the small-scale fading can be represented as follows:

$$\hat{h}_{i,j} = \sqrt{\frac{\lambda}{\lambda + 1} \bar{h}_{i,j} + \frac{1}{\lambda + 1} \tilde{h}_{i,j}} \quad (5)$$

Note that the Rician fading factor is denoted by λ , and it is represented by $\bar{h}_{i,j}$ in the Line of Sight (LoS) scenario, where $|\bar{h}_{i,j}| = 1$. In the Non-Line of Sight (NLoS) situation, it is denoted by $\tilde{h}_{i,j}$, where $\tilde{h}_{i,j} \sim \text{CN}(0, 1)$.

The ground terminals utilize Frequency Division Multiple Access (FDMA) to offload computation tasks to UAVs to prevent co-frequency interference. When the i -th GT offloads its tasks to the j -th UAV, the achievable uplink data rate for the GT will be

$$R_{l,u}^U = b_{i,j} B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j} B_U N_0} \right) \quad (6)$$

where B_U is the available bandwidth in the uplink, $b_{i,j}$ represents the allocated system bandwidth, P_i is the transmission power for the i -th GT in the uplink, and N_0 is the noise power. Where P_j is the transmission power of the j -th UAV, the downlink data rate for GTs will be

$$R_{u,l}^D = B_D \log \left(1 + \frac{P_j |\hat{h}_{i,j}|^2 \Delta_{i,j}}{B_D N_0} \right) \quad (7)$$

where B_D is the available bandwidth in the downlink.

2.2. Computation Model

We define a_{ij} as the association variable between the i -th GT and the j -th UAV. In particular, each GT can be associated with, at most, one UAV; i.e.,

$$a_{ij} \in \{0, 1\}, \forall i \in \mathcal{N}, j \in \mathcal{M} \quad (8)$$

where c_j^u represents the cache variable, and $c_j^u = 1$ indicates that the data has been cached on the UAV, and $c_j^u = 0$ otherwise.

$$c_j^u \in \{0, 1\}, \forall j \in \mathcal{M} \quad (9)$$

Considering the cache variable in the latency, in some scenarios, there are more computation tasks for the satellites and fewer tasks for ground user terminals. So, the paper temporarily considers caching variables only in the downlink. In addition, the satellite request latency can be ignored in certain situations, such as autonomously executed tasks or periodic data collection. The computation capability of a GT is denoted by f_i^l . When the GT processes task I_i locally, the delay consists of two parts: transmission delay and computation delay. It is essential to consider whether the UAV caches computation tasks from the satellite. The local computation delay can be defined as

$$T_i^l = \sum_{j=1}^M a_{i,j} \left(\frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} + \frac{D_i^s (1 - c_j^u)}{R_{u,l}^D} + \frac{D_i^s c_j^u}{R_{u,l}^D} + \frac{(D_i^l + D_i^s) E_i}{f_i^l} \right) \quad (10)$$

When the GT processes computation task I_i , the energy consumption can modeled as

$$E_i^l = k(f_i^l)^2 (D_i^1 + D_i^s) F_i \tag{11}$$

where k is a constant depending on the processor chip architecture.

When ground user terminals offload part of the computation task to a UAV, we can assume that the latency can be divided into three stages: (1) the transmission time from the GTs to the UAV, (2) the time it takes to process the task on the UAV server, and (3) the time it takes for the UAV to transmit the computation results back to the ground user terminals. According to [31], the processed computation results in a data volume much smaller than before processing; therefore, this part can be omitted. Therefore, the latency when processing computing tasks at the UAV can be expressed as

$$T_i^u = \sum_{j=1}^M a_{i,j} \left(\frac{D_i^2}{R_{l,u}^U} + \frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} + \frac{(D_i^2 + D_i^s) F_i}{f_i^u} \right) \tag{12}$$

where f_i^u represents the computing capability of the UAV. Therefore, the consumption of energy for transmission used during the process of offloading computation tasks to the j -th UAV can be expressed as

$$E_i^t = \sum_{i=1}^N a_{i,j} P_i \left(\frac{D_i^2}{R_{l,u}^U} \right) \tag{13}$$

The consumption of energy for carrying out computation tasks I_i by the UAV is given by

$$E_i^u = k(f_i^u)^2 (D_i^2 + D_i^s) F_i \tag{14}$$

We assumed that multiple UAVs hovered above the task area to provide relevant services for the GTs and neglected the propulsion energy of the UAVs for now [12,32]. However, except for the energy use of computing and transmission, the flying energy of the UAVs also needs to be considered [33].

$$P_j^{\text{hov}} = \frac{\zeta \sqrt{\zeta}}{\eta_j \sqrt{0.5 \pi q r^2 \rho}} \tag{15}$$

The thrust of the UAV's mass is represented by ζ , the power efficiency is denoted by η_j , the number of rotors is expressed by q , the rotor diameter is r , and the air density is ρ . Therefore, the energy consumption of the UAV hovering can be represented as

$$E_j^{\text{hov}} = P_j^{\text{hov}} T_i^u \tag{16}$$

So, the sum of the energy consumption of the UAV for computation data, transmitting data, and hovering is represented as

$$E_j^{\text{tot}} = E_j^{\text{hov}} + E_i^t + E_i^u \tag{17}$$

Similarly to [34], we consider fixed transmission rates for communication with the satellite. In addition, the delay in processing computing tasks in the satellite can be represented as

$$T_i^s = \sum_{j=1}^M a_{i,j} \left(\frac{D_i^3}{R_{l,u}^U} + \frac{D_i^3}{R_{u,s}^U} + \frac{(D_i^3 + D_i^s) F_i}{f_i^s} \right) \tag{18}$$

where f_i^s represents the computing capacity of the satellite, and P_u^s denotes the transmission power between the UAV and the satellite. Therefore, the energy consumption during the process of transmitting tasks from the UAV to the satellite is given by

$$E_i^s = \sum_{j=1}^M a_{i,j} P_u^s \left(\frac{D_i^3}{R_{l,u}^U} + \frac{D_i^3}{R_{u,s}^U} \right) \quad (19)$$

In summary, the total delay of computation tasks I_i during the processing can be expressed as

$$\begin{aligned} T_i^t &= T_i^l + T_i^u + T_i^s \\ &= \sum_{j=1}^M a_{i,j} \left(2 \frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} + 2 \frac{D_i^l}{R_{l,u}^U} + \frac{D_i^s}{R_{u,l}^D} + \frac{D_i^J}{R_{u,s}^U} \right. \\ &\quad \left. + \frac{(D_i^l + D_i^s) F_i}{f_i^l} + \frac{(D_i^2 + D_i^s) F_i}{f_i^u} + \frac{(D_i^3 + D_i^s) F_i}{f_i^s} \right) \end{aligned} \quad (20)$$

where the total energy consumption will be

$$E_i = E_i^l + E_i^{tot} + E_i^s \quad (21)$$

3. Problem Formulation and Optimization Solution

We focus on minimizing the total computation task delay while satisfying constraints on maximum energy consumption and latency. We formulated the association variable $\mathbf{a} = a_{i,j}$, the computation task $\mathbf{D} = \{D_i^l, D_i^s\}$, the bandwidth allocation $\mathbf{b} = b_{i,j}$, the caching decision variables $\mathbf{c} = c_j^u$, and the UAV coordinates $\mathbf{u}^U = \{H_j, x_j^{UAV}, y_j^{UAV}\}$. The formulated joint optimization problem is as follows:

$$\begin{aligned} &\min_{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{D}, \mathbf{u}^U} \sum_{i=1}^N \{T_i^t\} \\ &\text{s.t. C1: } T_i^t \leq T_i, \forall i \in \mathcal{N} \\ &\quad \text{C2: } E_i \leq E_i^{\max}, \forall i \in \mathcal{N} \\ &\quad \text{C3: } D_i^l + D_i^s = D_i, \forall i \in \mathcal{N} \\ &\quad \text{C4: } \sum_{i=1}^N a_{i,j} \leq 1, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ &\quad \text{C5: } \sum_{i=1}^N \sum_{j=1}^M a_{i,j} b_{i,j} \leq 1, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ &\quad \text{C6: } a_{i,j} \in \{0, 1\}, c_j^u \in \{0, 1\}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ &\quad \text{C7: } \sum_{i=1}^N c_j^u D_i^s \leq C^u, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ &\quad \text{C8: } \left\| (H_j, \mathbf{u}_j^{UAV}) - (H_m, \mathbf{u}_m^{UAV}) \right\|^2 \geq d_{\min}^2, \forall j \in \mathcal{M}, m \neq j \\ &\quad \text{C9: } H_{\min} \leq H_j \leq H_{\max} \\ &\quad \text{C10: } b_{i,j}, f_i^l, f_i^u, f_i^s, D_i^l, D_i^s \geq 0, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \end{aligned} \quad (22)$$

where T_i represents the maximum completion time for computation tasks, E_i^{\max} is the maximum available energy for the i -th GT, and C^u represents the cache capacity of each UAV not being able to exceed its maximum capacity. The constraint conditions in Equation (22) can be interpreted as follows: C1 represents the processing time of computation tasks not being able to exceed the maximum delay; C2 represents the energy consumption constraint for each GT; C3 implies the constraints on the allocation of computation tasks for each ground user terminal; C4 is the ability of each GT to communicate with no more than one UAV; C5 represents the constraints on bandwidth allocation ratios; C7 implies the limited

cache capacity of the UAV; C8 represents the the smallest distance between UAVs to avoid collisions; C9 denotes the constraints on the flight altitude.

We can observe that Equation (22) is non-convex, leading to difficulty obtaining a globally optimal solution. Based on the BCD method, we demonstrate an alternating algorithm that transforms Equation (22) into subproblems, including association control, task allocation, bandwidth allocation, and UAV position optimization, to obtain suboptimal solutions to problems with Equation (22).

3.1. Association Control and Caching Variables

Under a given $\{\mathbf{b}, \mathbf{D}, \mathbf{u}^U\}$, the association variable and the caching variable are tightly coupled in a multiplicative form, which makes the problem challenging to solve. We relax the integer variable $a_{i,j} \in \{0, 1\}, c_j^u \in \{0, 1\}$ to a continuous variable $0 \leq a_{i,j} \leq 1, 0 \leq c_j^u \leq 1$ and introduce a new variable $\mathbf{z} = \{z_{i,j}\}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}$. Therefore, the original problem can be reformulated as follows:

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{c}, \mathbf{z}} \sum_{i=1}^N \{T_i^t\} \\ \text{s.t. } z_{i,j} = (1 - c_j^u) a_{i,j}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ \text{C1 - C10} \end{aligned} \quad (23)$$

Then, according to the McCormick envelope theory, the non-convex constraint $z_{i,j} = (1 - c_j^u) a_{i,j}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}$ can be replaced by the McCormick convex relaxation condition, which can be expressed explicitly as

$$z_{i,j} \geq a_{i,j} - c_j^u, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \quad (24)$$

$$z_{i,j} \geq 0, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \quad (25)$$

$$z_{i,j} \leq a_{i,j}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \quad (26)$$

$$z_{i,j} \leq 1 - c_j^u, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \quad (27)$$

Therefore, Equation (23) can be re-expressed as

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{c}, \mathbf{z}} \sum_{i=1}^N \{T_i^t\} \\ \text{s.t. C1 - C4, C6, C7} \\ 0 \leq a_{i,j} \leq 1, 0 \leq c_j^u \leq 1 \\ (24)-(27) \end{aligned} \quad (28)$$

By utilizing the McCormick envelope theory and convex relaxation techniques, the original problem can be transformed into a convex optimization problem, and we can find the optimal solution using standard optimization algorithms.

3.2. Computation Task Assignment

Given $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{u}^U\}$, the original problem can be reformulated as follows:

$$\begin{aligned}
& \min_{\mathbf{D}} \sum_{i=1}^N \{T_i^t\} \\
& \text{s.t.} \\
& \text{C1}^2 : T_i^l = \sum_{j=1}^N a_{i,j} \left(\frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} + \frac{D_i^s (1 - c_j^u)}{R_{u,l}^D} + \frac{D_i^s c_j^u}{R_{u,l}^D} + \frac{(D_i^l + D_i^s) F_i}{f_i^l} \right) \leq T_i, \forall i \in \mathcal{N} \\
& \text{C2}^2 : T_i^u = \sum_{j=1}^N a_{i,j} \left(\frac{D_i^2}{R_{l,u}^U} + \frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} + \frac{(D_i^2 + D_i^s) F_i}{f_i^u} \right) \leq T_i, \forall i \in \mathcal{N} \\
& \text{C3}^2 : T_i^s = \sum_{j=1}^N a_{i,j} \left(\frac{D_i^3}{R_{l,u}^U} + \frac{D_i^3}{R_{u,s}^U} + \frac{(D_i^3 + D_i^s) F_i}{f_i^s} \right) \leq T_i, \forall i \in \mathcal{N} \\
& \text{C12}^2 : k(f_i^l)^2 (D_i^l + D_i^s) F_i + k(f_i^u)^2 (D_i^2 + D_i^s) F_i + \sum_{i=1}^N a_{i,j} P_i \left(\frac{D_i^2}{R_{l,u}^U} \right) + P_j^{\text{hov}} t_j^u \\
& \quad + \sum_{j=1}^M a_{i,j} P_u^s \left(\frac{D_i^3}{R_{l,u}^U} + \frac{D_i^3}{R_{u,s}^U} \right) \leq E_i^{\text{max}}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\
& \text{C1, C3, C10}
\end{aligned} \tag{29}$$

It can be concluded that Equation (29), the above expression, is non-convex and can be solved using various classical algorithms, such as interior-point methods.

3.3. Bandwidth Allocation

Under a given $\{\mathbf{a}, \mathbf{c}, \mathbf{D}, \mathbf{u}^U\}$, the original problem can be rewritten as follows:

$$\begin{aligned}
& \min_{\mathbf{b}} \sum_{i=1}^N \{T_i^t\} \\
& \text{s.t. C1}^3 : \frac{D_i^2}{T_i - \frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} - \frac{(D_i^2 + D_i^s) F_i}{f_i^u}} \leq b_{i,j} B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j} B_U N_0} \right), \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\
& \text{C2}^3 : \frac{P_j^{\text{hov}} D_i^2 + P_u^s D_i^3 + P_i D_i^2}{E_i^{\text{max}} - E_i^l - E_i^u - P_j^{\text{hov}} \left(\frac{D_i^s (1 - c_j^u)}{R_{s,u}^D} + \frac{(D_i^2 + D_i^s) F_i}{f_i^u} \right) - P_u^s \frac{D_i^3}{R_{u,s}^U}} \\
& \leq b_{i,j} B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j} B_U N_0} \right), \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\
& \text{C3}^3 : \sum_{i=1}^N b_{ij} \leq 1, \forall i \in \mathcal{N} \\
& \text{C10}
\end{aligned} \tag{30}$$

According to $T_i^u \leq T_i$ and $E_i \leq E_i^{\text{max}}$, C1^3 and C2^3 can be inferred. According to [35], the perspective operation preserves convex (concave) shapes. It can be seen that $b_{i,j} B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j} B_U N_0} \right)$ is a concave function of $\log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{N_0} \right)$. Hence, it can be proven that $b_{i,j} B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j} B_U N_0} \right)$ is a concave function of $b_{i,j}$; therefore, both C1^3 and C2^3 are convex. Combining other constraint conditions, we can conclude that Equation (30) is convex.

3.4. UAV Position

When $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{D}\}$ is given, the problem can be converted into the problem of the positions of UAVs.

$$\begin{aligned}
 & \min_{\mathbf{u}^U} \sum_{i=1}^N \{T_i^t\} \\
 \text{s.t. } & \text{C1}^4 : \frac{D_i^2}{T_i - \frac{D_i^s(1-c_j^u)}{R_{s,u}^D} - \frac{(D_i^2+D_i^s)F_i}{f_i^u}} \leq b_{i,j}B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_UN_0 \left(H_j^2 + \|\mathbf{u}_i^{GT} - \mathbf{u}_j^{UAV}\|^2 \right)} \right), \forall i \in \mathcal{N} \\
 & \text{C2}^4 : \frac{P_j^{\text{hov}}D_i^2 + P_u^sD_i^3 + P_iD_i^2}{E_i^{\text{max}} - E_i^l - E_i^u - P_j^{\text{hov}} \left(\frac{D_i^s(1-c_j^u)}{R_{s,u}^D} + \frac{(D_i^2+D_i^s)F_i}{f_i^u} \right) - P_u^s \frac{D_i^3}{R_{u,s}^U}} \\
 & \leq b_{i,j}B_UN_0 \left(H_j^2 + \|\mathbf{u}_i^{GT} - \mathbf{u}_j^{UAV}\|^2 \right), \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\
 & \text{C8, C9, C1}^2 - \text{C3}^2
 \end{aligned} \tag{31}$$

In order to address the non-convex constraints, the auxiliary variables $\varepsilon, \lambda, \eta$ and x_i are introduced. We define that $H_j^2 + \|\mathbf{u}_i^l - \mathbf{u}_j^U\|^2 \leq x_i$, $L_i^c \leq \varepsilon$, $R_{t,u}^U \geq \lambda$, and $R_{u,l}^D \geq \eta$. Furthermore, C1^4 can be reformulated as follows:

$$\frac{D_i^2}{T_i - \frac{D_i^s(1-c_j^u)}{R_{s,u}^D} - \frac{(D_i^2+D_i^s)F_i}{f_i^u}} \leq b_{i,j}B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_UN_0 x_i} \right) \tag{32}$$

Based on the SCA approach, we perform a first-order Taylor expansion at the local point \hat{x}_i . The Equation (32) can be inferred as follows:

$$b_{i,j}B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_UN_0 x_i} \right) \geq b_{i,j}B_U \log \left(x_i + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_UN_0} \right) - b_{i,j}B_U \log(\hat{x}_i) - \frac{b_{i,j}B_U}{\hat{x}_i} (x_i - \hat{x}_i) \tag{33}$$

Next, we need to address the non-convex constraint C8. By giving the local points $(\hat{H}_j, \hat{\mathbf{u}}_j^{UAV})$ and $(\hat{H}_m, \hat{\mathbf{u}}_m^{UAV})$, it can be inferred that the lower bound is on the left-hand side of C9. The formula can be expressed as Equation (34):

$$\begin{aligned}
 & \left\| (H_j, \mathbf{u}_j^{UAV}) - (H_m, \mathbf{u}_m^{UAV}) \right\|^2 \geq - \left\| (\hat{H}_j, \hat{\mathbf{u}}_j^{UAV}) - (\hat{H}_m, \hat{\mathbf{u}}_m^{UAV}) \right\|^2 + \\
 & 2 \left((\hat{H}_j, \hat{\mathbf{u}}_j^{UAV}) - (\hat{H}_m, \hat{\mathbf{u}}_m^{UAV}) \right)^T \left((H_j, \mathbf{u}_j^{UAV}) - (H_m, \mathbf{u}_m^{UAV}) \right)
 \end{aligned} \tag{34}$$

The lower bound at the given local point is replaced by the left-hand side of $\text{C1}^4, \text{C2}^4$, and C9, so we can obtain the following convex problem:

$$\begin{aligned}
& \min_{\lambda, \eta, \mathbf{u}^U, \varepsilon} \sum_{i=1}^N \{\varepsilon\} \\
& \text{s.t. C1}^5 : \frac{D_i^2}{T_i - \frac{D_i^s(1-c_j^u)}{R_{s,u}^D} - \frac{(D_i^2+D_i^s)F_i}{f_i^u}} \leq b_{i,j}B_U \log \left(x_i + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_U N_0} \right) \\
& \quad - b_{i,j}B_U \log(\hat{x}_i) - \frac{b_{i,j}B_U}{\hat{x}_i} (x_i - \hat{x}_i), \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\
& \text{C2}^5 : \frac{P_j^{\text{hov}} D_i^2 + P_u^s D_i^3 + P_i D_i^2}{E_i^{\text{max}} - E_i - E_i^u - P_j^{\text{hov}} \left(\frac{D_i^s(1-c_j^u)}{R_{s,u}^D} + \frac{(D_i^2+D_i^s)F_i}{f_i^u} \right) - P_u^s \frac{D_i^3}{R_{u,s}^U}} \\
& \quad \leq b_{i,j}B_U \log \left(x_i + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_U N_0} \right) - b_{i,j}B_U \log(\hat{x}_i) + \frac{b_{i,j}B_U}{\hat{x}_i} (x_i - \hat{x}_i), \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \tag{35} \\
& \text{C3}^5 : H_j^2 + \|\mathbf{u}_i^I - \mathbf{u}_j^U\|^2 \leq x_i, \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\
& \text{C4}^5 : \left\| (H_j, \mathbf{u}_j^{\text{UAV}}) - (H_m, \mathbf{u}_m^{\text{UAV}}) \right\|^2 \geq - \left\| (\hat{H}_j, \hat{\mathbf{u}}_j^{\text{UAV}}) - (\hat{H}_m, \hat{\mathbf{u}}_m^{\text{UAV}}) \right\|^2 + \\
& \quad 2 \left((\hat{H}_j, \hat{\mathbf{u}}_j^{\text{UAV}}) - (\hat{H}_m, \hat{\mathbf{u}}_m^{\text{UAV}}) \right)^T \left((H_j, \mathbf{u}_j^{\text{UAV}}) - (H_m, \mathbf{u}_m^{\text{UAV}}) \right) \\
& \text{C11} : b_{i,j}B_U \log \left(1 + \frac{P_i |\hat{h}_{i,j}|^2 \Delta_{i,j}}{b_{i,j}B_U N_0 x_i} \right) \geq \lambda \\
& \text{C12} : B_D \log \left(1 + \frac{P_j |\hat{h}_{i,j}|^2 \Delta_{i,j}}{B_D N_0 x_i} \right) \geq \eta \\
& \text{C10}
\end{aligned}$$

The problem mentioned above can be resolved by utilizing the MATLAB CVX toolbox. At the m -th iteration, $(x_i[m], H_j[m], \mathbf{u}_j^{\text{UAV}}[m])$ can be obtained from $\hat{x}_i = x_i[m-1]$, $\hat{H}_j = H_j[m-1]$, and $\hat{\mathbf{u}}_j^{\text{UAV}} = \mathbf{u}_j^{\text{UAV}}[m-1]$. As the subset of (31) is the feasible set of Equation (35), the upper bound solution for Equation (31) can be obtained by solving an approximate problem of Equation (35). The solution process for problem (31) is depicted in Algorithm 1.

Algorithm 1 Optimization algorithm for the problem of Equation (31)

- 1: **Initialize:** Set $(x_i[0], H_j[0], \mathbf{u}_j^{\text{UAV}}[0])$, and $m = 1$;
 - 2: **repeat**
 - 3: Set $\hat{x}_i = x_i[m-1]$, $\hat{H}_j = H_j[m-1]$, $\hat{\mathbf{u}}_j^{\text{UAV}} = \mathbf{u}_j^{\text{UAV}}[m-1]$;
 - 4: Solve (35) by the MATLAB CVX toolbox and obtain $(x_i[m], H_j[m], \mathbf{u}_j^{\text{UAV}}[m])$;
 - 5: $m = m + 1$;
 - 6: **until** Converge
 - 7: **return** UAV position $H_j, \mathbf{u}_j^{\text{UAV}}[0]$.
-

In conclusion, we propose a comprehensive alternating optimization algorithm using the BCD method to solve Equation (22). Specifically, we divide the variables in Equation (22) into different blocks, which are $\mathbf{a}, \mathbf{c}, \mathbf{b}, \mathbf{D}$, and \mathbf{u}^U . Subsequently, we alternately optimize association control, cache decision, computing task allocation, bandwidth allocation, and

UAV position optimization by solving Equations (23) and (29)–(31) while keeping the other variable blocks fixed. Algorithm 2 summarizes the detailed process of this method.

In the first step of Algorithm 2, initializing $\{\mathbf{a}[0], \mathbf{c}[0], \mathbf{b}[0], \mathbf{D}[0], \mathbf{u}^{UAV}[0]\}$ is necessary. To guarantee the feasibility of the initial point, the initialization strategy should satisfy the constraints of Equation (23). Algorithm 3 provides the initialization method.

Algorithm 2 Optimization algorithm for Equation (22)

- 1: **Initialize:** Set $(a_{i,j}[0], c_j^u[0], b_{i,j}[0], D_i[0], H_j[0], \mathbf{u}_j^{UAV}[0])$ and $r = 1$;
 - 2: **repeat**
 - 3: Obtain $a_{i,j}[r], c_j^u[r]$ by solving (28);
 - 4: Obtain $D_i[r]$ by solving (29);
 - 5: Obtain $b_{i,j}[r]$ by solving (30);
 - 6: UAV positions $(H_j[r], \mathbf{u}_j^{UAV}[r])$ obtained by Algorithm 1;
 - 7: Update the iterative number;;
 - 8: **until** Converge to a prescribed accuracy;
 - 9: **return** Optimal solution $(a_{i,j}, c_j^u, b_{i,j}, D_i, H_j, \mathbf{u}_j^{UAV})$.
-

Algorithm 3 Initialization method

- 1: **Input:** $N, M, D_i, T_i, E_i^{\max}, \mathbf{u}_i^{GT}, d_{\min}, H_{\min}, H_{\max}$;
 - 2: **a[0] Initialize:** $\mathbf{a}[0] = \{a_{ij}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\}$;
 - 3: **c[0] Initialize:** $\mathbf{c}[0] = \{c_{ij}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\}$;
 - 4: **D[0] Initialize:** The computation task data volume originated from both GTs and satellites. The data produced by GTs can be processed locally on the ground, UAVs, and satellites, respectively; i.e., $\mathbf{D}[0] = \{D_i^l + D_i^s, D_i^l = D_i^1 + D_i^2 + D_i^3, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\}$;
 - 5: **b[0] Initialize:** $\mathbf{b}[0] = \{b_{ij}, \forall i \in \mathcal{N}, \forall j \in \mathcal{M}\}$;
 - 6: **u^{UAV}[0] Initialize:** Set the coordinates of the UAVs, and the height of the UAV is $\frac{H_{\min} + H_{\max}}{2}$;
 - 7: **Output:** $\mathbf{a}[0], \mathbf{c}[0], \mathbf{b}[0], \mathbf{D}[0], \mathbf{u}^{UAV}[0]$.
-

4. Results and Analysis

In this section, we conducted extensive simulations to validate the proposed scheme's and algorithm's effectiveness. We consider a system with $K = 20$ that are located at the horizontal coordinates $[0; 20], [20; 5], [20; 0], [5; 20], [20; 20], [10; 10], [0; 30], [30; 5], [30; 0], [5; 30], [30; 20], [10; 20], [-20; 20], [-10; 10], [0; -30], [-30; 5], [-30; 0], [-5; 30], [-30; 20],$ and $[-10; 20]$. Following [21,36,37], the simulation parameters are shown in Table 2.

Table 2. Simulation parameters.

Parameter	Value	Parameter	Value
F_i	1000 cycles/bit	N_0	−170 dBm
C^u	1 Gb [38]	α	2
κ	10^{-28}	B	20 MHz
H_{\min}	40 m	H_{\min}	60 m
ζ	10 N [36]	q	4 [36]
r	0.254 m [38]	η_j	70% [38]
f_i^l	10^8 bit/s	f_i^u	10^9 bit/s
f_i^s	10^{10} bit/s	P_j	0.01 W

We have set up the following three basic schemes for ease of comparison.

1. Local Computation: The data volume of computing tasks is entirely processed by local GTs, and neither the UAVs nor the satellite (LEO) provide computing services for ground GTs.
2. Satellite Computation: The data volume of the computation task is entirely processed by the satellite, and ground GTs and UAVs do not provide computation services for the satellite.
3. Fixed UAV Positions: The data volume of computing tasks can be handled at ground GTs, UAVs, and the LEO satellite. In contrast with the proposed method, this method concurrently optimizes association control, bandwidth allocation, cache decision, and the distribution of computing tasks while maintaining fixed UAV positions.
4. Fixed Bandwidth Allocation: All ground user terminals are allocated equal bandwidth, while the GTs, UAVs, and satellites provide services normally [20].
5. Ground–Air Cooperation: The computation tasks of ground user terminals can be offloaded to UAVs for processing, whereas the satellite does not provide services for GTs and UAVs [21].

Figure 2 illustrates the convergence of the algorithm. It can be observed that the proposed algorithm converges rapidly and obtains the optimal solution. Based on the convergence analysis conducted using the Block Coordinate Descent method [39], it can be observed that when the number of UAVs (M) remains constant and the number of ground user terminals (N) rises, there will be a corresponding increase in the computation delay of each ground user terminal. This is due to the inevitable decrease in the allocation of edge computing resources to each ground user terminal, which leads to an increase in the total computation task delay.

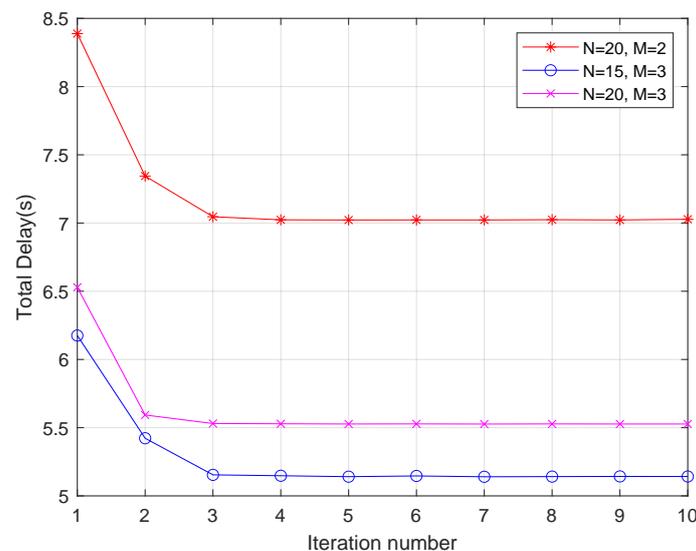


Figure 2. Total delay versus iteration number.

When N remains constant and M increases, the total computation task delay decreases. The reason is that the increase in the number of UAVs leads to a corresponding increase in the computational capacity on the UAV side, which allows UAVs to process tasks quickly and reduces the latency for task processing. In addition, the ground user terminals have the capability to allocate tasks to more UAVs, thereby reducing the processing workload on each UAV and consequently reducing the total computation task delay.

As shown in Figure 3, we plot the relationship between the total computational latency and the local computation capacity. It can be observed that there is a negative correlation between the computing capability of ground user terminals and the total computation task delay. The reason is that the enhancement in computation capability reduces the local processing delay. However, the enhancement of local computation capability does not affect the situation in which all computation tasks are executed on the satellite. In

addition, it is worth noting that, compared with local computation, satellite computation, fixed UAV positions, fixed bandwidth allocation, and ground–air cooperation schemes, the optimization algorithm proposed in this paper effectively reduces the delay.

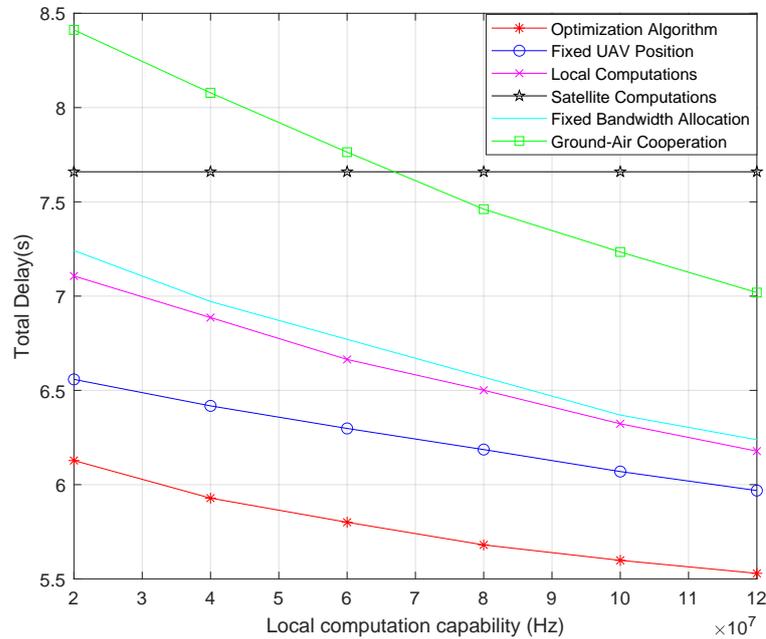


Figure 3. Total delay versus local computation capability.

As shown in Figure 4, we plot the relationship between the total computational latency and the satellite computation capacity. From Figure 4, it can be observed that the total computation task delay decreases with the increase in satellite computation capability. This is due to the increase in satellite computation capacity reducing the processing latency of the tasks handled by the satellite. The enhanced computation capacity of the satellites facilitates collaboration between space, air, and ground, thereby reducing the total computation task delay. In addition, we can conclude that the proposed optimization approach reduces the latency effectiveness compared with local computation, satellite computation, fixed UAV position, fixed bandwidth, and ground–air cooperation schemes.

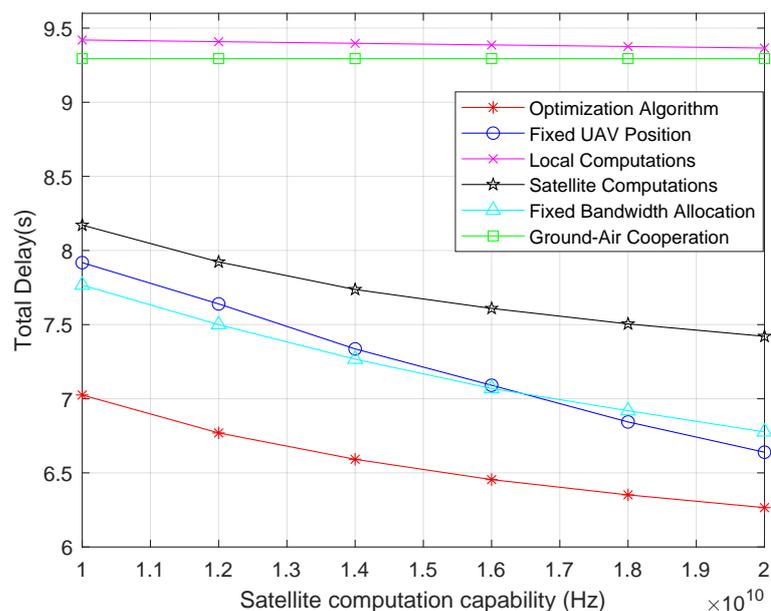


Figure 4. Total delay versus satellite computation capability.

As depicted in Figure 5, we plot the relationship between the total computation task delay and the amount of the computation task data. It can be seen that the total computation task delay increases with the increase in the amount of computation task data in Figure 5. The increase in computation tasks leads to an increase in the transmission latency from ground user terminals to UAVs and from UAVs to satellites. In addition, it increases the number of computation tasks handled by ground user terminals, UAVs, and satellites, leading to an increase in the processing latency of these tasks. Consequently, the total computation task latency increases. The reason for this is that an increase in the amount of computation tasks leads to an increase in the transmission latency from ground user terminals to UAVs and from UAVs to satellites, which, at the same time, increases the computation tasks handled by ground user terminals, UAVs, and satellites, increasing the processing latency of these tasks, thus leading to an increase in the total computation task delay. Besides, it is evident that the optimization scheme proposed in this paper surpasses other baseline approaches.

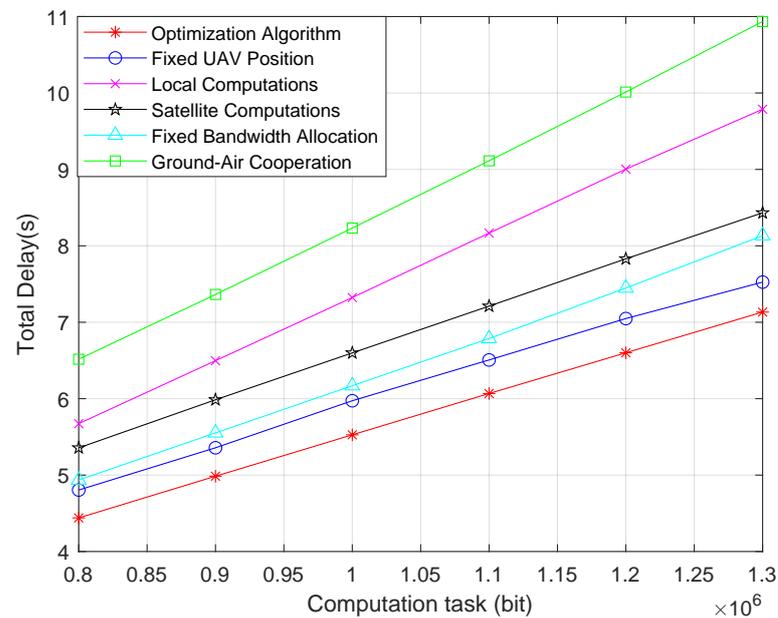


Figure 5. Total delay versus computation task.

We investigate the relationship between the total computation latency and the transmission power under the condition of a Rician fading factor of $\lambda = 3$, as depicted in Figure 6. It is evident that when the transmission power increases, there is a corresponding decrease in the total computation task delay. The reason for this occurrence is that the augmentation in transmission power results in a corresponding augmentation in the transmission rate between ground user terminals and UAVs, hence diminishing the transmission latency both on the ground and in the air. Moreover, the paper's optimization algorithm effectively reduces the total computation task delay when compared with other baseline solutions.

Figure 7 depicts the correlation between the total computation task delay and the bandwidth of the system when the Rician fading factor $\lambda = 3$. From Figure 7, it can be inferred that there is a negative correlation between the total computation task delay and the system bandwidth. The reason for this phenomenon is that an increase in system capacity leads to a rise in the transmission rate between ground user terminals and UAVs, which reduces transmission delay. This has a similar effect on increasing transmission power, which fosters cooperation between terrestrial user terminals and aerial UAVs, further reducing the total computation task delay.

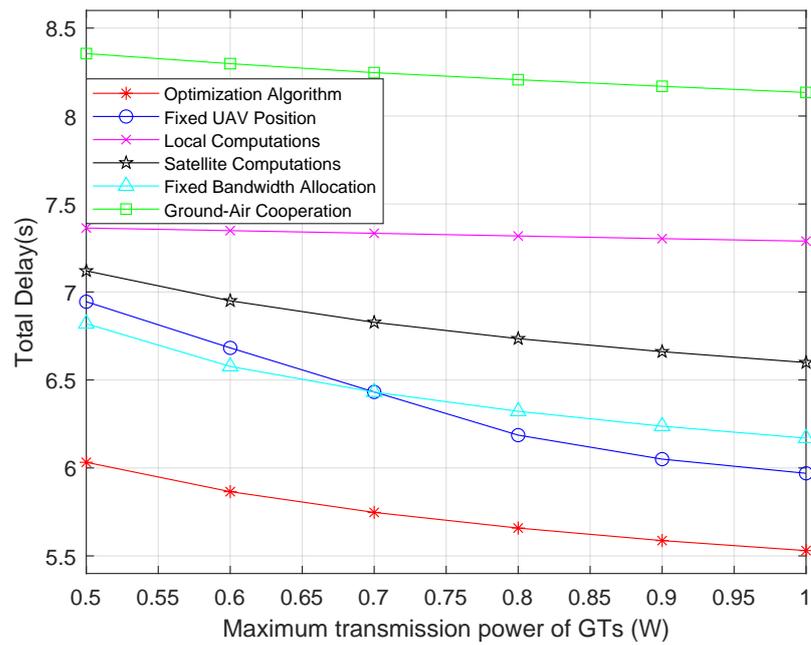


Figure 6. Total delay versus maximum transmission power of GTs.

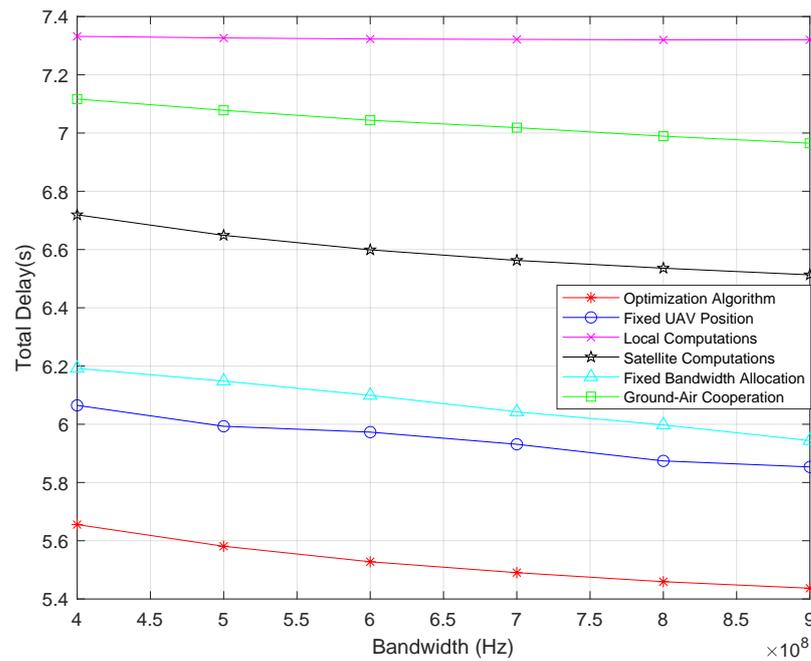


Figure 7. Total delay versus system bandwidth.

5. Discussion

In contrast with conventional research, the novel UAV-assisted SAGIN model proposed in this paper is quite interesting. In order to improve the precision of our findings, we partitioned the computation tasks generated by ground user terminals, enabling separate processing on the ground, UAVs, and satellites. In general, three-layer networks do not typically comprehensively consider optimizing cache variables [23,25]. Moreover, to the best of our knowledge, the existing research mainly focuses on optimizing caching variables in two-tier mobile edge networks [15]. We are the first to optimize caching variables in the three-tier network model of UAV-assisted SAGIN. The optimization of resource distribution to improve network performance has been an important topic of research for scholars globally in the field of MEC. Correspondingly, the proposed framework model in

the paper effectively makes use of the flexibility of UAVs, which improves the performance of the MEC system.

In order to minimize the total computation task delay, we present an iterative algorithm that improves association control, bandwidth allocation, computation task assignment, caching decisions, and the UAV position. However, the present problem is mixed-integer nonlinear programming (MINLP), which is difficult to solve directly. The BCD is frequently utilized to resolve optimization problems with many variables. The core idea is to decompose the optimization problem into several subproblems, wherein all variables are partitioned into distinct blocks, each containing one or more variables. After that, these blocks are updated alternately. During each iteration, a specific block is selected, and the variables in the selected block are considered independent variables, forming a subproblem. The variables in the chosen block are updated, and this process is repeated iteratively until a convergence criterion is satisfied. Specifically, we decompose the proposed problem into four blocks using the BCD method. In the first block, we solve for the association control and caching decision variables while maintaining other variables constantly. The McCormick envelope theory is utilized to decouple the variables, followed by applying convex optimization techniques to convert them into convex problems to solve them. In the above manner, the process is repeated iteratively until the final block, where the only need is to solve for the positions of the UAVs. The problem is converted into a convex problem by introducing new variables for relaxation and conducting a Taylor expansion around local points to get the lower bounds, which can be solved using MATLAB. Finally, the efficacy of the proposed optimization algorithm is validated through simulations.

The results show that the proposed UAV-assisted SAGIN model presented in this paper is significant. Firstly, this model can be applied in various scenarios, including but not limited to emergency communication, post-disaster recovery, communication coverage in remote areas, and military operations. It has the capability to offer flexible, efficient, and dependable communication. The parameters that were studied in the model, including cache variables and UAV positions, are also instructive and can help designers make better MEC-related systems in the future. Comparing the proposed optimization method with different baseline schemes has shown that it works well, demonstrating the significance of the optimized variables in this paper. This indicates their applicability in various real-world scenarios. Finally, the efficacy of the proposed optimization algorithm is validated through simulations.

Future work still has many directions worthy of our exploration, such as (1) multi-agent systems, (2) reconfigurable intelligent surfaces, and (3) artificial intelligence.

- (1) **Multi-Agent Systems:** Multi-agent systems can be deployed on MEC servers on UAVs or at base stations, where each ground user reports its parameters to the agents via UAVs. The agents can regularly update the state of the associated users, determine bandwidth allocation and transmission strategies, and perform other functions based on the current user status. Based on the Markov Decision Process, the actor learns the policy function and the critic learns the value function through interaction with the environment, aiming to maximize resource utilization.
- (2) **Reconfigurable Intelligent Surfaces (RIS):** Obstacles like trees and buildings may inevitably obstruct communication between ground users and UAVs. We can achieve the control and optimization of signals by appropriately configuring the phase and magnitude of the controllable units of RIS. An RIS can expand coverage and enhance signal transmission effectiveness, and compensate for signal coverage blind spots and shadow areas, which improve the coverage range and quality of UAV-assisted SAGIN networks.
- (3) **Artificial Intelligence:** In the UAV-assisted SAGIN model proposed by the paper, we can utilize artificial intelligence algorithms to design resource allocation schemes, which include deep neural networks and deep reinforcement learning, including but not limited to planning UAV paths, making decisions about navigation, searching for targets, sensing the surroundings, and avoiding obstacles. This enables the UAV-assisted SAGIN to allocate communication resources dynamically based on real-

time communication demands and environmental changes, thereby improving the performance and efficiency of the model in various application scenarios.

6. Conclusions

The paper presents a three-tier network model comprising satellites, UAVs, and ground user terminals. In the three-tier network model, the computation tasks of ground user terminals can be divided, and the tasks can be processed by ground stations, UAVs, and satellites. To be more specific, to reduce the total computation task delay, we proposed an iterative approach that jointly optimizes association control, bandwidth allocation, cache decision-making, computation task assignment, and UAV positioning. However, the proposed problem is an MINLP, which makes it difficult to solve directly. We utilize BCD to decompose it into four different blocks, each containing different variables. We keep the other variables constant to solve the variables we need in each block. The McCormick envelope theory separates the associated control and caching decision variables, thereby transforming the non-convex problem into a convex one. A Taylor expansion is conducted on local points to determine the lower bounds and introduce new variables for relaxation to solve the positions of the UAVs. Once all the different blocks have been transformed into convex problems, MATLAB is employed to solve them. Finally, it is shown through simulations that the approach suggested in this paper surpasses other baseline solutions and ensures convergence, effectively lowering the total computation task delay.

Author Contributions: Conceptualization, methodology, Z.X.; investigation, writing—original draft preparation Q.Y.; review and editing, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by the National Key Research and Development Program (308), and in part by the R&D Program of Beijing Municipal Education Commission (KM202211232006) and Major Research Fostering Project to Promote the Colleges Classified Development of Beijing Information Science & Technology University (No. 2121YJPY222).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Mohan, M.; Richardson, G.; Gopan, G.; Aghai, M.M.; Bajaj, S.; Galgamuwa, G.A.P.; Vastaranta, M.; Arachchige, P.S.P.; Amorós, L.; Della Corte, A.P.; et al. Uav-supported forest regeneration: Current trends, challenges and implications. *Remote Sens.* **2021**, *13*, 2596. [[CrossRef](#)]
2. Shakhatreh, H.; Sawalmeh, A.H.; Al-Fuqaha, A.; Dou, Z.; Almaita, E.; Khalil, I.; Othman, N.S.; Khreishah, A.; Guizani, M. Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges. *IEEE Access* **2019**, *7*, 48572–48634. [[CrossRef](#)]
3. Sheng, J.; Cai, X.; Li, Q.; Wu, C.; Ai, B.; Wang, Y.; Kadoch, M.; Yu, P. Space-air-ground integrated network development and applications in high-speed railways: A survey. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 10066–10085. [[CrossRef](#)]
4. Mozaffari, M.; Saad, W.; Bennis, M.; Nam, Y.; Debbah, M. A tutorial on uavs for wireless networks: Applications, challenges, and open problems. *IEEE Commun. Surv. Tutorials* **2019**, *21*, 2334–2360. [[CrossRef](#)]
5. Kato, N.; Fadlullah, Z.M.; Tang, F.; Mao, B.; Tani, S.; Okamura, A.; Liu, J. Optimizing space-air-ground integrated networks by artificial intelligence. *IEEE Wirel. Commun.* **2019**, *26*, 140–147. [[CrossRef](#)]
6. Afghah, F.; Razi, A.; Chakareski, J.; Ashdown, J. Wildfire monitoring in remote areas using autonomous unmanned aerial vehicles. In Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 29 April–2 May 2019; pp. 835–840.
7. Khawaja, W.; Guvenc, I.; Matolak, D.W.; Fiebig, U.; Schneckenburger, N. A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles. *IEEE Commun. Surv. Tutorials* **2019**, *21*, 2361–2391. [[CrossRef](#)]
8. Zhang, P.; Wang, C.; Kumar, N.; Liu, L. Space-air-ground integrated multi-domain network resource orchestration based on virtual network architecture: A drl method. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 2798–2808. [[CrossRef](#)]
9. Zhang, W.; Li, L.; Zhang, N.; Han, T.; Wang, S. Air-ground integrated mobile edge networks: A survey. *IEEE Access* **2020**, *8*, 125998–126018. [[CrossRef](#)]

10. Hu, X.; Wong, K.; Yang, K.; Zheng, Z. Uav-assisted relaying and edge computing: Scheduling and trajectory optimization. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 4738–4752. [[CrossRef](#)]
11. Pervez, F.; Sultana, A.; Yang, C.; Zhao, L. Energy and latency efficient joint communication and computation optimization in a multi-uav assisted mec network. *IEEE Trans. Wirel. Commun.* **2023**, *23*, 1728–1741. [[CrossRef](#)]
12. Nway Nway Ei, Seok Won Kang, Madyan Alsenwi, Yan Kyaw Tun, and Choong Seon Hong. Multi-uav-assisted mec system: Joint association and resource management framework. In Proceedings of the 2021 International Conference on Information Networking (ICOIN), Jeju Island, Republic of Korea, 13–16 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 213–218.
13. Abbasi, O.; Yanikomeroglu, H.; Ebrahimi, A.; Yamchi, N.M. Trajectory design and power allocation for drone-assisted nr-v2x network with dynamic noma/oma. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 7153–7168. [[CrossRef](#)]
14. Zhou, Y.; Zhou, F.; Zhou, H.; Ng, D.W.K.; Hu, R.Q. Robust trajectory and transmit power optimization for secure uav-enabled cognitive radio networks. *IEEE Trans. Commun.* **2020**, *68*, 4022–4034. [[CrossRef](#)]
15. Zhao, N.; Cheng, F.; Yu, F.R.; Tang, J.; Chen, Y.; Gui, G.; Hikmet Sari Caching uav assisted secure transmission in hyper-dense networks based on interference alignment. *IEEE Trans. Commun.* **2018**, *66*, 2281–2294. [[CrossRef](#)]
16. Wu, W.; Zhou, F.; Hu, R.Q.; Wang, B. Energy-efficient resource allocation for secure noma-enabled mobile edge computing networks. *IEEE Trans. Commun.* **2019**, *68*, 493–505. [[CrossRef](#)]
17. Messous, M.O.; Senouci, S.I.; Sedjelmaci, H.; Cherkaoui, S. A game theory based efficient computation offloading in an uav network. *IEEE Trans. Veh. Technol.* **2019**, *68*, 4964–4974. [[CrossRef](#)]
18. Apostolopoulos, P.A.; Fragkos, G.; Tsiropoulou, E.E.; Papavassiliou, S. Data offloading in uav-assisted multi-access edge computing systems under resource uncertainty. *IEEE Trans. Mob. Comput.* **2021**, *22*, 175–190. [[CrossRef](#)]
19. Zhao, N.; Ye, Z.; Pei, Y.; Liang, Y.; Niyato, D. Multi-agent deep reinforcement learning for task offloading in uav-assisted mobile edge computing. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 6949–6960. [[CrossRef](#)]
20. Sriharsha, C.; Murthy, C.S.R. Energy-efficient computation offloading in 6g space-air-ground integrated networks. In Proceedings of the 2022 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Gandhinagar, India, 18–21 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
21. Mao, S.; He, S.; Wu, J. Joint uav position optimization and resource scheduling in space-air-ground integrated networks with mixed cloud-edge computing. *IEEE Syst. J.* **2020**, *15*, 3992–4002. [[CrossRef](#)]
22. Shi, Y.; Zhang, J.; Gao, Y.; Xia, Y. Inter-server computation offloading and resource allocation in multi-drone aided space-air-ground integrated iot networks. *J. Commun. Netw.* **2022**, *24*, 324–335. [[CrossRef](#)]
23. Chen, B.; Li, N.; Li, Y.; Tao, X.; Sun, G. Energy efficient hybrid offloading in space-air-ground integrated networks. In Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022; pp. 1319–1324.
24. Tun, Y.K.; Kim, K.T.; Zou, L.; Han, Z.; Dán, G.; Hong, C.S. Collaborative computing services at ground, air, and space: An optimization approach. *IEEE Trans. Veh. Technol.* **2023**, *73*, 1491–1496. [[CrossRef](#)]
25. Asheralieva, A.; Niyato, D.; Wei, X. Ultra-reliable low-latency slicing in space-air-ground multi-access edge computing networks for next-generation internet of things and mobile applications. *IEEE Internet Things J.* **2023**, *11*, 3956–3978. [[CrossRef](#)]
26. Pervez, F.; Zhao, L.; Yang, C. Joint user association, power optimization and trajectory control in an integrated satellite-aerial-terrestrial network. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 3279–3290. [[CrossRef](#)]
27. Tun, Y.K.; Dán, G.; Park, Y.M.; Hong, C.S. Joint uav deployment and resource allocation in thz-assisted mec-enabled integrated space-air-ground networks. *arXiv* **2024**, arXiv:2401.11419.
28. Chowdhury, M. Superactive: A priority, latency, and sla-aware resource management scheme for software defined space-air-ground integrated networks. *Int. J. Sens. Netw.* **2023**, *41*, 23–41. [[CrossRef](#)]
29. Swaminathan, R.; Sharma, S.; Vishwakarma, N.; Madhukumar, A.S. Haps-based relaying for integrated space-air-ground networks with hybrid fso/rf communication: A performance analysis. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 1581–1599.
30. Zeng, Y.; Wu, Q.; Zhang, R. Accessing from the sky: A tutorial on uav communications for 5 g and beyond. *Proc. IEEE* **2019**, *107*, 2327–2375. [[CrossRef](#)]
31. Feng, J.; Yu, F.R.; Pei, Q.; Chu, X.; Du, J.; Zhu, L. Cooperative computation offloading and resource allocation for blockchain-enabled mobile-edge computing: A deep reinforcement learning approach. *IEEE Internet Things J.* **2019**, *7*, 6214–6228. [[CrossRef](#)]
32. Pang, X.; Tang, J.; Zhao, N.; Zhang, X.; Qian, Y. Energy-efficient design for mmwave-enabled noma-uav networks. *Sci. China Inf. Sci.* **2021**, *64*, 1–14. [[CrossRef](#)]
33. Stolaroff, J.K.; Samaras, C.; O'Neill, E.R.; Lubers, A.; Mitchell, A.S.; Ceperley, D. Energy use and life cycle greenhouse gas emissions of drones for commercial package delivery. *Nat. Commun.* **2018**, *9*, 409. [[CrossRef](#)]
34. Cheng, N.; Lyu, F.; Quan, W.; Zhou, C.; He, H.; Shi, W.; Shen, X. Space/aerial-assisted computing offloading for iot applications: A learning-based approach. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1117–1129. [[CrossRef](#)]
35. Boyd, S.P.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
36. Ei, N.N.; Alsenwi, M.; Tun, Y.K.; Han, Z.; Hong, C.S. Energy-efficient resource allocation in multi-uav-assisted two-stage edge computing for beyond 5g networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 16421–16432. [[CrossRef](#)]
37. Selim, M.M.; Rihan, M.; Yang, Y.; Ma, J. Optimal task partitioning, bit allocation and trajectory for d2d-assisted uav-mec systems. *Peer-Peer Netw. Appl.* **2021**, *14*, 215–224. [[CrossRef](#)]

38. Monwar, M.; Semiari, O.; Saad, W. Optimized path planning for inspection by unmanned aerial vehicles swarm with energy constraints. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6.
39. Kang, Z.; You, C.; Zhang, R. 3d placement for multi-uav relaying: An iterative gibbs-sampling and block coordinate descent optimization approach. *IEEE Trans. Commun.* **2020**, *69*, 2047–2062. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.