Article

Joint Power and Bandwidth Allocation for UAV Backhaul Networks: A Hierarchical Learning Approach

Tingting Yang [1], Kailing Yao [1], Youming Sun [2], Fei Song [1], Yang Yang [1] and Yuli Zhang [1]

[1] the College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China; m15358101536@163.com (T.Y.); kailing_yao@126.com (K.Y.); songfei180517@sina.com (F.S.); sheep_1009@163.com (Y.Y.); yulipkueecs08@126.com (Y.Z.);

[2] No. 61062 Troops of PLA, Beijing China, 10091; sunyouming10@163.com

Abstract: Unmanned Aerial Vehicles (UAVs) severing as the relay is an effective technology method to extend the coverage. It can also alleviate the congestion and increase the throughput, especially applied in UAV networks. However, since the energy of UAVs is limited and the resources in UAV networks are scarce, how to optimize the network delay performance under these constraints should be well investigated. Besides, the relationship among different resources, e.g. power and bandwidth, is coupled which makes the optimization more complex. This article investigates the problem of joint power and bandwidth allocation in UAV backhaul networks, which considers both the delay performance and the resource utilization efficiency. Considering the heterogenous locations characteristics of different UAVs, we formulate the optimization problem as a Stackelberg game. The relay UAV acts as the leader and extended UAVs act as followers. Their utility functions take both the delay durance and the resource consumption into account. To capture the competitive relationship among followers, the sub-game is proved to be an exact potential game and exists Nash equilibriums (NE). The Stackelberg Equilibrium (SE) is proved afterwards. We utilize a hierarchical learning algorithm (HLA) to find out the best resource allocation strategies, which also reduces the computational complexity. Simulation results demonstrate the effectiveness of the proposed method.

Keywords: UAV backhaul networks; Stackelberg game; delay durance; resource allocation; energy efficiency

1. Introduction

In recent years, Unmanned Aerial Vehicles (UAVs) have been widely applied in military, commercial and civilian activities [1]. They can gather and deliver images, texts or videos to ground center stations (GCSs) [2], either can conduct disaster rescuing or provide commodities in extreme environment where infrastructure or necessary services are absent [3]. In order to finish tasks, the UAV networks tend to be in large scale where the performance guarantee turns out to be a tough problem. Besides, since the energy of each UAV is constrained and the available spectrum resource is scarce, how to allocate resources in a reasonable manner should be well studied. Therefore, this article focuses on the power and bandwidth allocation problems to enhance the performance of the UAV network.

Backhaul is used to being the bottleneck of a ground network and several researchers resorted to UAVs to tackle this problem. For example, UAVs were used to alleviate the congestion in high-demand and overloaded situations of small cell (SC) networks [4] [5] [6], either to allocate resources on-demand for users [7] [8], or to extend coverage at the disaster scenes [9]. However, as far as we are concerned, the backhual scheme in UAV networks has not been studied in existing works, which results in some challenges. For example, the battery-powered UAVs need energy to support the hardware and realize mobility. However, the energy is finite and it is urgent to manage them efficiently to extend the lifespan of UAV networks [10] [11]. Another challenge is that the heavy communication load and the large scale of UAV networks make the scarcity of resources more severe. Hence, the nice performance can

 doi:10.20944/preprints201810.0100.v1

not be blindly pursued without taking the resource utilization efficiency into account. Besides, the relationship among resources of different UAVs, e.g. power and bandwidth, are coupled, which will make the resource optimization challengeable. Therefore, how to allocate resources reasonably in UAV networks based on the backhaul scheme is a tough task.

For the sake of considering resource utilization efficiency and delay improvement simultaneously, a new resource allocation strategy is proposed in this paper, which is based on the tradeoff between the resource consumption and delay durance. To realize the above objective, we construct the utility function with delay and resource consumption to mutual restrain. In order to fully utilize the bandwidth, we make the relay small UAV (R-UAV) use a half-duplex mechanism and propose a two-phrase process to help the extended small UAVs (E-UAVs) deliver information. Specifically, in the first phrase, the E-UAVs to R-UAV phrase, all of E-UAVs divide total bandwidth evenly and select the transmission powers to arrive at R-UAV using equal time as much as possible. In the second phrase, the R-UAV to the cluster head large UAV phrase, the R-UAV delivering all the received information can use total bandwidth. Even if the amount of information is heterogeneous, the corresponding E-UAVs could save energy and make full use of bandwidth through selecting appropriate transmission powers in the first phrase. Such resource allocation method makes UAV backhaul network more efficient. It can realize an optimal resource allocation and delay improvement.

The Stackelberg game is always used to solve the distributed decision problem in wireless network [12] [13] [14] [15] where entities have heterogeneous characteristics. In this article, due to different locations of small UAVs, they are classified into R-UAVs and E-UAVs. The small UAV who can deliver information one-hop to the cluster head large UAV is located in the core network serving as a relay. Others who are located in the expanded network require the relay UAV to help delivering information. The heterogeneity of these two kinds of UAVs motivates us to apply the Stackelberg game. The E-UAVs act as followers while the R-UAV acts as the leader. The backward induction method [16] is used to find the Stackelberg Equilibrium (SE) of the game. The lower level sub-game of followers are solved firstly is proved to be an exact potential game which has at least a Nash Equlibrium (NE). The leader find the optimal strategy afterwards based on the observation of the responses of followers. To solve the problem in unknown channel environment, we use a hierarchical learning algorithm (HLA) [17], which helps users learn from their history reward and converge to a SE.

The main contributions can be summarized as follows:

- We investigate the joint optimization of delay durance and resource consumption for UAV backhaul networks, where the coverage of the cluster head is extended. The small UAV locates in the core networks could act as a relay. It can help small UAVs in the extended network deliver information to the cluster head.
- Considering the heterogeneity of UAVs in backhaul networks, we formulate the resource allocation optimization problem and the delay improvement problem as a Stackelberg game. The existence of SE is proved by using the backward induction. The lower level sub-game is proved to be an exact potential game, which certifies the existence of the NE. After that the best strategy of the leader is given based on the best responses of the followers.
- To solve the problem in unknown environment, a hierarchical learning algorithm (HLA) is proposed to reach the SE, by using the interactive learning in different levels. Users update their strategies according to their historical reward value. Simulation results verify the effectiveness of the proposed method.

The rest of this paper is organized as follows: In section 2, we summarize the related work. In section 3, we describe the system model and problem formulation. The section 4 formulates the Stackelberg game for optimizing the delay durance and resource allocation. In section 5, we propose HLA to find out the SE of the game. Simulation results are discussed in section 6. At last, section 7 describes our conclusion.

 doi:10.20944/preprints201810.0100.v1

2. Related Work

The UAV backhaul communication has been studied in many literatures [18] [19] [20] [21]. In information collection and transmission scenes, a UAVs network which is connected by the relay can be more robust than the one where several UAVs respond for the failures individually. However, some problems, e.g.,the delay, may appear. Authors of [18] proposed a new cyclical multiple access mechanism where the users transmitted information to the relay UAV in a cyclical time-division manner, which expanded the coverage of the relay UAV. In [19], the UAV relay connected with the sink and ensured without any delay only within a desired time window. In [20], a resource allocation mechanism was proposed for packet delay minimization but used the total constraint resource. These works all used to optimize the delay durance, but they are at the expense of decreasing throughput performance and increasing resource consumption. In large-scale UAV backhaul networks, spectrum and power resource is scarce and limited, so that it urgent needs to design an optimization mechanism for reasonable resource allocation and delay performance guarantee.

The problem of optimizing resource allocation in UAV relay network has drawn much attentions [10] [11] [23] [22]. Authors of [11] proposed jointly optimizing transmission power, bandwidth and position of the relay to maximize the throughput. In [22], the author proposed a schedule to maximize the minimum signal to interference-plus-noise ratio through optimizing power and time-frequency block. In literature [23], UAVs constitute floating relay cells inside the macro cell to realize frequency reuse and coverage extension. In [10], an UAV aims to maximize its energy efficiency through optimizing power and share spectrum with the primary network. The defect of the above works is that they did not consider the delay enhancement performance, which is not suitable for the scenes of information collection and transmission monitoring environment.

There are several literatures solving the optimization issue from the game theoretic perspective [24] [25]. In [24], they proposed a network formation game to solve the problem of connecting SBSs to the core network by multi-hop UAVs devices and to optimize the delay and throughput. Literature [25] selected the optimal relay and allocated power using a Stackelberg game without the complete channel information (CSI), which used a auction mechanism to obtain a comparable performance. Game theoretic method uses formulas to represent the interaction of incentives and studies the optimal decision-making problems under conflict confrontation. Different from other games, Stackelberg game is suitable for solving the problems where players have heterogeneity. In this paper, we employ Stackelberg game to analyze the interactions among players' decisions due to the heterogeneity of R-UAV and E-UAVs. The leader is superior to followers, so takes actions firstly. In this UAV backhaul networks, the R-UAV acts as the leader to decide total bandwidth and the E-UAVs are followers to decide the transmission power, through which to realize the tradeoff between resource allocation and delay durance.

The existing works studied the optimization of UAV relay through the placement, resource allocation and throughput performance. The differences of our work with theirs can be concluded as follows:

- We proposed consider the delay performance enhancement and the optimization of resource allocation simultaneously. To obtain this objective, we resorted to design utility functions and utilize Stackelberg game method where is adaptive to solve the interaction of different selfish players.
- We consider the tradeoff between energy consumption and delay durance, which is beneficial to extend the lifespan of large scale battery-powered UAV networks.
- Face the severe trend of scarcity of spectrum resources in UAV networks, we utilize the bandwidth in two phrase and enhance the bandwidth utilization through changing the power strategies.
- We consider the coupled relationship among different resources, which is beneficial to the efficient allocation of resources.

.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

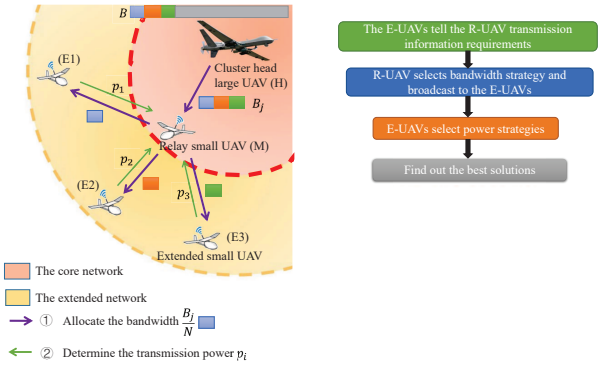

Figure 1. The resource allocation in UAV backhaul networks system model.

3. System Models and Problem Formulation

3.1. System model

We consider a UAV backhaul network involving a cluster head large UAV H, a relay small UAV (R-UAV) M and the extended small UAVs (E-UAVs) $\mathcal{N}$. The R-UAV extends the communication coverage of cluster head large UAV, which could help more small UAVs transmit the information to the ground control station (GCS). Suppose that the number of E-UAVs does not exceed the load of the cluster head. In addition, the small UAV who can acts as a relay must meet two conditions: a) UAV is idle; b) UAV is located in the core network. The core network means the small UAV can deliver information to the cluster head through one hop. We propose a backhaul allocation scheme, where the E-UAVs can deliver the information to the cluster head through the R-UAV.

As shown in Fig. 1, one R-UAV M and E-UAVs users $\mathcal{N}$ constitute our system. In order to eliminate interference, all E-UAVs transmit in the orthogonal channels. Denote the E-UAVs' set as $\mathcal{N} = \{1, 2, ..., N\}$ and each E-UAV's available power profile is $\mathcal{P} = \{p_1, p_2, ..., p_M\}$. For simplicity, we assume all the E-UAVs start to transmit their information to the R-UAV at the same time with the information transmission demand $\mathcal{T} = \{t_1, t_2, ..., t_N\}$. The R-UAV M works in a half-duplex mode. Firstly, the R-UAV decides the total bandwidth $B_j \in \mathcal{B}$ and broadcasts to all the E-UAVs after the E-UAVs told the information demands $t_i \in \mathcal{T}$ to the R-UAV. Secondly, each E-UAV i decides its own transmission power p_i and transmits all the information demand to the R-UAV with the delay as similar to others. In this phrase, all the E-UAVs are allocated the bandwidth evenly, which means $b_i = \frac{B_j}{N}$. In the third stage, the R-UAV M transmits all information received to the cluster head large UAV H using the total bandwidth B_j after the last E-UAV finished its information transmission.

Referring from the UAV-to-UAV multi-hop links model given by [24], we assume that UAVs transmit information over the sub-6 GHz band and the free-space path loss model ξ is

$$\xi(dB) = 20\log_{10}(f_c) + 20\log_{10}(d_{i,M}) - 147.55, \tag{1}$$

where f_c is the system center frequency (in Hz) and $d_{i,M}$ is the distance between the E-UAV i and the R-UAV M.

Referring from literature [26], the signal-to-noise rate (SNR) between E-UAV i and R-UAV M is

$$r_{i,M} = \frac{p_i}{10^{L_{i,M}/10}\sigma^2}, \tag{2}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

where p_i is transmission power from the E-UAV i to the R-UAV M, and σ is the noise. The rate of the A2A link is $R_{i,M} = b_i \log_2(1 + r_{i,M})$. The delay d_i of the E-UAV i delivering its information to the R-UAV M can be depicted as

$$d_i = \frac{t_i}{\frac{B_j}{n}\log_2(1 + \frac{p_i}{10^{L_{iM}/10}\sigma^2})}, \tag{3}$$

where $L_{i,M} = \xi_{i,M} + \eta_{LoS}$. We default that all the small UAVs propagate at line-of-sight (LoS) links.

3.2. Problem formulation

The objective of our study is to jointly optimize the delay durance and energy consumption. Inspired by the literature [16], the utility function can be defined as profit minus cost. In our system, we hope to make a tradeoff between the total delay and power or bandwidth consumption. The bigger power or bandwidth is, the bigger information transmission rate will be, and the smaller delay will be.

We wish all the E-UAVs to finish their information transmission as much as possible at the same time to make full use of the bandwidth resource, which is because they has divided the total bandwidth B_j evenly. The R-UAV M waits for the last E-UAV to finish its information transmission, and then it can deliver the received information to the cluster head large UAV H. The power selected by the E-UAV i and the bandwidth selected by the R-UAV M are denoted as $p_i \in \mathcal{P}$ and $B_j \in \mathcal{B}$, respectively. Specifically, the utility function of each E-UAV i can be defined as:

$$u_i = -\max[d_i] - cp_i d_i, \tag{4}$$

where c is the normalization coefficient of the power. The first term of equation (4) represents the delay of the last E-UAV one-hop transmitting to the R-UAV M. The physical meaning of the second term is the energy consumption of the E-UAV i, which is calculated by the power multiplying delay. Therefore, we optimize the selection of the power and bandwidth to maximize the utility function of each E-UAV i. The corresponding optimization can be depicted as follows:

$$(P1): p_{opt} = \arg\max u_i = \arg\max(-\max[d_i] - cp_i d_i). \tag{5}$$

For the R-UAV M, the delay of information transmission from R-UAV to cluster head can be depicted as:

$$d(M,H) = \frac{\sum_{i=1}^{N} t_i}{B_j \log_2(1 + \frac{p_L}{10^{L_{M,H}/10}\sigma^2})}. \tag{6}$$

Its utility function is composed of the opposite value of the system total delay and its bandwidth cost, which can be depicted as:

$$u_0(B_j, \mathbf{p}) = -\max[d_i] - d(M,H) - B_j C_L, \tag{7}$$

where C_L and P_L denote the bandwidth cost coefficient and power for the R-UAV M, respectively. The second term represents the delay from the R-UAV M to the cluster head H. We can express the optimization of the bandwidth selection as follows:

However, due to the available power and bandwidth are discrete values and the environment is uncertain, the optimization issues $P1$ and $P2$ cannot be solved by traditional optimization methods. In the following, we solve this problem as a Stackelberg game, and HLA is proposed to find out the SE.

4. Stackelberg Game of UAV Backhaul Networks

In this section, we formulate jointly optimizing the delay durance and resource allocation problems as a Stackelberg game. The power selection game in the E-UAVs has Nash equilibriums,, which is proved by an exact potential game. Then, the existence of Stckelberg equilibrium is certified through

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

combining with the finite bandwidth strategies selections game. At last, a hierarchical stochastic learning algorithm is proposed to find out the optimal solution.

4.1. Stckelberg game model

Mathematically, the resource allocation game could denote as $\mathcal{G} = \{\mathcal{N} \cup M, \mathcal{B}, \mathcal{P}, \{u_i\}_{i\in\mathcal{N}}, u_L\}$, in which $\mathcal{N} = \{1, 2, ..., N\}$ and M represent the set profile of the E-UAVs and the R-UAV. $\mathcal{P}$ represents the set of selectable power strategies for all the E-UAVs and $\mathcal{B}$ is the set of selectable bandwidth strategies for the R-UAV M. $\{u_i\}_{i\in\mathcal{N}}$ and $\{u_L\}$ are the utility function of the E-UAV i and R-UAV, respectively. Specifically, the R-UAV M acts as the leader, who can take an action firstly. The E-UAVs act as followers, who would select the best response dynamic rationally based on the action of the leader. We consider the Stackelberg game model with one leader and multiple followers. The way of we solving the Stackelberg game is to find out the Stackelberg equilibrium (SE).

From the E-UAV's side, it is objective to optimize the maximal delay of the followers' information transmission using as little as possible energy cost, and its utility function can be defined as equation (4). The objective of the E-UAV i is to change its power strategy to maximize its utility function, and the optimization could be depicted as $p_i = \arg\max_{p_i} u_i(p_i, P_{-i}, B_j)$.

From the perspective of the E-UAVs, the lower hierarchical sub-game can be depicted as:

$$\mathcal{G}_f = \{\mathcal{N}, \mathcal{P}, u_i(p_i, p_{-i}, B_j)\}, \tag{8}$$

where all the E-UAVs $\mathcal{N}$ are players, and their available strategy set is power set $\mathcal{P}$. Every user wants to maximize its own utility function through selecting an optimal power strategy independently and selfishly.

For the R-UAV M, the objective is to jointly optimize the delay of system and the system bandwidth cost. Its utility function can be depicted as equation (7). The optimization of the R-UAV M is $B_j = \arg\max_{B_j} u_0(\mathbf{P}, B_j)$. The sub-game of the leader can be depicted as:

$$\mathcal{G}_l = \{\mathcal{B}, M, u_0(\mathbf{P}, B_j)\}, \tag{9}$$

where $\mathcal{B}$ is the selectable bandwidth strategies set of the R-UAV M. Each E-UAV has the same power strategies set $\mathbf{P}$.

4.2. Stackelberg Game Analysis

Due to the heterogeneity of UAV in backhaul networks, we formulate a Stackelberg game to solve the joint optimization. In this model, The R-UAV acts as the leader and E-UAVs act as followers. The leader make action firstly. Followers rationally choose the best dynamic response based on the observed upper-level leader actions. In this subsection, we proved the potential game and analyzed the NE. Finally, we proved the existence of SE.

Definition 1 (Exact Potential Game) [27]: A strategy formulated game $\mathcal{G}_f$ is an exact potential game, when there exists a potential function ϕ and any player's unilateral deviation causing the variation in ϕ equals to the variation in the utility function. We can depict it mathematically,

$$\begin{aligned} &\phi(\tilde{p}_i, p_{-i}, B_j) - \phi(p_i, p_{-i}, B_j) \\ &= u_i(\tilde{p}_i, p_{-i}, B_j) - u_i(p_i, p_{-i}, B_j), \\ &\forall i \in N, \forall p_i \in \mathbf{P}, \tilde{p}_i \neq p_i, \end{aligned} \tag{10}$$

where $\tilde{p}_i$ is the action of user i after unilateral deviation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

Definition 2 (Nash Equilibrium) [16]: We can define the strategy collection $\mathbf{p}^* = (p_1^*, p_2^*, ..., p_N^*)$ as a pure strategy Nash equilibrium if and only if no player can improve its utility by deviating unilaterally. It can be expressed mathematically:

$$u_i(p_i^*, p_{-i}^*, B_j) \geq u_i(p_i, p_{-i}^*, B_j), \forall i \in \mathcal{N}, p_i \in \mathcal{P}, p_i^* \neq p_i, \tag{11}$$

where p_{-i}^* represents the set of actions for all participants except participant i.

Theorem 1: The follower sub-game $\mathcal{G}_f$ with given bandwidth strategy B_j is an exact potential game, and have at least one pure strategy NE point.

Proof. Referencing literature [16], the details are as follows. We construct a potential function as follows:

$$\phi(p_k, \mathbf{p}_{-k}, B_j) = -\max[d_k] - \sum_{k=1}^{N} p_k d_k. \tag{12}$$

If the E-UAV k unilaterally changes his action from p_k to $\tilde{p}_k$, the change of its utility function is:

$$\begin{aligned} &u_i(\tilde{p}_k, p_{-k}, B_j) - u_k(p_k, p_{-k}, B_j) \\ &= -\max[\tilde{d}_k] - \tilde{p}_k \tilde{d}_k + \max[d_k] + p_k d_k. \end{aligned} \tag{13}$$

At the same time, the change of the potential function is:

$$\begin{aligned} &\phi(\tilde{p}_k, p_{-k}, B_j) - \phi(p_k, p_{-k}, B_j) \\ &= -\max[\tilde{d}_k] - \textstyle\sum_{i \in \{\mathcal{N}/\{k\}\}} p_i d_i - \tilde{p}_k \tilde{d}_k \\ &-\{-\max[d_k] - \textstyle\sum_{i \in \{\mathcal{N}/\{k\}\}} p_i d_i - p_k d_k\} \\ &= -\max[\tilde{d}_k] - \tilde{p}_k \tilde{d}_k + \max[d_k] + p_k d_k\} \\ &= u_i(\tilde{p}_k, p_{-k}, B_j) - u_k(p_k, p_{-k}, B_j). \end{aligned} \tag{14}$$

We notice that the change of the action of the E-UAV k has no influence on others. In addition, the actions of other E-UAVs have no influence on the change of the potential function before and after the E-UAV k changing the power action. Thus, the game $\mathcal{G}_f$ formulated is an exact potential game, and at least have one pure strategy NE point. We can find out the optimal pure strategy NE of $\mathcal{G}_f$, which is the global optimal solution of problem $P1$. □

Definition 3 (Stackelberg Equilibrium) [16]: If ${B_j}^*$ could make the utility function of the R-UAV M maximal and $\mathbf{p}^*$ are the best response of all the E-UAVs, we call the strategy combination $({B_j}^*, \mathbf{p}^*)$ as a Stackelberg equilibrium point of the Stacklberg game. Mathematically, for any combination of strategies $(B_j, \mathbf{p})$, the conditions of follows are always satisfying:

$$u_0({B_j}^*, \mathbf{p}^*) \geq u_0(B_j, \mathbf{p}^*), \tag{15}$$

$$u_i(p_i^*, p_{-i}^*, {B_j}^*) \geq u_i(p_i, p_{-i}^*, {B_j}^*), \forall i \in N, \tag{16}$$

where p_{-i}^* represent that all the E-UAVs except the E-UAV i adopt the best response of the strategy vector. Stackelberg game is a non-cooperation game, and belongs to a multi-stage dynamic game. When the system is at a SE point, all the participants in the hierarchical structure cannot increase their own utility by only changing own strategy.

Next, we will analyze the existence of SE in the proposed Stackelberg game.

Theorem 2: The Stackelberg game $\mathcal{G}$ formulated by the optimization for the delay durance and resource consumption always exists a SE point.

Proof. Given any one of the stationary bandwidth strategy B_j, the Stackelberg game degenerates into a non-cooperation game $\mathcal{G}_f = \{\mathcal{N}, \mathcal{P}, u_i(p_i, p_{-i}, B_j)\}$. We have certified the existence of Nash Equilibrium in lower hierarchial sub-game through an exact potential game. Therefore, there is always

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

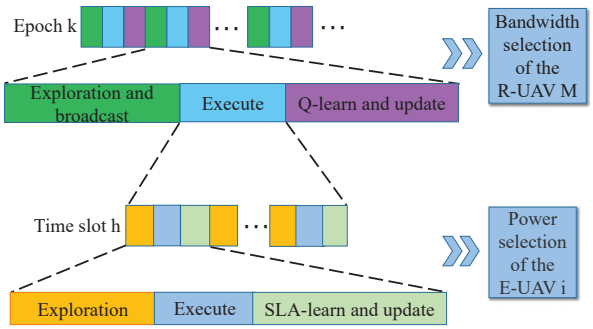

Figure 2. The hierarchical learning algorithm model.

existing a $NE(B_j)$. In the leader sub-game, there is a finite bandwidth strategies set, and we can find out an optimal bandwidth policy ${B_j}^* \in \mathcal{B}$ of the R-UAV:

$$ {B_j}^* = \arg_{B_j} \max u_0(B_j, NE(B_j)). \tag{17} $$

Therefore, $({B_j}^*, NE({B_j}^*))$ constitutes a SE in the proposed Stackelberg game. □

5. Hierarchical Learning Algorithm

5.1. Algorithm Description

In this subsection, we propose a hierarchical learning algorithm to solve the joint optimization of delay durance and resource allocation. Under the learning framework, the smart agent could observe the state of the environment and then select an optimal action strategy. Each smart agent can optimize its future return through selecting the corresponding action. We define the selection of bandwidth and power with a mixed strategies form requested by the proposed HLA. The hierarchical learning algorithm mentioned in this article could be analyzed with a time frame structure in Fig. 2 referring literature [28]. The R-UAV M updates its strategies selection probabilities at each epoch h, and each epoch could be decomposed into K time slots, so as to the E-UAVs update their strategy selection probabilities at each time slot. The R-UAV's available bandwidth strategies set is $\mathcal{B} = \{b_1, ..., b_M\}$ and at epoch h the mixed strategies can be denoted as $\omega_0(h) = (\omega_{01}(h), ..., \omega_{0m}(h), ..., \omega_{0M}(h))$ and $\sum_{m \in \mathcal{B}} \omega_{0m}(h) = 1$. The E-UAVs' available power strategy set is $\mathcal{P} = \{p_1, ..., p_L\}$. At time slot k, the mixed strategies of E-UAV i are $\omega_i(k) = (\omega_{i1}(k), ..., \omega_{il}(k), ..., \omega_{iL}(k))$ and $\sum_{l \in \mathcal{P}} \omega_{il}(k) = 1$. In this way, the HLA proposed could be initialized. In the lower hierarchical sub-game, we use the stochastic learning automata (SLA) algorithm [17] to help the E-UAVs select the power strategy and deliver the information to the R-UAV M. The advantage of this algorithm is that it does not require the interaction of user's information. Each user updates his own strategies selection probabilities according to his own profit. The SLA algorithm selects the optimal strategy by repeated iterating in a random environment, so it is widely used for the decision problems in the field of wireless communication [29] [30] [31] [32]. At the kth time slot, the random profit of the E-UAV i can be defined as follows:

$$ u_i = e^{(-\max[d_i] - c p_i d_i)/l}, \tag{18} $$

where l denotes the adjustment factor and it ensures the value of profit of the E-UAV i between 0 and 1. The trend of change is the same as that of the original utility function. We put the utility function of the E-UAV i in exponential position to ensure its non-negativity and effectiveness of the proposed algorithm.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

Algorithm 1: Hierarchical learning algorithm (HLA)

Step 1: Initialization: set $h = 0$, $k = 0$ and the R-UAV M with all the E-UAV users initialize their selectable mixed strategy probabilities with the average value $\omega_{0m}(h) = 1/|\mathcal{B}|$, $\omega_{il}(k) = 1/|\mathcal{P}|, \forall m \in \mathcal{B}, \forall l \in \mathcal{P}$.
Step 2: In the hth epoch, the R-UAV stochastically selects total bandwidth $a_m(h)$ according to its strategy profile $\omega_0(h)$, and broadcasts it to all the E-UAVs.
Step 3: Learning process of all the E-UAVs
(1) At the beginning of time slot k, each E-UAV selects its transmission power a_i stochastically according to its current strategy selection probability set $\omega_i(k)$.
(2) Each E-UAV i calculates its profit $U_i(a_0, a_i^k, a_{-i}^k)$.
(3) Each E-UAV updates its strategies selection probabilities according to the following rules:

$$\begin{aligned} \omega_{il}(k+1) &= \omega_{il}(k) + \eta\tilde{u}_i(k)(1-\omega_{il}(k)), l = a_i(k), \\ \omega_{il}(k+1) &= \omega_{il}(k) - \eta\tilde{u}_i(k)\omega_{il}(k), l \neq a_i(k), \end{aligned} \tag{20}$$

where $0 < \eta < 1$ is a learning step, and $\tilde{u}_i(k)$ is the normalized profit, which value is between 0 and 1.
Step 4: The R-UAV M calculates its utility $u_0(h)$.
Step 5: The R-UAV M updates its Q value according to its Q function as follows.

$$Q_0^{h+1}(a_m) = (1-\kappa_0^h)Q_0^h(a_m) + \kappa_0^h u_0(h) \tag{21}$$

where $\kappa_i \in [0,1)$ is a learning rate, and it meets the condition of $\sum_{h=0}^{\infty}\kappa_i = \infty$, $\sum_{h=0}^{\infty}(\kappa_i)^2 < \infty$.

$$\omega_{0m}(h) = \frac{\exp[Q_0^h(a_m)/\tau_0]}{\sum_B \exp[Q_0^h(a_m)/\tau_0]} \tag{22}$$

where the temperature $\tau_0 > 0$, and it can make a tradeoff between exploration and exploitation. When τ_0 is bigger, the relay would select the strategy more randomly, otherwise the relay would select the strategy which can make the value of Q maximum. We make the trend of τ_0 from big to small to approximate the optimal solution.
Step 6: The R-UAV selects an action according to the newest strategy selection probability.
Step 7: Update $k = k+1$ until $k = k_{\max}$.

In the upper hierarchical sub-game, we propose a Q-learning algorithm for the R-UAV M selecting the bandwidth strategy. In the process of Q learning, all the actions of the R-UAV M are represented by Q values, which represent the relative payoff value of the selected actions when the R-UAV interacts with the environment. We make the action of the corresponding high return value strengthen continuously through updating the Q function of the R-UAV M repetitively, which could help find out the optimal bandwidth strategy. Correspondingly, the profit of the R-UAV M is given by the equation as follows at the hth epoch.

$$u_0 = e^{(-\max[d_i]-d(M,H)-B_jC_L)/s}, \tag{19}$$

where s denotes the coefficient to ensure the utility function of the R-UAV M between 0 and 1, which makes the Q value update within a reasonable range.

As shown in Fig. 2, it depicts the proposed hierarchy learning algorithm clearly. At last, we default that the stop criterions is satisfying the maximum number of iterations. The proof of convergence for the proposed HLA can be referred from literature [16], which has also proved the HLA can always find a SE.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

5.2. The Proof of the Convergence

In this subsection, we show the theoretical proof of the algorithm convergence as follows. The specific proof process can refer to literature [16], because the hierarchical framework of the algorithm is similar.

Lemma 1 [33]: The SLA algorithm in the low level's sub-game will converges to a pure NE point, when the learning step η tends to zeros.

Lemma 2 [28]: The Q-learning algorithm in the high level's sub-game can converge to $(\mathrm{B}_j{}^*, NE(\mathbf{p}^*))$.

Lemma 3 [16]: The proposed HLA can always find out a SE.

6. Simulation Results and Discussions

6.1. Scenario setup

In this subsection, we show the simulation results of our proposed HLA. The parameters in this simulation refer from literature [24] about multi-hop UAV backhaul communication, which is described specifically at TABLE. 1.

Table 1

Parameters	Values
Transmit power ($\mathcal{P}$)	$[0.5, 1.5, 2.2, 3, 5]$W
Channel bandwidth ($\mathcal{B}$)	$[11, 13, 15, 17, 19]$MHz
Noise variance (σ^2)	$9 * 10^{-13}W$
Carrier frequency (f_c)	2GHz
η_{LoS}	$5dB$
Distance ($d_{i,M}$, $d_{M,H}$)	($random * 100\ m$, 800m)
Information demand	$5 \sim 8$ Mbit

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

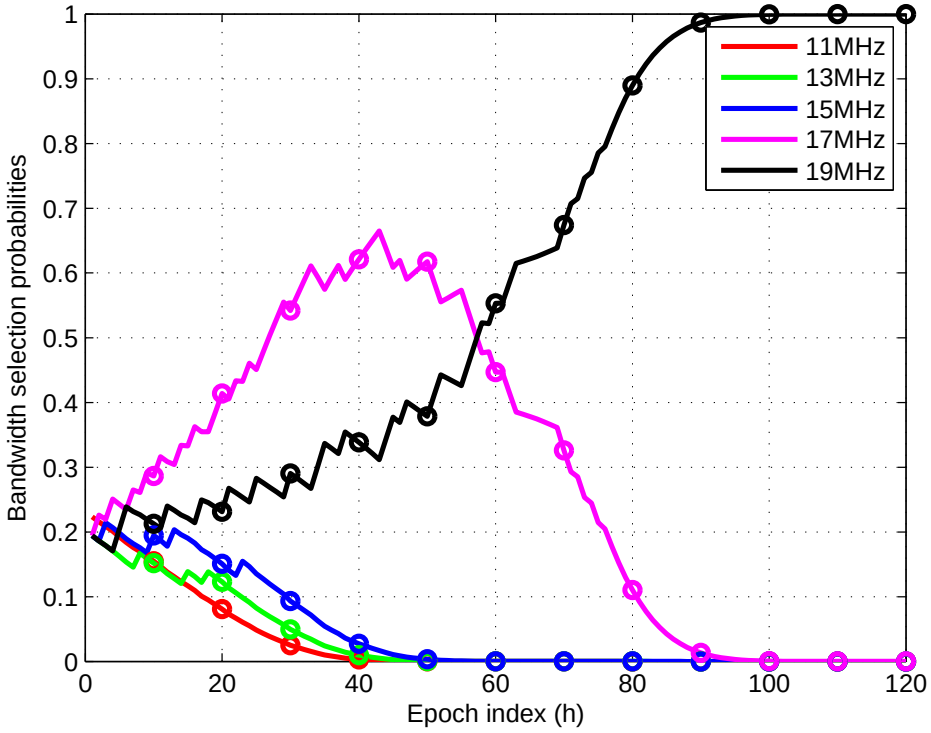

Figure 3. Convergence trend of bandwidth selection probabilities of the relay small UAV *M*.

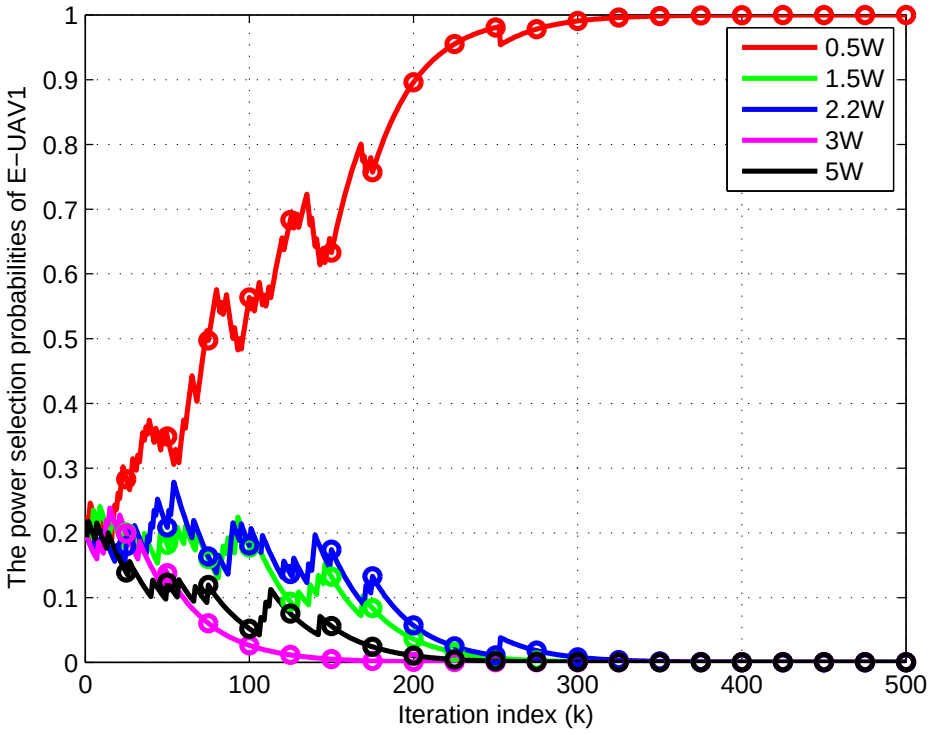

Figure 4. Convergence trend of power selection probabilities of user 1 in the first epoch.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

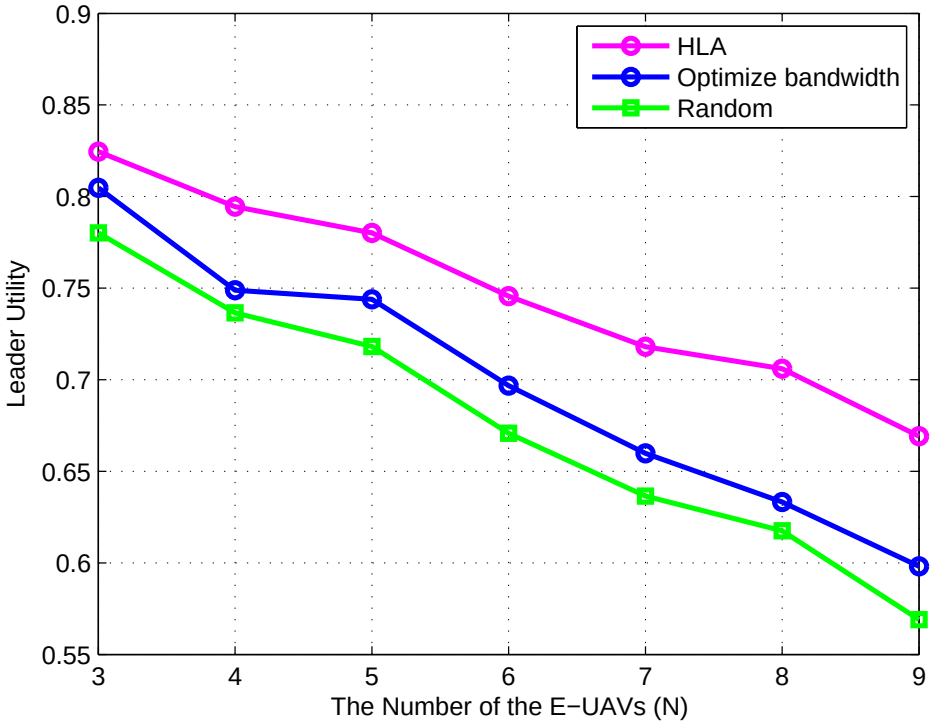

Figure 5. Performance comparison of the utility of leader for different resource strategy choose algorithm.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

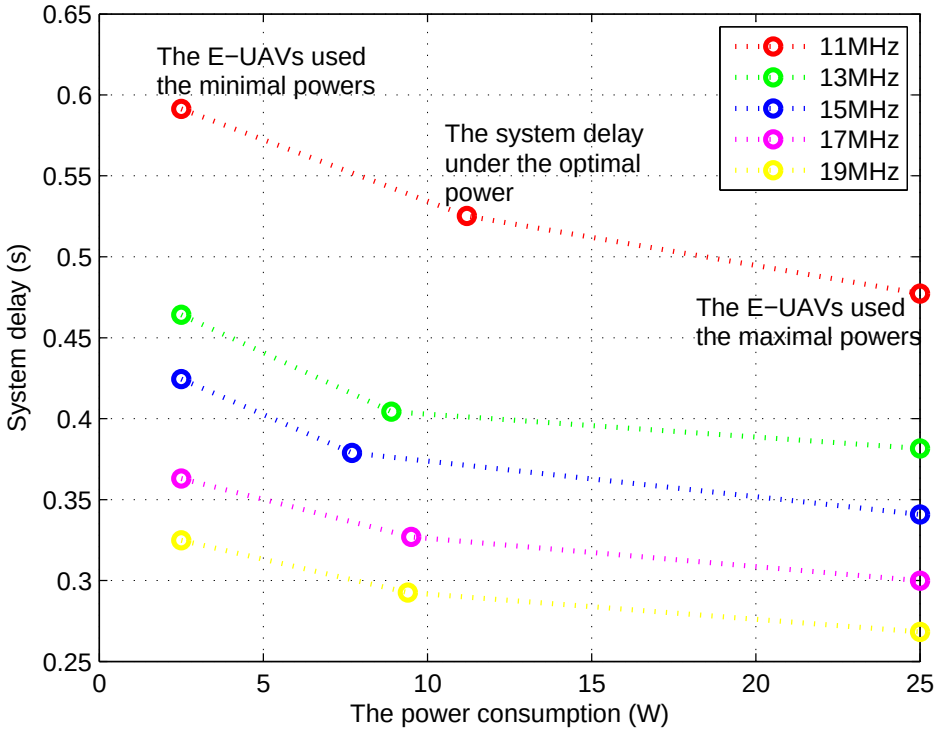

Figure 6. The tradeoff between system delay and power consumption for different bandwidth strategies.

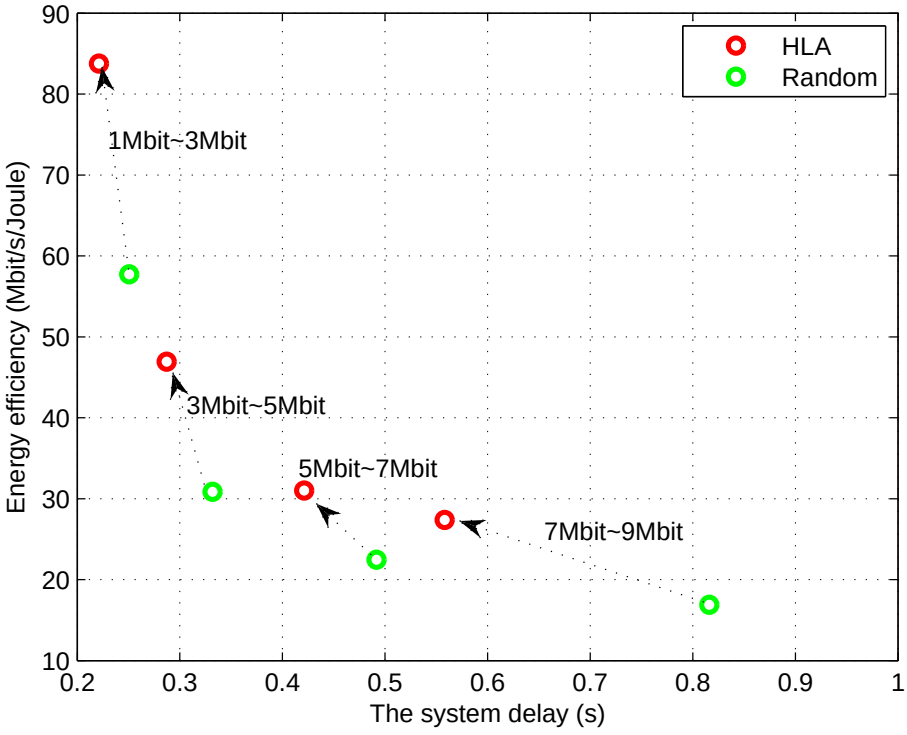

Figure 7. The compromise between energy efficiency and the system delay.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2018 doi:10.20944/preprints201810.0100.v1

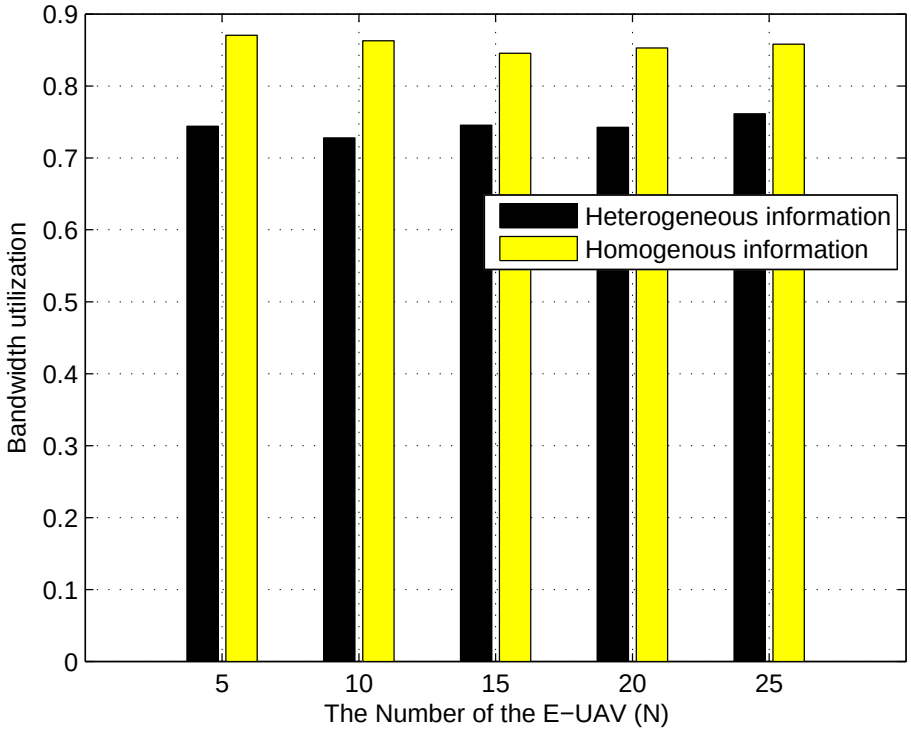

Figure 8. The bandwidth utilization for the system.

We can observe the convergence of the proposed HLA in a single simulation presented in Fig. 3 and Fig. 4. In the leader's sub-game, the R-UAV M repeats 120 epochs to select the optimal bandwidth strategy. The selection probability ω_0 of bandwidth strategy 5 (19 MHz) converges to 1 using about 100 epochs, and other strategies converge to 0 in Fig. 3. The power strategies convergence performance of the E-UAV user 1 in the first epoch is shown in Fig. 4, which begin with equal probabilities. E-UAV user 1 converges to power 1 (0.5 W) about 325 times iterations, and selective probabilities of other strategies converge to 0.

Fig. 5 shows that the leader's utility optimized power and system bandwidth by HLA, only optimized power by SLA for the lower-level user's strategy and selected power and system bandwidth by the random strategies. It can be seen from the Fig. 5 that our proposed method obtains the largest leader's utility. The larger the leader's utility is, the smaller the sum of the system delay and the resource consumption will be, which embodies the superiority of our proposed algorithm.

Fig. 6 shows the proposed method has a good enhancement for delay performance. The five curves represent different bandwidths strategies. The three nodes of each curve represent the corresponding system delay under the current bandwidth strategy in three conditions: a) all E-UAVs in the lower layer select the minimum powers; b) all E-UAVs select the optimal powers obtained by HLA; c) all E-UAVs select the maximum powers. We can see that each line is downward convex, which reflects the superiority of our method. It realized the tradeoff between the system delay and the power consumption with different system bandwidth strategies.

Fig. 7 shows the tradeoff between energy efficiency and system latency obtained using HLA and random allocation algorithm of resources. The four pairs points in red and green represent that other parameters are fixed and the amount of information for E-UAVs is [1∼3, 3∼5, 5∼7, 7∼9] Mbit, respectively. HLA is better than random allocation of resources because it not only increases the energy efficiency, but reduces the system delay as well.

 doi:10.20944/preprints201810.0100.v1

The energy efficiency can be defined as follows [34]:

$$\eta_{EE} = \frac{Total\ data\ rate}{Total\ energy\ consumption}. \tag{23}$$

Based on the half-duplex relay mode, the EE of the UAV backhaul networks is given by

$$\eta_{EE} = \frac{R_{SD}}{\sum_{i=1}^{N} p_i d_i + P_{relay} d_{M,H}}, \tag{24}$$

$$R_{SD} = \frac{\sum_{i=1}^{N} t_i}{\sum_{i=1}^{N} d_i + d_{M,H}} \tag{25}$$

where R_{SD} denotes the rate from the E-UAV to the cluster head large UAV. $p_i d_i$ denotes the energy consumption of each E-UAV delivering information to the R-UAV.

Fig. 8 shows the bandwidth utilization for the system. The big gap between the amount of carried information by E-UAVs is heterogeneous information. Similarly, The small gap between the amount of carried information by E-UAVs is homogeneous information. The homogeneous information caused higher system bandwidth utilization. The reason is that the heterogeneous information makes some E-UAVs take longer time to wait. In addition, increasing the number of the E-UAVs would not have a significant impact on system bandwidth utilization.

7. Conclusion

In this paper, we investigated the joint optimization for delay durance and the resource allocation in UAV bakhaul networks. Different from previous works only pursuing the network performance, we also considered resource utility efficiency. The reason is that the energy is limited and the resource is scarce in the large scale UAV network. Besides, the coupled relationship between different resources increases the complexity of the problem. We maked a tradeoff between the delay and the resource consumption in UAV backhaul networks by formulated as a Stackelberg game. The R-UAV acted the leader and the E-UAVs acted as followers. The lower hierarchical sub-game exists a NE solution proved by an exact potential game, which is combined with the best bandwidth strategy constituting the SE solution. Finally, a hierarchical learning algorithm was proposed and simulation results revealed that the proposed backhaul allocation method was efficient for delay improvement and resource utilization.

Author Contributions: Tingting Yang and Yuli Zhang conceived and designed the model; Tingting Yang performed the game analysis and simulation; Yuli Zhang analyzed the game proof and simulation result; Tingting Yang wrote the paper; and Kailing Yao, Youming Sun and Yang Yang provided some suggestions and revised the paper.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 61771488, No. 61671473 and No. 61631020, in part by the Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province under Grant No. BK20160034, and in part by the Open Research Foundation of Science and Technology on Communication Networks Laboratory.

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

References

1. Peng, W.; Gu, Q.; Sun, D. Wireless sensor network data collection by connected cooperative UAVs. In Proceedings of American Control Conference, pp. 5911-5916, 2013.
2. Pinto, L.; Almeida, L.; Rowe, A. Video Streaming in Multi-hop Aerial Networks: Demo Abstract. in Proceedings of ACM/IEEE International Information Processing in Sensor Networks Conference, Pittsburgh, PA, USA, April 18-21, pp.283-284, INSPEC Accession Number: 16962349.
3. Merwaday, A.; and Guvenc, I. UAV Assisted Heterogeneous Networks for Public Safety Communications. In Proceedings of IEEE Wireless Communications and Networking Conference, New Orleans, LA, USA, March 9-12, 2015, pp. 329-334, DOI: 10.1109/WCNCW.2015.7122576.

4. Siddique, U.; Tabassum, H.; Hossain, E.; and Dong, I. K. Wireless Backhauling of 5g Small Cells: Challenges and Solution Approaches. *Journal of IEEE Communications Society of Wireless Communications*, **2017**, *22*, 22-31, DOI: 10.1109/MWC.2015.7306534.
5. Jaber, M.; Imran, M. A.; Tafazolli, R.; and Tukmanov, A. 5g Backhaul Challenges and Emerging Research Directions: A Survey. *IEEE Access* **2017**, *4*, 1743-1766, DOI: 10.1109/ACCESS.2016.2556011.
6. Semiari, O.; Saad, W.; Daw, Z.; and Bennis, M. Matching Theory for Backhaul Management in Small Cell Networks with MmWave Capabilities. In Proceedings of IEEE International Communications Conference, London, UK, June 8-12, 2015, pp. 3460-3465, DOI: 10.1109/ICC.2015.7248860.
7. Ahdi, F.; and Subramanian, S. Using Unmanned Aerial Vehicles as Relays in Wireless Balloon Networks. In Proceedings of IEEE International Communications Conference, London, UK, June 8-12, 2015, pp.3795-3800, DOI: 10.1109/ICC.2015.7248915.
8. Chen, M.; Mozaffari, M.; Sadd, W.; Yin, C.; Debbah, M.; and Hong, C. Caching in The Sky: Proactive Deployment of Cache-Enabled Unmanned Aerial Vehicles for Optimized Quality-of-Experience. *Journal of Selected Areas in Communications* **2017** *35*, 1046-1061, DOI: 10.1109/JSAC.2017.2680898.
9. Merwaday, A.; and Guvenc, I. UAV assisted heterogeneous networks for public safety communications. In Proceedings of IEEE Wireless Communications and Networking Conference, New Orleans, LA, USA, March 9-12, 2015, pp. 329-334, DOI: 10.1109/WCNCW.2015.7122576.
10. Lokman, S.; Hakim, G.; Zouheir, R.; Mohamed-Slim, A. Energy-Efficient Power Allocation for UAV Cognitive Radio Systems. In Proceedings of IEEE 86th Vehicular Technology Conference, Toronto, ON, Canada, Septmber 24-27, 2017, pp.1-5, DOI: 10.1109/VTCFall.2017.8287971.
11. Rongfei, F.; Jiannan, C.; Song, J.; Kai, Y.; Jianping, A. Optimal Node Placement and Resource Allocation for UAV Relaying Network. *IEEE Communiation Letters* **2018**, *22*, 808-811, DOI: 10.1109/LCOMM.2018.2800737.
12. Guruacharya, S.; Niyato, D.; Kim, D.; et al. Hierarchical Competition for Downlink Power Allocation in OFDMA Femtocell Networks. *IEEE Transactions on Wireless Communications*, **2013**, *12*, 1543-1553, DOI: 10.1109/TWC.2013.022213.120016.
13. Zhang, H.; Xiao, Y.; Cai, L.; et al., Multi-Leader Multi-Follower Stackelberg Game for Resource Management in LTE Unlicensed. *IEEE Transactions on Wireless Communications*, **2017**, *16*, 348-361, DOI: 10.1109/TWC.2016.2623603.
14. Xiao, L.; Chen, T.; Liu, J.; et al. Anti-jamming Transmission Stackelberg Game with Observation Errors. *IEEE Communcation Letters*. **2015**, *19*, 949-952, DOI: 10.1109/LCOMM.2015.2418776.
15. Jia, L.; Yao, F.; Sun, Y.; et al. Bayesian Stackelberg Game for Anti-jamming Transmission with Incomplete Information. *IEEE Communcation Letters*, **2016**, *20*, 1991-1994, DOI: 10.1109/LCOMM.2016.2598808.
16. Yao, F.; Jia, L.; Sun, Y.; et al. A Hierarchical Learning Approach to Anti-Jamming Channel Selection Strategies. *Wireless Networks*, **2017**, *5*, 1-13, DOI: org/10.1007/s11276-017-1551-9.
17. Sastry, P.; Phansalkar, V.; Thathachar, M. Decentralized Learning of Nash Equilibrium in Multi-Person Stochastic Games with Incomplete Information. *IEEE Transactions on Systems*, **1994**, *24*, 769-777, DOI: 10.1109/21.293490.
18. Lyu, J.; Zeng, Y.; Zhang, R. Cyclical Multiple Access in UAV-Aided Communications: A Throughput-Delay Tradeoff. *IEEE Wireless Communication Letters*, **2016**, *99*, 600-603, DOI: 10.1109/LWC.2016.2604306.
19. Ponda, S.; Johnson, L.; Kopeikin, A.; Choi, H. Distributed Planning Strategies to Ensure Network Connectivity for Dynamic Heterogeneous Teams. *IEEE Journal on Selected Areas in Communications*, **2012**, *30*, 861-869, DOI: 10.1109/JSAC.2012.120603.
20. Li, J.; Han, Y. Optimal Resource Allocation for Packet Delay Minimization in Multi-Layer UAV Networks. *IEEE Communication Letters*, **2017**, *21*, 580-583, DOI: 10.1109/LCOMM.2016.2626293.
21. Zeng, Y.; Zhang, R.; and Lim, T. Wireless Communications with Unmanned Aerial Vehicles: Opportunities and Challenges. *IEEE Communications Magazine*, **2016**, *54*, 36-42, DOI: 10.1109/MCOM.2016.7470933.
22. Xue, Z.; Wang, J.; Shi, Q.; et al. Time-Frequency Scheduling and Power Optimization for Reliable Multiple UAV Communications. *IEEE Access*, **2018**, *99*, 3992-4005, DOI: 10.1109/ACCESS.2018.2790933.
23. Li, Y.; Cai, L.; Li, Y.; et al. UAV-Assisted Dynamic Coverage in a Heterogeneous Cellular System. *IEEE Network*, **2018**, *31*, 56-61, DOI: 10.1109/MNET.2017.1600280.
24. Challita, U.; Saad, W. Network Formation in the Sky: Unmanned Aerial Vehicles for Multi-Hop Wireless Backhauling. In Proceedings of IEEE Global Communications Conference, Singapore, Singapore, December 4-8, 2017, pp. 1-6, DOI: 10.1109/GLOCOM.2017.8254715.

25. Wang, B.; Han, Z.; Liu, K. Distributed Relay Selection and Power Control for Multiuser Cooperative Communication Networks Using Stackelberg Game. *IEEE Transactions on Mobile Computing*, **2009**, *8*, 975-990, DOI: 10.1109/TMC.2008.153.
26. Ozduran, V.; Soleimani-Nasab, E.; Yarman, B. S. Sum-rate Based Relay Selection with Feedback Delay for A Multiple Half/Full-Duplex Two-Way Relaying. In Proceedings of International ITS Telecommunications Conference, Warsaw, Poland, May 29-31, 2017, pp. 1-7, DOI: 10.1109/ITST.2017.7972196.
27. Monderer, D.; Shapley, L. S. Potential Games. *Games and Economic Behavior*, **1996**, *14*, pp. 124-143, DOI: org/10.1006/game.1996.0044.
28. Sun, Y.; Shao, H.; Liu, X. Traffic Offloading in Two-Tier Multi-Mode Small Cell Networks over Unlicensed Bands. *Ksii Transactions on Internet Information System*, **2016**, *9*, 4291-4310.
29. Zhong, W.; Tao, M.; et al. Game Theoretic Multimode Precoding Strategy Selection for MIMO Multiple Access Channels. *IEEE Signal Processing Letters*, **2010**, *17*, 563-566, DOI: 10.1109/LSP.2010.2047315.
30. Zheng, J.; Cai, Y.; Lu, N.; et al. Stochastic Game-Theoretic Spectrum Access in Distributed and Dynamic Environment. *IEEE Transactions on Vehicular Technology*, **2015**, *64*, 4807-4820, DOI: 10.1109/TVT.2014.2366559.
31. Xu, Y.; Wang, J.; Wu, Q.; et al. Opportunistic Spectrum Access in Unknown Dynamic Environment: A Game-Theoretic Stochastic Learning Solution. *IEEE Transactions on Wireless Communications*, **2012**, *11*, 1380-1391, DOI: 10.1109/TWC.2012.020812.110025.
32. Xu, Y.; Xu, Y.; Anpalagan, A. Database-Assisted Spectrum Access in Dynamic Networks: A distributed Learning Solution. *IEEE Access*, **2015**, *3*, 1071-1078, DOI: 10.1109/ACCESS.2015.2453266.
33. Sastry, P.; Phansalkar, V.; Thathachar, M. Decentralized Learning of Nash Equilibria in Multi-person Stochastic Games with Incomplete Information. *IEEE Transactions on Systems*, **1994**, *24*, 769-777, DOI: 10.1109/21.293490.
34. Yue, X.; Liu, Y.; Kang, S. l.; Nallanathan, A.; Chen, Y.; Modeling and Analysis of Two-Way Relay Non-Orthogonal Multiple Access Systems. *IEEE Transactions on Communications*, **2018**, *66*, 3784 - 3796, DOI: 10.1109/TCOMM.2018.2816063.