Data Collection Maximization in IoT-Sensor Networks Via an Energy-Constrained UAV

Yuchen Li, Weifa Liang, Senior Member, IEEE, Wenzheng Xu, Member, IEEE, Zichuan Xu, Member, IEEE, Xiaohua Jia, Fellow, IEEE, Yinlong Xu, and Haibin Kan

Abstract—In this paper, we study sensory data collection of IoT devices in a sparse IoT-sensor network, using an energy-constrained Unmanned Aerial Vehicle (UAV), where the sensory data is stored in IoT devices while the IoT devices may or may not be within the transmission range of each other. We formulate two novel data collection problems to fully or partially collect data stored in IoT devices using the UAV, by finding a closed tour for the UAV that consists of hovering locations and the sojourn duration at each of the hovering locations such that the accumulative volume of data collected per tour is maximized, subject to the energy capacity on the UAV, where the UAV consumes energy on both hovering for data collection and flying from one hovering location to another hovering location. To this end, we first propose a novel data collection framework that enables the UAV to collect sensory data from multiple IoT devices simultaneously if these IoT devices are within the coverage range of the UAV, through adopting the orthogonal frequency division multiple access (OFDMA) technique. We then formulate two data collection maximization problems to deal with full or partial data collection from IoT devices at each hovering location, and show that both defined problems are NP-hard. We instead devise approximation and heuristic algorithms for the problems. We finally evaluate the performance of the proposed algorithms through experimental simulations. Simulation results demonstrated that the proposed algorithms are promising.

Index Terms—Wireless sensor networks; a single UAV; approximation algorithms; energy-constrained optimization; UAV trajectory finding; collecting data from multiple sensors simultaneously; full and partial data collection; IoT applications; the orienteering problem.

1 INTRODUCTION

Due to its high flexibility, low cost and ease of deployment, the Unmanned Aerial Vehicle (UAV) has been a key enabling technology that has received significant attention recently, and been widely been applied in civilian environments such as natural disaster rescuing, good deliveries, crop health assessment, and so on [11]. On the other hand, with the increasing popularity of Internet of Things such as various sensors, wearable, traffic and other monitoring devices, more and more applications of smart homes/ smart cities, e-health care, and artificial transportations built upon IoT devices become part of our daily life. However, most IoT devices (e.g., mobile phones, security cameras, meter collection devices, temperature sensors) usually have very limited energy, computational and storage capacities due to their portable sizes. Sometimes, it is unrealistic to allow these devices to transmit or relay their sensing data to a base station through multihop relays due to the significant transmission energy consumptions. In the worst case, they may not be within the transmission ranges of each other. Thus, it is very challenging to collect sensing data from these IoT devices for processing on time to better help human decision-making and respond to the monitoring scenarios.

In this paper, we study sensory data collection from IoT devices on the ground using a UAV. Specifically, we consider a sparse sensor network that consists of many IoT devices for sensing their surroundings. Some of the IoT devices serve as aggregate sensor nodes to store their own and neighbors’ sensing data. The stored data at an aggregate sensor node will then be collected periodically by a UAV for further processing. As the volume of data stored at different aggregate sensor nodes are significantly different, the hovering durations of the UAV for data collection at different hovering locations are different, and the amounts of energy consumed by the UAV at different hovering locations thus are different too. In addition, when the UAV travels from one hovering location to another hovering location, it does also consume the energy consumption. As the energy capacity of the UAV is fixed, this poses a great challenge. That is, how to find a closed tour for the UAV including its depot for data collection such that the total volume of data collected by the UAV at different hovering locations in the tour is maximized, subject to its energy capacity. Furthermore, how long the UAV should stay at each of the hovering locations in the tour to ensure all (or partial) data stored at the IoT sensor devices under its coverage will be collected. To address this challenge, in this paper we aim to explore a non-trivial energy tradeoff between the amount of energy allocated to hovering and the amount of energy allocated to traveling of the UAV. We will focus on...
developing efficient heuristic and approximation algorithms for the data collection optimization problems.

The novelties of this work lie in the provisioning of a novel framework of data collection from multiple IoT sensor devices simultaneously via an energy-constrained UAV. We formulate two data collection maximization problems and develop efficient approximation and heuristic algorithms for the problems that strive for a fine tradeoff between the energy usages of the UAV on hovering and traveling, respectively. To the best of our knowledge, this is the first time that the use of a UAV for full or partial data collection from multiple IoT devices simultaneously is thoroughly studied. Efficient algorithms for finding a data collection trajectory for the UAV is developed.

The main contributions of this paper are summarized as follows. We study the full and partial data collection maximization problems, by deploying an energy-constrained UAV. We first propose a novel data collection framework that enables the UAV to collect data from multiple IoT devices simultaneously. We then formulate two data collection maximization problems to fully or partially collect data stored from IoT devices based upon the proposed data collection framework, and show that the two defined problems are NP-hard. We instead devise efficient approximation and heuristic algorithms for the problems. We finally evaluate the performance of the proposed algorithms through experimental studies. Simulation results reveal that the proposed algorithms are promising.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 introduces the system model, notions, and notations. Section 4 presents the problem definitions and shows the NP-hardness of the defined problems. Section 5 devises approximation algorithms for the full and partial data collection maximization problems without hovering coverage overlapping when the problem size is large. Section 6 proposes efficient heuristic algorithms for the full and partial data collection maximization problems with hovering coverage overlapping. Section 7 evaluates the proposed algorithms empirically, and Section 8 concludes the paper.

2 RELATED WORK

The use of mobile charging vehicles or mobile data collection vehicles on the ground for sensor charging and sensory data collection has been widely studied in the past [13], [14], [16], [27], [25], [26]. Most studies focused on finding the trajectories of charging or data collection for one or multiple mobile vehicles. However, due to various obstacles in the sensor networks including building, ponds, rivers block roads in the monitoring region, mobile vehicles cannot travel in the region for sensor charging or data collection smoothly. In recent years, there is growing interest in the employment of UAVs for sensory coverage and data collection for wireless sensor networks, as UAVs have freedoms for data collection while avoiding the mentioned obstacles by flying over them [17], [15], [18]. For example, Mozaffari et al. [17] considered trajectory path findings for multiple UAVs with the aim to minimize the total transmission energy consumption of IoT devices to upload their data to the UAVs, where the UAVs are treated as aerial base stations. They proposed a clustering method to cluster IoT devices into different clusters and then find trajectories for multiple UAVs that sojourn only at the cluster centers. However, the clustering method only minimizes the sum of distances between each sojourn location of the UAV and device locations, it does not consider the distances among sojourn locations. Zhan et al. [29] jointly considered working states of sensors and the trajectory of the UAV by utilizing a fading channel model for the sensor-UAV links to minimize the maximum energy consumption of sensors. Chorbet et al. [9] focused on the identification of hovering locations of a single UAV for data collection from a cluster of sensors with the aim to minimize the energy consumption of the UAV. Liang et al. [15] considered a coverage quality problem via a UAV. They assumed that the hovering time of the UAV at each hovering location is identical, for which they proposed an approximation algorithm for the coverage quality maximization problem. Binol et al. [3] aimed at finding time-optimal paths for multiple UAVs for data collection. Zhan and Zeng [28] studied the time minimization problem of data collection, using multiple UAVs. Sikertidis and Tsiropoulou et al. [21] proposed a novel framework that improves the energy efficiency in a UAV-supported Public Safety Networks. Guo et al. [10] considered the problem of a fleet of UAVs for disaster surveillance by finding trajectories for the UAVs such that the longest time among the UAVs is minimized. Recently Zhang et al. [30] investigated how to minimize the number of UAV deployment when the tour duration of each UAV is bounded by a budget, for which they proposed constant approximation algorithms under different data collection models. Samir et al. [20] considered data collection with data uploading deadlines, where a UAV is dispatched for data collection from time-constrained devices and each device has a data uploading deadline, they aimed to maximize the network throughput. Chen et al. [5], [6] considered the data collection maximization problem without considering traveling energy consumption of a UAV, they devised approximation algorithms for the problem, they also considered the data collection under different data transmission rate by proposing a heuristic algorithm [7].

The essential differences between the work in the paper and the aforementioned studies lie in the following two aspects. One is that we consider the partial or full data collection from multiple IoT devices simultaneously by the deployment of a UAV that has not been considered previously. Another is that an approximate solution for the problem without hovering coverage overlapping is developed, which is the first algorithm with a provable performance guarantee for the problem. It must be mentioned that this paper is an extension of its conference version [12], which contains the approximate solution to the partial data collection maximization problem without hovering coverage overlapping that has not been covered in its conference version.

3 PRELIMINARIES

In this section, we introduce the system model, notions and notations.
3.1 System model

We consider an IoT application scenario where many different types of IoT devices are deployed in a given region for monitoring purposes, some of the IoT devices (sensors) are chosen as aggregate sensor nodes that can store both their own sensing data and their neighbors’ sensory data, assuming those IoT devices that have not been chosen as aggregate sensor nodes can forward their sensory data to one of their neighboring aggregate sensor nodes. In case there are multiple aggregate sensor neighbors, it can choose one of such neighbors for the storage of its sensory data. Since aggregate sensor nodes are sparsely distributed, they may or may not be within the communication range with each other. The sensory data collected at each aggregate sensor node thus cannot be transferred to the base station through multi-hop relays, or there are obstacles between the aggregate sensor nodes, e.g., ponds, or buildings that prevent the relays. Furthermore, they are energy-constrained on aggregate sensor nodes, e.g., ponds, or buildings that prevent the relays. The UAV takes hovering locations for data collection from aggregate sensor nodes in its hovering coverage range and traveling (flying) from one hovering location to another hovering location. For the sake of convenience, in the rest of this paper we term the aggregate sensor nodes as IoT devices or sensor nodes interchangeably if no confusion arises. The aggregate sensor nodes in a monitoring region form a sparse sensor network \( G = (V \cup \{d\}, E) \), where \( V \) is the set of aggregate sensor nodes, and there is an edge \( e \in E \) between each pair of aggregate sensor nodes. The depot of the UAV is \( d \) in which the UAV will be recharged and its collected data will be downloaded for further processing.

To ensure that the UAV can return the depot \( d \) after each tour, its data collection tour must be a closed tour that includes depot \( d \). The duration of a tour of the UAV will be determined by the tour length and the volume of data stored in the IoT devices covered by the UAV at each hovering location in the tour. Assume that the UAV takes \( T \) time units to finish its tour, in which \( T_h \) and \( T_t \) are the amounts of time spent on hovering and traveling, respectively, then \( T = T_h + T_t \) and the total amount of energy consumed by the UAV in the tour must meet that \( T_h \cdot \eta_h + T_t \cdot \eta_t \leq \mathcal{E} \), where \( \eta_h \) and \( \eta_t \) are the energy consumption rates of the UAV on hovering and traveling, respectively [22].

3.2 Data collection framework from IoT devices using a UAV

We assume that each aggregate sensor node in a monitoring region is labeled by its coordinates \((x_i, y_i, 0)\). Denote by \((x', y', H)\) the coordinates of a hovering location of the UAV, where \( H \) is the flying altitude of the UAV, which is no greater than the transmission range \( R \) of each aggregate sensor node, and \( B \) is the data transmission rate of any aggregate sensor node. An aggregate sensor node \( v \in V \) can transmit its stored data to the UAV with the data transmission rate \( B \) if the UAV is within its transmission range \( R \). The hovering altitude \( H \) of the UAV for data collection thus is no greater than \( R \), i.e., \( H \leq R \), and we further assume that the altitude \( H \) of the UAV does not change [1].

Following the data collection model that if all IoT devices within the reception range of the UAV, then their transmitted data can be collected by the UAV, assuming that each IoT device uses a different communication channel [17]. We assume that the reception range of the UAV is a ball centered at its hovering location, this ball is projected to the ground where the IoT devices are located to form a circle with the same radius of the ball, the data of all aggregate sensor nodes within the projected circle can be collected by the UAV when it hovers at the center of the ball, through the use of the orthogonal frequency division multiple access (OFDMA) technique [17].

Since there are infinite numbers of hovering locations for the UAV in the sky, to make the problem tractable, we assume that the hovering locations of the UAV are finite by partitioning its hovering region into finite numbers of equal squares with edge length \( \delta > 0 \). In the rest of discussion, we assume that the measurement unit of the UAV movement in its hovering region is the edge length \( \delta \) of each square. Or the coordinates of potential hovering locations within a square are indistinguishable. For the sake of convenience, we further assume that the center of each square is the potential hovering location of the UAV within that square area. For a given data collection period \( T \), assume that the volume \( D_v \) of data stored at each aggregate sensor node \( v \in V \) is given, which consists of its own sensory data and the data forwarded by its neighboring IoT devices that have not been chosen as aggregate sensor nodes. Fig. 1 is an illustrative example of an aggregate sensor network \( G \) with a UAV for data collection.

We assume that the hovering region of a UAV is partitioned into \( M \) squares: \( s_1, s_2, \ldots, s_M \), and the UAV performs data collection at the centers of these squares \( s_j = (x_j, y_j, H) \) as its hovering locations, let \( C(s_j) \) be the set of aggregate sensor nodes whose distances to the projected location \((x_j, y_j, 0)\) of \( s_j \) on the ground are no more than the data reception range \( R_0 \) of the UAV, i.e., the data stored at an aggregate sensor node \( v_j \in V \) with coordinates \((x_i, y_i, 0)\) can be collected by the UAV if it is within the ball of the UAV centered at hovering location \( s_j \), i.e., \( \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq R_0 \) and \( R_0 = \sqrt{R^2 - H^2} \) (see Fig. 1 (b)). For any two potential hovering locations \( s_i \in M \) and \( s_j \in M \) with \( i \neq j \), if \( C(s_i) \cap C(s_j) = \emptyset \), we refer to the set \( M \) as a set of hovering locations without hovering coverage overlapping; otherwise, we refer to the set \( M \) as a set of hovering locations with hovering coverage overlapping. Notice that the data transmission duration of an aggregate sensor node from the ground to the UAV usually is determined by its distance to the UAV, its data volume, and its data transmission rate. Given two aggregate sensor nodes at different locations in the coverage circle of the UAV, it is well known that their data transmission durations and the data transmission rate will be different even if they have the same amounts of data to be transmitted. However, such differences between them are negligible if the UAV altitude \( H \) is relatively low (not too high). For the sake of discussion simplicity, in this paper we assume that all sensors within
the hovering coverage range of the UAV will have the same data transmission rate $B$.

The hovering (sojourn) duration of the UAV for data collection at hovering location $s_j$ thus is

$$t(s_j) = \max_{v_i \in V} \left\{ \frac{D_{v_i}}{B} \mid \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq R_0 \right\},$$

(1)

the total volume of data collected at hovering location $s_j$ is

$$V(s_j) = \sum_{v_i \in V} \left\{ \frac{D_{v_i}}{B} \mid \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq R_0 \right\},$$

(2)

and the total amount of energy consumption on data collection by the UAV at $s_j$ is

$$w_1(s_j) = t(s_j) \cdot \eta_h.$$  

(3)

### 3.3 Approximation ratio

Given an instance of a maximization optimization problem $P$, let $A$ and $OPT$ be a feasible solution delivered by an approximation algorithm and the optimal solution of the problem respectively if $\frac{A}{OPT} \geq \frac{1}{\alpha}$ with $\alpha > 1$, then $\frac{1}{\alpha}$ is referred to the approximation ratio of the approximation algorithm for $P$.

### 4 Problem Formulations and NP-hardness of the Defined Problems

In this section, we define two novel data collection maximization problems using a single UAV, under the assumptions that (i) sensory data from multiple IoT devices can be collected simultaneously by a UAV at a hovering location; and (ii) the data stored at each IoT device can be either fully collected or partially collected by the UAV at each hovering location. Both problems are defined as follows.

**Definition 1.** Given an aggregate sensor network $G(V \cup \{d\}, E)$ and a UAV with energy capacity $E$, each aggregate sensor node $v \in V$ has a volume $D_v$ of data for collection, the full data collection maximization problem is to find a closed tour for the UAV such that the accumulative volume of data collected by the UAV at all hovering locations in the closed tour is maximized, subject to the energy capacity $E$ on the UAV, assuming that the data stored at each IoT device in the hovering coverage of the UAV will be fully collected.

We deal with the full data collection maximization problem under two different restrictions. That is, whether the hovering coverage ranges of the UAV at any two different hovering locations are allowed to be overlapping. We thus have two cases for the problem: the full data collection maximization problem without hovering coverage overlapping and the full data collection maximization problem with hovering coverage overlapping, respectively.
energy units, while the total energy consumption on the partial data collection at the two locations is 

\[
(10 + 6) \eta_n
\]

overlapping between any two hovering locations. Even for this special data collection maximization problem, we show that it is equivalent to an orienteering problem in \( G \) as follows.

Given a node- and edge-weighted, undirected graph \( G(V, E) \), in which each node \( v \in V \) has a positive award \( p(v) = D_v \) and each edge \( (u, v) \in E \) has a positive integral length \( l(u, v)/m \), and a given integral length \( L \), the orienteering problem is to find a closed tour in \( G \) including a specified node (the depot \( d \)) such that the total award collected from the nodes in the closed tour is maximized, subject to the tour length no greater than \( L \) [2], [4], [24].

We reduce the orienteering problem in \( G \) to this special full data collection maximization problem as follows. The award collected at each hovering location \( s_j \in S \) (the coordinates of \( s_j \) are \( (x_j, y_j, H) \) assuming that the coordinates of \( v_j \in V \) are \( (x_j, y_j, 0) \)) is \( V(s_j) = D_{v_j} \), \( L = [\frac{E}{2}] \), and the hovering energy consumption at each potential hovering location \( s_j \in S = V \) by the assumption) is zero. As the orienteering problem is NP-hard [24], the data collection maximization problem is NP-hard.

The full data collection maximization problem is a special case of the partial data collection maximization problem when \( K = 1 \), the latter is NP-hard, too.

5 Approximation Algorithms for the Full and Partial Data Collection Maximization Problems Without Hovering Coverage Overlapping

In this section, we deal with the full and partial data collection maximization problems without hovering coverage overlapping. We first formulate an Integer Programming solution for the problem when the problem size is small or moderate. We then propose approximation algorithms for the problems under the assumption that different hovering locations of the UAV within each square are indistinguishable. We finally analyze the correctness of the solutions and the time complexity of the proposed algorithms.

5.1 ILP formulation

In the following we formulate an integer programming solution (ILP) to the full data collection maximization problem without hovering coverage overlapping. Given the aggregate sensor network \( G = (V, E) \) and the UAV, we partition the hovering region of the UAV into \( M \) squares, denoted by

\[
S = \{s_1, s_2, \ldots, s_M\} \cup d
\]

Let \( x_j \) be indicator variables which determine whether the UAV will hover at square \( s_j \) and collect the data, and let \( y_{i,j} \) be the indicator variables which determine whether the UAV flies from square \( s_i \) to \( s_j \) and \( y_{0,1} \) and \( y_{0,0} \) indicates if the UAV flies from depot \( d \) to \( s_i \) and if the UAV flies from \( s_j \) to depot \( d \), respectively, where

\[
1 \leq i, j \leq M \text{ and } i \neq j.
\]

We formulate the ILP solution as follows:

\[
\text{Maximize } \sum_{j=1}^{M} V(s_j) \cdot x_j
\]
subject to:

\[
(1), (2) \\
x_j = \sum_{i=0}^{M} y_{i,j} \\
\sum_{i=1}^{M} y_{i,j} \leq 1, \quad \forall j \\
\sum_{j=1}^{M} y_{i,j} \leq 1, \quad \forall i \\
\sum_{i=0}^{M} y_{i,j} = \sum_{i'=0}^{M} y_{i',j} \quad j \neq 0 \\
\sum_{i=0}^{M} \sum_{j=0}^{M} y_{i,j} \cdot \frac{t(s_i,s_j)}{\nu} \cdot \eta_t + \sum_{j'=1}^{M} t(s_j') \cdot \eta_h \leq \mathcal{E} \\
\sum_{i=1}^{M} y_{i,0} = 1 \\
\sum_{j=1}^{M} y_{0,j} = 1 \\
\sum_{s_i,s_j \in S'} y_{i,j} \leq |S'| - 1, \forall S' \subseteq S \setminus \{d\}, \quad S' \neq \emptyset \\
x_j \in \{0, 1\} \quad \forall s_j \in S \\
y_{i,j} \in \{0, 1\} \quad \forall s_i, s_j \in S \cup \{d\},
\]

where \( V(s_j) \) and \( t(s_j) \) are defined in Eq. (2), and Eq. (1), \( t(s_i,s_j) \) is the Euclidean distance between \( s_i \) and \( s_j \).

The objective (6) is to maximize the sum of profits of nodes on the data collection tour. Constraint (7) restricts that the UAV can only hover at \( s_j \) if \( s_j \) is on the flying path. Constraints (8) and (9) ensure that each hovering location can be reached or left once. Constraint (10) ensures that if the UAV arrives at node \( s_j \), it must leave from node \( s_j \). Constraint (11) enforces that the sum of the energy consumption is no greater than the UAV energy capacity \( \mathcal{E} \). Constraints (12) and (13) enforce that the UAV must start from depot \( d \) and end at \( d \). Constraint (14) prevents subtours disconnected from the subtour containing depot \( d \). Constraints (15) and (16) restrict the ranges of decision variable \( x_j \) and \( y_{i,j} \) to be either 0 or 1.

5.2 Approximation algorithm for the full data collection maximization problem without hovering coverage overlapping

The basic idea behind the proposed algorithm is to reduce the problem to the orienteering problem [24]. The challenge of such a reduction lies in that the hovering energy consumptions of the UAV at different hovering locations are different, we aim to find a closed tour including depot \( d \) such that the accumulative volume of data collected by the UAV at all hovering locations in the tour is maximized, while the total amount of energy consumed of the UAV on both hovering and traveling is no greater than its energy capacity \( \mathcal{E} \). We address this challenge by finding a closed tour in an auxiliary graph in which each edge is assigned an energy weight for both hovering at the endpoints of the edge and traveling along the edge as follows.

Since there are infinite numbers of potential hovering locations in the hovering plane of the UAV, to enable the problem to be tractable, we partition the hovering region of the UAV (or the corresponding IoT device deployment region) into a number of squares with edge length \( \delta > 0 \). We assume that the distance differences among potential hovering locations in each square are negligible if the value of \( \delta \) is sufficiently small. We further assume that the center of each square is a potential hovering location for the UAV when it is in the square.

Having partitioned the hovering region of the UAV into \( M \) squares, we now construct a node and edge weighted, undirected graph \( G_s = (S, E_s; p(\cdot), w_1(\cdot), w_2(\cdot, \cdot)) \) as follows. \( S \) is the set of potential hovering locations of the UAV, and \( E_s \) is the set of edges that the UAV hovering from one hovering location to another hovering location. The functions related to nodes and edges in \( G_s \) are defined as follows. \( p : S \rightarrow \mathbb{R}^{\geq 0} \) is the award function, \( w_1 : S \rightarrow \mathbb{R}^{\geq 0} \) is the hovering energy consumption function, and \( w_2 : E_s \rightarrow \mathbb{R}^{\geq 0} \) is the energy consumption function on both hovering and traveling. There is an edge \( (s_i, s_j) \in E_s \) between each pair of nodes \( s_i \) and \( s_j \) in \( S \).

For each potential hovering location \( s_j \in S \), the award (the amount of data collected) is

\[
p(s_j) = \sum_{v_i \in C(s_j)} D_{v_i},
\]

where \( C(s_j) \) is the set of IoT devices in the hovering coverage range of the UAV when it hovers at hovering location \( s_j \), i.e., \( C(s_j) = \{ v_i \mid v_i \in V \land \sqrt{(x_i-x_j)^2 + (y_i-y_j)^2} \leq R_0 \} \), assuming that the UAV is at location \( s_j = (x_j, y_j, H) \). If \( C(s_j) = \emptyset \), then \( p(s_j) = 0 \), \( t(s_j) = 0 \) and \( w_1(s_j) = 0 \). \( D_{v_i} \) is the volume of data stored at IoT device \( v_i \) that is a function of the monitoring duration \( T \) and the sensing data generation rates of neighboring sensors of \( v \) within the monitoring period.

The hovering duration \( t(s_j) \) of the UAV at hovering location \( s_j \) for collecting data from IoT devices in \( C(s_j) \) is

\[
t(s_j) = \max_{v_i \in C(s_j)} \left\{ \frac{D_{v_i}}{B} \right\},
\]

where \( B \) is the data transmission rate of an aggregate sensor node (an IoT device).

The amount of energy consumption of data collection by the UAV at hovering location \( s_j \) thus is

\[
w_1(s_j) = t(s_j) \cdot \eta_h.
\]

We assign a weight \( w_2(s_j, s_k) \) to each edge \( (s_j, s_k) \in E_s \) as follows.

\[
w_2(s_j, s_k) = \frac{w_1(s_j) + w_1(s_k)}{2} + \frac{l(s_j, s_k) \cdot \eta_h}{\nu}
\]

where the first term in the right hand side of Eq. (20) is half the sum of the amounts of hovering energy consumed by the UAV for data collection at locations \( s_j \) and \( s_k \) respectively, the second term is the amount of traveling energy consumption of the UAV along edge \( (s_j, s_k) \), and \( l(s_j, s_k) \) is the Euclidean distance between locations \( s_j \) and \( s_k \).

Having the constructed auxiliary graph \( G_s(S, E_s; p(\cdot), w_1(\cdot), w_2(\cdot, \cdot)) \), the orienteering problem in \( G_s \) is to find a closed tour including depot \( d \) such that
the total award collected from the hovering locations in the tour is maximized, subject to that the tour length (measured in terms of energy) is no greater than the energy capacity $\mathcal{E}$ of the UAV. It can be seen that a solution to the orienteering problem in $G_s$ returns a solution to the full data collection maximization problem in $G$ without hovering coverage overlapping.

The detailed algorithm for the full data collection maximization problem is given in Algorithm 1.

**Algorithm 1** Approximation algorithm for the full data collection maximization problem without hovering coverage overlapping  

**Input:** An aggregate sensor network $G = (V,E)$ with a set $V$ of aggregate sensor nodes, a UAV with energy capacity $\mathcal{E}$ at depot $d$, and each node $v \in V$ has data volume $D_v$, and a given constant $\delta > 0$ but $\delta \leq R_0$.  

**Output:** Find a closed tour including the depot $d$ for the UAV such that the volume of data collected from all aggregate sensor nodes covered by the UAV at hovering locations in the tour is maximized, subject to the energy capacity of the UAV.

1. Partition the monitoring region into $M$ squares $s_1,s_2,\ldots,s_M$ with the edge length of each square being $\delta$; let $S = \{d,s_1,s_2,\ldots,s_M\}$;
2. Compute $t(s_j), p(s_j)$, and $w_1(s_j)$ for each $s_j \in S$ with $1 \leq j \leq M$;
3. Construct an auxiliary graph $G_s = (S \cup \{d\},E_s \cup \{(v,d') \mid (v,d) \in E_s\};p(\cdot),w_1(\cdot),w_2(\cdot))$, where $d'$ is a dummy depot;
4. Find a simple path $P$ in $G_s$ between the depot $d$ and the dummy depot $d'$ such that the total award collected in the path is maximized (as there is no coverage overlapping between any two hovering locations by the assumption), subject to the energy capacity $\mathcal{E}$ of the UAV, by the approximation algorithm for the orienteering problem in metric graphs [2];
5. return A closed tour $C$ derived from $P$ for the UAV, which contains the hovering locations and the sojourn time at each of the hovering locations.

5.3 Approximation algorithm for the partial data collection maximization problem without hovering coverage overlapping  

We now deal with the partial data collection maximization problem by reducing the problem to the orienteering problem as well, and by adopting the proposed algorithm Algorithm 1 with minor modifications. Specifically, for a given integer $K \geq 1$ and each potential hovering location of the UAV, the $K$ virtual hovering locations are created for the potential hovering location, i.e., $K$ virtual hovering locations $s_{j,1},s_{j,2},\ldots,s_{j,K}$ are generated for each potential hovering location $s_j \in S$. Recall that $t(s_j)$ the hovering duration at $s_j$ for all sensor data collection of sensors under this hovering coverage, we divide this duration into $K$ equal time durations, and assign a hovering duration $t(s_{j,k}) = \frac{t(s_j)}{K}$ at virtual hovering location $s_{j,k}$, i.e.,

$$t(s_{j,k}) = \frac{t(s_j)}{K}, \quad \forall k \in \{1,2,\ldots,K\} \quad (21)$$

The volume $V'(s_{j,k})$ of data collected by the UAV at $s_{j,k}$ with duration $t(s_{j,k})$ is

$$V'(s_{j,k}) = \sum_{v \in C(s_j)} \min\{\frac{B \cdot t(s_j)}{K}, D_v - \frac{B \cdot (k-1) \cdot t(s_j)}{K}\}$$

$$\frac{D_v}{B} > (k-1) \cdot t(s_j)/K. \quad (22)$$

It can be seen that $V'(s_{j,1}) \geq V'(s_{j,2}) \geq \ldots \geq V'(s_{j,K})$ because some sensors in $C(s_j)$ have no data for transmissions with the time progress.

Let $S'$ be the set of all virtual potential hovering locations derived from all potential hovering locations, while the latter is obtained the partitioning of the hovering region into a number of squares as we did for the full data collection maximization problem. An auxiliary graph $G_s$ similar to $G_s$ in the previous subsection, is then constructed. It can be shown that $G_s$ is a metric graph as well, an approximate solution to the corresponding orienteering problem in $G_s'$ can be found, by applying Algorithm 1, and an approximate solution - a closed tour $C'$ then is delivered. However, tour $C'$ may not be a feasible solution to the partial data collection maximization problem, since there is very likely to have a node $s_{j,k} \in C'$ but $s_{j,k'} \not\in C'$ with $k > k'$. Meanwhile, we know that there are no distinctions between node $s_{j,k}$ and $s_{j,k'}$ in terms of their locations and hovering durations as both of them are derived from the same node $s_j$. We also know that $V'(s_{j,k}) \leq V'(s_{j,k'})$. Thus, a better solution $C''$ to the problem can be obtained, by replacing node $s_{j,k}$ in $C'$ with node $s_{j,k'}$. Now, assume that there are $l \geq 2$ nodes from $s_j$ included in tour $C''$, denoted by $s_{j,i_1},s_{j,i_2},\ldots,s_{j,i_l}$ with $1 \leq i_1 \leq K$ and $1 \leq i_l \leq K$, they will be replaced by nodes $s_{j,1},s_{j,2},\ldots,s_{j,l}$ respectively, and the resulting solution still is a feasible solution to the problem, this node replacement procedure continues until no further replacement is needed. An approximate solution to the problem is obtained in the end. We term this algorithm as Algorithm 2.

5.4 Algorithm analysis  

In the following we show the correctness and time complexity of the proposed algorithms. We first show that the auxiliary graph $G_s$ is a metric graph, as the given approximation algorithm for the orienteering problem is only applicable to metric graphs, we then analyze the time complexity of the two proposed approximation algorithms and their approximation ratios. Notice that both Algorithm 1 and Algorithm 2 are approximation algorithms under the assumption that there is no distinction among different hovering locations in each square; otherwise, the problems are intractable due to infinite numbers of potential hovering locations even for a small square area.

**Lemma 1.** The auxiliary graph $G_s$ is a metric graph.

**Proof:** Since there is an edge for each pair of nodes in $G_s$, we show that the edge weights in $G_s$ meet the triangle
Given an aggregate sensor network $G = (V, E)$ with a set $V$ of aggregate sensor nodes, a UAV with energy capacity $\mathcal{E}$ at depot $d$, and each node $v \in V$ has data volume $D_v$, and a given constant $\delta > 0$ but $\delta \leq R_0$.

**Output:** Find a closed tour including the depot $d$ for the UAV such that the volume of data collected from all aggregate sensor nodes covered by the UAV at hovering locations in the tour is maximized, subject to the energy capacity of the UAV.

1. Partition the monitoring region into $M$ squares $s_1, s_2, \ldots, s_M$ with the edge length of each square being $\delta$; let $S = \{d, s_1, s_2, \ldots, s_M\}$;
2. $S' \leftarrow \bigcup_{s \in S} \{s_j, s \mid s_j \in S\}$;
3. Compute $t(s_{j,k}), p(s_{j,k})$, and $w_1(s_{j,k})$ for each $s_{j,k} \in S$ with $1 \leq j \leq M$;
4. Construct an auxiliary graph $G'_s = (S', \bigcup\{d', E_v \cup \{(v, d') \mid \forall v \in E_v\}; p(\cdot), w_1(\cdot), w_2(\cdot), \cdot\}$, where $d'$ is a dummy depot;
5. Find a simple path $P$ in $G'_s$ between the depot $d$ and the dummy depot $d'$ such that the total award collected in the path is maximized (as there is no coverage overlapping between any two hovering locations by the assumption), subject to the energy capacity $\mathcal{E}$ of the UAV, by the approximation algorithm for the orienteering problem in metric graphs [2];
6. for each $j$ from 1 to $M$
7. if $\exists k, k' \in S$, then $k > k'$, $s_{j,k} \in P, s_{j,k'} \notin P$ then
8. replace node $s_{j,k}$ by node $s_{j,k'}$
9. end if
10. end for
11. return A closed tour $C$ derived from $P$ for the UAV, which contains the hovering locations and the sojourn duration at each of the hovering locations.

Inequality. For the three edges formed by any three nodes $s_j, s_k$, and $s_l$ in $S$, we have

$$w_2(s_j, s_k) + w_2(s_k, s_l)$$

$$= \left(\frac{w_1(s_j) + w_1(s_k)}{2} + \frac{l(s_j, s_k) \cdot \eta_1}{\nu}\right) + \left(\frac{w_1(s_k) + w_1(s_l)}{2} + \frac{l(s_k, s_l) \cdot \eta_1}{\nu}\right)$$

$$= \frac{w_1(s_j) + w_1(s_l)}{2} + w_1(s_k) + \left(\frac{l(s_j, s_k) + l(s_k, s_l)}{\nu}\right) \cdot \eta_1$$

$$\geq \frac{w_1(s_j) + w_1(s_l)}{2} + \left(\frac{l(s_j, s_l)}{\nu}\right) \cdot \eta_1$$

$$= w_2(s_j, s_l).$$

Thus, $G_s$ is a metric graph. □

**Theorem 2.** Given an aggregate sensor network $G(V, E)$ with each node $v \in V$ having a data volume $D_v$, for collection, and a UAV with energy capacity $\mathcal{E}$ and its depot $d$, assuming that the moving unit of the UAV is measured by a value of $\delta > 0$ and its coverage range at each hovering location is a circle with radius $R_0$, there is a $\frac{1}{2}$-approximation algorithm, Algorithm 1, for the full data collection maximization problem in $G$ without hovering coverage overlapping, assuming that the distance difference among potential hovering locations within each square is negligible. The algorithm takes $O(T_{ort}(\frac{\pi R_0}{\delta}, |V|, \frac{\pi^2 R_0^2}{\delta} \cdot V^2))$ time, where $T_{ort}(|V'|, |E'|)$ is the time complexity of the approximation algorithm of Bansal et al. [2] for the orienteering problem in a graph with $|V'|$ nodes and $|E'|$ edges.

**Proof:** We first show that the solution obtained by Algorithm 1 is feasible. It can be seen that the closed tour $C$ is a simple closed tour including the depot $d$. We show that the total energy consumption of the UAV on the closed tour $C$ is no greater than $\mathcal{E}$. As the total length of $C$ is no greater than $\mathcal{E}$, the energy consumption of the UAV on $C$ (homing at nodes and traveling on edges) is the weighted sum of energies in $C$, which is no greater than its energy capacity. Furthermore, for each hovering location $s_j$ in $C$, assume that $s_{j1}$ and $s_{j2}$ are its two neighboring hovering locations in $C$, then the hovering energy consumption $w(s_j)$ of the UAV at location $s_j$ is distributed to its two incident edges $(s_{j1}, s_j)$ and $(s_j, s_{j2})$ as their parts, i.e., the energy weights of the two edges are $w_2(s_{j1}, s_j) = \frac{w_1(s_{j1}) + w_1(s_j)}{\nu} + \frac{l(s_{j1}, s_j) \cdot \eta_1}{\nu}$ and $w_2(s_j, s_{j2}) = \frac{w_1(s_j) + w_1(s_{j2})}{\nu} + \frac{l(s_j, s_{j2}) \cdot \eta_1}{\nu}$, respectively. Thus, the UAV has sufficient energy at each hovering location $s_j$ to collect all data from the IoT devices in $C(s_j)$.

We then analyze the time complexity of the proposed algorithm, Algorithm 1. We notice that the number of squares, $M$, is a linear function of the number of aggregate sensor nodes in $V$. For example, assume that the coverage range of the UAV at a hovering location is a circle with radius $R_0$, then, the number of its potential hovering locations for collecting data from an IoT device $v \in V$ will be no greater than $\lceil \frac{\pi R_0^2}{\delta} \rceil$ in terms of the number of squares covering $v$. Thus, the maximum number of squares in $G_s$ is no greater than $\sum_{v \in V} \lceil \frac{\pi R_0^2}{\delta} \rceil \leq \frac{\pi R_0^2}{\delta} \cdot |V|$ as both $R_0$ and $\delta$ usually are constants. There exists an edge between every pair of the node in $G_s$, therefore, $G_s$ contains $\mathcal{E}$ = $O(|S|^2) = O(\frac{\pi^2 R_0^2}{\delta} \cdot |V|^2)$ edges. Finding a $\frac{1}{2}$-approximate solution (a closed tour $C$) for the orienteering problem in $G_s$ starting at node $d$ takes $O(T_{ort}(\frac{\pi R_0}{\delta}, |V|, \frac{\pi^2 R_0^2}{\delta} \cdot |V|^2))$ time, by the approximation algorithm due to Bansal et al. [2], assuming that the distances of hovering locations within each square are negligible, where $T_{ort}(|V'|, |E'|)$ is the time complexity of the approximation algorithm of Bansal et al. [2] for the orienteering problem in a graph with $|V'|$ nodes and $|E'|$ edges. □

We finally show that the solution delivered by Algorithm 2 for the partial data collection maximization problem is feasible, and analyze its approximation ratio as follows.

**Theorem 3.** Given an aggregate sensor network $G(V, E)$ with each node $v \in V$ having a data volume $D_v$, for collection, and a UAV with energy capacity $\mathcal{E}$ and its depot $d$, assuming that the moving unit of the UAV is measured by a value of $\delta > 0$ and its coverage range at each hovering location is a circle with radius $R_0$, there is a $\frac{1}{2}$-approximation algorithm, Algorithm 2, for the partial data collection maximization problem in $G$ without hovering coverage overlapping, assuming that the
distance difference among potential hovering locations within each square is negligible. The algorithm takes $O(T_{ort}(\frac{R_0^2}{\delta^3}, |K \cdot V|, \frac{R_0^2}{\delta^3}, |K \cdot V|^2))$ time, where $T_{ort}(|V'|, |E'|)$ is the time complexity of the approximation algorithm of Bansal et al. [2] for the orienteering problem in a graph with $|V'|$ nodes and $|E'|$ edges.

Proof: The proof of Theorem 3 is almost identical to the one for Theorem 2, and it can be shown that the auxiliary graph $G_s$ is also a metric graph. The only difference lies in the analysis of the approximation ratio of the approximation algorithm. Assume that there is an optimal solution for the partial data collection maximization problem $OPT_K$, as the solution $A_K$ delivered by the approximation algorithm is $OPT_K/3$, while a feasible solution to the problem is a solution by replacing all virtual nodes derived from a hovering location node with its first virtual nodes, and the accumulative volume of the removed virtual nodes is no more than the accumulative volume of the added virtual nodes, and the total volume of data collected in the resulting tour is no less than that of the initial solution. Thus, the solution is no less than one third of the optimal one.

The proof of time complexity is similar to the one in the proof body of Theorem 3, omitted.

6 Heuristic algorithm for the full and partial data collection maximization problems with hovering coverage overlapping

In this section, we deal with the full and partial data collection maximization problems with hovering coverage overlapping, by proposing efficient heuristic algorithms for them. We also conduct the computing complexity analysis of the proposed algorithms.

6.1 Algorithm for the full data collection maximization problem with hovering coverage overlapping

The basic idea behind the proposed algorithm is to find a closed tour for the UAV iteratively. Initially, the closed tour consists of only the depot. Within each iteration, a new hovering location is added to the tour. For the sake of convenience, we assume that a partially closed tour that consists of hovering locations $s_0, s_1, \ldots, s_{j-1}$ has been constructed. Let $S_{j-1} = \{s_0, s_1, s_2, \ldots, s_{j-1}\}$ be the set of chosen hovering locations for the UAV so far and $s_0 = d$, which implies that the sum of energy consumptions on these $j$ hovering locations and traveling along the closed tour $TSP(S_{j-1})$ is no more than $E$, where $TSP(S_{j-1})$ is the length (energy) of the closed tour induced by the nodes in set $S_{j-1}$, which is obtained by applying Christofides’s algorithm for the Traveling Salesman Problem [8]. Recall that the coordinates of location $s_j$ are $(x_j, y_j, H)$, and $C(s_j) = \{v_i | v_i \in V \land \sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \leq R_0\}$ is the set of IoT devices covered by the UAV at hovering location $s_j$, i.e., each IoT device in $C(s_j)$ is able to transmit its data to the UAV located at $s_j$. The rest is to determine the next hovering location $s_j$ as follows.

The volume of data collected by the UAV when it is located at $s_j$ is

$$V'(s_j) = \sum \{D_v | v \in C(s_j) \cup \bigcup_{i=0}^{j-1} C(s_i')\}$$

6.2 Algorithm for the partial data collection maximization problem with hovering coverage overlapping

In the following, we consider the partial data collection maximization problem with hovering coverage overlapping, by adopting the similar technique for the full data collection maximization problem with hovering coverage overlapping. We show how to modify Algorithm 3 for this purpose. Specifically, we treat each virtual hovering location derived from each (real) potential hovering location as a potential hovering location of the UAV as we did for Algorithm 3. However, we only allow one virtual hovering location of each potential hovering location to be included in the closed tour, so the tour must be a simple closed tour. If two virtual hovering locations $s_{j,k_1}$ and $s_{j,k_2}$ derived from the same hovering location $s_j$ are chosen to be included in the closed tour with $1 \leq k_1 < k_2 \leq K$, then only location $s_{j,k_1}$ will be included while location $s_{j,k_2}$ will be removed from the tour. Notice that all data that are supposed to be collected by the UAV at hovering location $s_{j,k_1}$ for sojourn duration $t'(s_{j,k_1})$ will be collected by the UAV at location $s_{j,k_2}$ since the UAV at $s_{j,k_2}$ takes a longer sojourn duration $t'(s_{j,k_2})$ for data collection while $t'(s_{j,k_2}) > t'(s_{j,k_1})$. It is also noted that the volume of data in a sensor $v \in V$ may be collected by the UAV at multiple hovering locations if the stored data in $v$ has not been fully collected yet. We here use an example to illustrate this scenario. Assume that the duration of collecting all data
**Algorithm 3** A heuristic algorithm for the full data collection maximization problem with hovering coverage overlapping

**Input:** An aggregate sensor network $G = (V, E)$ with a set $V$ of aggregate sensor nodes, a UAV with energy capacity $\mathcal{E}$ at depot $d$, and each node $v \in V$ has data volume $D_v$, and a given constant $\delta > 0$.

**Output:** Find a closed tour including depot $d$ for the UAV such that the volume of data collected from all aggregate sensor nodes covered by it at its hovering locations in the tour is maximized, subject to the energy capacity of the UAV.

1: Partition the monitoring region into $M$ squares $s_0, s_1, \ldots, s_M$, let $S = \{s_0, s_1, \ldots, s_M, d\}$;
2: Construct the closed tour for the UAV iteratively, $S_0 = \{d\}$ initially;
3: $j \leftarrow 1$;
4: while $\sum_{s_j \in S_{j-1}} t'(s_j) \cdot \eta_h + \frac{TSP(S_{j-1}) \cdot \eta_s}{\nu} < \mathcal{E}$ do
5: Choose the next hovering location $s_j \in S \setminus S_{j-1}$ such that the ratio $\rho(s_j)$ is the maximum one, i.e., $s_j = \text{argmax}_{s_j' \in S \setminus S_{j-1}, \sum_{s_j' \in S_{j-1} \cup \{s_j'\}} t'(s_j') \cdot \eta_h + \frac{TSP(S_j) \cdot \eta_s}{\nu} \leq \mathcal{E}}\}$;
6: $S_j \leftarrow S_{j-1} \cup \{s_j\}$;
7: $j \leftarrow j + 1$;
8: end while;
9: return the closed tour with the hovering location sequence $s_0, s_1, \ldots, s_j$.

from a sensor $v$ is $t(v) = D_v/B$ time units. We further assume that $v$ is covered by the UAV at three different hovering locations $s_{j1}, s_{j2}$ and $s_{j3}$ with sojourn time $t_1, t_2$ and $t_3$, respectively, i.e., $v \in C(s_{j1}), v \in C(s_{j2})$, and $v \in C(s_{j3}).$ If $t(v) \geq t'(s_{j1}) + t'(s_{j2}) + t'(s_{j3}) = t_1 + t_2 + t_3$, then the rest of data stored at sensor $v$ can still be collected by the UAV at these three hovering locations. Therefore, the residual data volume and the hovering duration at some virtual hovering locations must be recalculated after a virtual hovering location $s_j$ is added to the closed tour, since the data of some aggregate sensor nodes in $C(s_j)$ are contained by these potential virtual hovering locations and their data have been partially collected in the previous hovering locations already, where node $s_{j,k}$ is defined as follows.

$$s_{j,k} = \text{argmax}_{s_j' \in S_j \setminus S_{j-1}, \sum_{s_j' \in S_{j-1} \cup \{s_j'\}} t'(s_j') \cdot \eta_h + \frac{TSP(S_j) \cdot \eta_s}{\nu} < \mathcal{E}}\}.$$ (27)

The detailed algorithm for the full data collection maximization problem with hovering coverage overlapping is given in Algorithm 4.

6.3 Analysis of the proposed algorithm

The rest is dedicated to the correctness proof and time complexity analysis of the proposed algorithms. In the following we first analyze the time complexity of the proposed algorithm, Algorithm 3.

**Theorem 4.** Given an aggregate sensor network $G(V, E)$ with each node $v \in V$ having a data volume $D_v$ for collection, and a UAV with energy capacity $\mathcal{E}$ and depot $d$, assuming that the moving unit of the UAV is measured by $\delta > 0$ and its hovering coverage range of the UAV at each hovering location is a circle with radius $R_0$, there is an efficient heuristic algorithm, Algorithm 3, for the full data collection maximization problem in $G$ with hovering coverage overlapping. The algorithm takes $O(\frac{\pi R_0^2}{\delta^2} \cdot |V|^4)$ time.

**Proof:** Algorithm 3 proceeds iteratively, and the number of iterations is bounded by $|S| = M$. Within iteration $j$ with $1 \leq j \leq M$, identifying the next hovering location $s_j$ is performed through the calculation of $\rho(s_j)$ for every location $s_j \in S \setminus S_{j-1}$. This takes $O(|S| \cdot |S_{j-1}| \cdot |V| + |S|^2) = O(M \cdot |V| + M^2)$ time, due to the fact that the calculation of $TSP(S_j)$ takes $O(|S_j|^3)$ time by Christofides’ algorithm [8]. The proposed algorithm thus takes $O(M^2 \cdot |V| + M^4) = O(\frac{\pi R_0^2}{\delta^2} \cdot |V|^4)$ time, since we proved that $M \leq (\frac{\pi R_0^2}{\delta^2} + 1) \cdot |V|$ in Section 7. Notice that in
practice, the values of both \( \delta \) and \( R_0 \) are constants, the time complexity of Algorithm 3 thus is \( O(|V|^4) \).

We then analyze the correctness of Algorithm 4. To this end, we show that the amount of data that is supposed to be collected by the UAV at \( s_{j,k_1} \) will be collected by the UAV at \( s_{j,k_2} \) if \( k_1 < k_2 \) by the following lemma.

**Lemma 2.** For a given hovering location \( s_j \), if one of its virtual hovering locations \( s_{j,k_i} \) is included in the closed tour and its another virtual hovering location \( s_{j,k_2} \) is chosen to be added to the tour later with \( k_2 > k_1 \), then \( s_{j,k_1} \) will be removed from the closed tour by the proposed algorithm, Algorithm 4. By doing so will not reduce the amount of data collected when the UAV at hovering location \( s_j \), and the correctness of the proposed algorithm holds.

**Proof:** We note that for each potential hovering location, at most one of its virtual hovering locations is included in the closed tour. Following Algorithm 4, if at most one virtual hovering location derived from a potential hovering location in \( S \) is included in the closed tour, the solution is feasible. Otherwise, assume that a virtual hovering location \( s_{j,k_1} \), of \( s_j \) at iteration \( i_1 \) of the algorithm has been added to the closed tour already, another virtual location \( s_{j,k_2} \), of \( s_j \) at iteration \( i_2 \) is added to the closed tour again, where \( 1 \leq i_1 < i_2 \leq |S| \). To ensure that at most one virtual hovering location for each hovering location is added to the closed tour, the hovering location \( s_{j,k_1} \) will be removed while \( s_{j,k_2} \) is added to the tour, and all the other virtual hovering locations \( s_{j,k'} \) of \( s_j \) with \( k' < k_2 \) will be removed from \( S \) for further consideration by the algorithm. Despite that the volume of data collected and sojourn durations at each hovering location from iterations \( i_1 + 1 \) to \( i_2 - 1 \) could be changed due to the removal of \( s_{j,k_1} \) from the closed tour, we will not update these hovering locations and their sojourn durations as the residual volume of data at each IoT device in \( C(s_{j,k_1}) \) (iteration \( i_1 \)) in fact is larger than when the virtual hovering location \( s_{j,k_1} \) is included in the tour. The volume of all that data was supposed to be collected by the UAV at location \( s_{j,k_1} \) with sojourn duration \( k_1 \cdot t(s_j)/K \) will be later collected by the UAV at location \( s_{j,k_2} \) (iteration \( i_2 \)) with sojourn duration \( k_2 \cdot t(s_j)/K \). Thus, the total volume of data collected from the closed tour by the removal of \( s_{j,k_1} \) does not change.

We then have the following theorem.

**Theorem 5.** Given an aggregate sensor network \( G(V,E) \) with each node \( v \in V \) having a data volume \( D_v \) for collection, and a UAV with energy capacity \( E \) and depot \( d \), assuming that the moving unit of the UAV is measured by \( \delta > 0 \) and its hovering coverage range of the UAV at each hovering location is a circle with radius \( R_0 \), there is a heuristic algorithm, Algorithm 4, for the partial data collection maximization problem with hovering coverage overlapping, assuming that the distance differences among the potential hovering locations within each square are negligible. The algorithm takes \( O(\frac{n \cdot R^3}{\delta^2} \cdot K^4 \cdot |V|^4) \) time, where \( K \) is a given integer with \( K \geq 1 \).

**Proof:** The correctness of the solution delivered by Algorithm 4 is shown by Lemma 2, omitted. The time complexity analysis of Algorithm 4 is similar to the one in the proof body of Theorem 4. The only difference between the two algorithms lies in the fact that we now have \( M' = K \cdot M \) virtual squares instead of \( M \) squares in Theorem 4, omitted.

## 7 Performance Evaluation

In this section, we evaluate the performance of the proposed algorithms for the full and partial data collection maximization problems through experimental simulations. We also investigate the impact of important parameters on the algorithm performance.

### 7.1 Experimental Settings

We consider a sparse sensor network that consists of 500 aggregate sensor nodes randomly deployed in a 1,000 meters \( \times \) 1,000 meters square. The data volume of each aggregate sensor node is randomly drawn from 100 MB to 1,000 MB. Without loss of generality, we assume that the transmission range \( R \) of each aggregate sensor node is 70 meters and the data transmission rate of the sensor is 150 Mbps [23]. We assume that a UAV is deployed at a depot \( d \) initially. The UAV has energy capacity \( E = 3 \times 10^6 \text{joules} \) at constant flying speed \( v = 10 \text{ m/s} \) and the flying altitude \( H = 50 \text{ meters} \). The energy consumption rates of the UAV on traveling and hovering are \( \eta_t = 100 \text{ J/s} \) and \( \eta_h = 150 \text{ J/s} \), respectively [19]. The value in each figure is the mean of the results out of 50 network instances of the same size. The running time of an algorithm is obtained based on a machine with 3.6 GHz Intel i7 single-core CPU and 16 GB RAM. Unless otherwise specified, these parameters will be adopted in the default setting. Table 1 lists the default settings of the parameters in this paper.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensing field</td>
<td>1,000 m \times 1,000 m</td>
</tr>
<tr>
<td>Network size</td>
<td>500 aggregate sensor nodes</td>
</tr>
<tr>
<td>Node data volume ( D_v )</td>
<td>100 MB - 1,000 MB</td>
</tr>
<tr>
<td>Transmission range ( R )</td>
<td>70 m</td>
</tr>
<tr>
<td>Flying altitude ( H )</td>
<td>50 m</td>
</tr>
<tr>
<td>Data rate ( B )</td>
<td>150 Mbps</td>
</tr>
<tr>
<td>Energy capacity ( E )</td>
<td>3 \times 10^6 \text{joules}</td>
</tr>
<tr>
<td>Flying speed ( nu )</td>
<td>1</td>
</tr>
<tr>
<td>Traveling consumption rate ( \eta_t )</td>
<td>100 J/s</td>
</tr>
<tr>
<td>Hovering energy rate ( \eta_h )</td>
<td>150 J/s</td>
</tr>
</tbody>
</table>

TABLE 1: Table of Parameter Settings
7.2 Performance evaluation of different algorithms for the full and partial data collection maximization problems without hovering coverage overlapping

We first investigate the performance of different algorithms for the full data collection maximization problems without hovering coverage overlapping. As shown in Fig. 3(a), Algorithm 1 outperforms the benchmark algorithm, and collects at least 80% of the data in the optimal solution. For example, when $E = 3 \times 10^5$ Joules, the volume of data collected by Algorithm 1 is around twice the amount by the benchmark algorithm, and the volume gap between them becomes larger with the increase of the energy capacity on the UAV. Fig. 3(b) plots the running time curves of the two mentioned algorithms and ILP solver, from which it can be seen that the running time of Algorithm 1 is much shorter than ILP solver.

We also study the performance of the proposed algorithm for the partial data collection maximization problem. Fig. 4(a) shows that with the growth of $K$, the UAV can collect more data per tour. For instance, when $E = 3 \times 10^5$ Joules, the volume of data collected by Algorithm 2 when $K = 4$ is 149.8 GB, while the volume of data collected by Algorithm 1 is 131.9 GB. However, the growth of $K$ results in a longer running time of Algorithm 2 and the ILP solver. E.g., when $K = 1$, the running time for Algorithm 2 and ILP solver is around 100 ms and 2,000 ms respectively, but when $K = 16$, the time consuming is around 24,830 ms and $4 \times 10^7$ ms respectively. When $K \geq 32$, the ILP solver cannot deliver a solution within a reasonable amount of time.

7.3 Performance evaluation of algorithms for the full and partial data collection maximization problems with hovering coverage overlapping

We then study the performance of Algorithm 3 and Algorithm 4 for the full data collection maximization problem with hovering coverage overlapping. It can be seen from Fig. 5(a) that Algorithm 3 and Algorithm 4 outperform the benchmark algorithm significantly. Furthermore,
Algorithm 4 is superior to Algorithm 3 as the latter is a special case of the former when $K = 1$. When $\delta = 5$ meters, the collected data volume by Algorithm 3 and Algorithm 4 ($K = 2$) are 132.8 GB and 147.7 GB, respectively, which are 79.1% and 99% higher than that of the solution delivered by the benchmark algorithm (74.14 GB).

We also study the performance of Algorithm 4 by varying the value of $K$. It can be seen from Fig. 5(a) that a larger $K$ will result in more data collected per tour, this is due to more accurate planning for data collection per unit energy consumption. For instance, the volume of data collected increases from 147.7 GB to 150.7 GB when the value of $K$ increases from 2 to 4. Fig. 5(b) indicates that Algorithm 4 takes a longer running time than that of Algorithm 3 with the growth of $K$. For example, the running time of Algorithm 4 is around 54.1 minutes when $K = 4$, which is around 50 times of the running time 1.61 minutes of Algorithm 3.

7.4 Impact of important parameters on the performance of the proposed algorithms for the full and partial data collection maximization problems with and without hovering coverage overlapping

The rest is to evaluate the impact of parameters $\delta, E, |V|$ on the performance of the proposed algorithms for the full and partial data collection maximization problems without hovering coverage overlapping. We start with the impact of the parameters on the performance of the proposed algorithms for the full data collection maximization problem without hovering coverage overlapping.

(a) We analyze the impact of energy capacity $E$ of the UAV, by varying it from $3 \times 10^5$ to $9 \times 10^5$ Joules while fixing other parameters. Fig. 3(a) shows that the volume of data collected by all algorithms increases with the growth of energy capacity. When $E = 9 \times 10^5$, Algorithm 1 collects 73.09% ($\approx (228.3 - 131.9)/131.9$) more data than the one when $E = 3 \times 10^5$. A larger battery capacity implies that the UAV can visit more hovering locations and hovering
Fig. 7. Performance of different algorithms, by varying the battery capacity of the UAV from $3 \times 10^5$ joules to $9 \times 10^5$ joules.

Fig. 8. Performance of different algorithms by varying the number of sensors from 500 to 1,000.

longer at hovering locations, hence increasing the running time of the algorithm. On the other hand, from Fig. 3(b), it can be seen that the running time of Algorithm 1 and Algorithm 2 increases with the growth of the UAV energy capacity, because more hovering locations needs planning. While the running time of the benchmark algorithm decreases with the growth of the UAV energy capacity, because less nodes will be pruned from the initial TSP tour as more energy can be consumed in the tour.

(b) We study the impact of the number of aggregate sensor nodes $|V|$ on the performance of Algorithm 1 and Algorithm 2. Fig. 6(a) depicts the data collections of the UAV while varying the number of aggregate sensor nodes in the monitoring region from 500 to 1,000, which shows the amount of collected data goes up when there are more aggregate sensor nodes. When there are 1,000 nodes, Algorithm 1 collects 196.3 GB, which is approximately 64.4GB higher than that when there are 500 nodes. Fig. 6(b) plots the running time curve of different algorithms. It can be seen that Algorithm 1 and Algorithm 2 remain stable when the number of nodes increases from 500 to 1000. The reason is that the hovering sojourn duration at each hovering location is determined by the sensor within the coverage range of the UAV with the maximum data volume.

We then evaluate the impact of parameters $\delta$, $E$ and $|V|$ on the performance of Algorithm 3 and Algorithm 4 for the full and partial data collection maximization problems with hovering coverage overlapping.

(a) We evaluate the impact of $\delta$ on the performance of the two mentioned algorithms. It can be seen from Fig. 5(a) that for a fixed $K \geq 1$ of number of partitionings on the sojourn duration at each hovering location, the total volume of data collected per tour reduces, so do the running times of the algorithms because the number of potential hovering locations for the UAV becomes smaller, and less data will be collected. E.g., when $K = 4$, it can be seen from Fig. 5(a) that when $\delta = 5$ meters, the collected data volume is about 13.9% higher than that by itself when $\delta = 30$ meters. Therefore, when $\delta$ is sufficiently small, the total volume of data collected by the UAV is maximized. Fig. 5(b) depicts the running times of the two mentioned algorithms.

(b) We investigate the impact of the energy capacity $E$...
of the UAV on the performance of different algorithms, by varying it from $3 \times 10^5$ joules to $9 \times 10^5$ joules while fixing the value of $\delta$ at 10 meters. Fig. 7(a) illustrates that the collected data volume goes up with the increase on the energy capacity of the UAV, since the UAV can visit more hovering locations with longer hovering durations to collect more data from its hovering locations. For example, when $K = 4$, the collected data is increased by 82% with the growth of the energy capacity of the UAV from $3 \times 10^5$ joules to $9 \times 10^5$ joules. Fig. 7(b) demonstrates the impact of the battery capacity of UAV on the running time of the algorithm. A larger battery capacity implies that Algorithm 3 and Algorithm 4 can visit more hovering locations, hence increasing their running time. On the other hand, with more energy capacity, the benchmark algorithm will remove fewer nodes from the initial closed tour $C$, which leads to a shorter running time. Algorithm 4 thus needs more running time compared with that of the benchmark algorithm with the growth of the energy capacity on the UAV.

(c) We study the impact of aggregate sensor nodes $|V|$ on the performance of Algorithm 3 and Algorithm 4, by varying the number of aggregate sensor nodes from 500 to 1,000. Fig. 8(a) shows that volume of data collected grows with the increase on the number of aggregate sensor nodes. On the other hand, Fig. 8(b) depicts that the running time of each comparison algorithm does not change too much, by increasing the number of sensors. The rationale behind is that the hovering sojourn duration at each hovering location is determined by the sensor with the maximum data volume, due to the OFDMA technique.

8 Conclusions
In this paper, we studied the full and partial data collection maximization problems for IoT sensor networks, using an energy-constrained UAV. We first proposed a novel data collection framework that enables the UAV to collect sensor data from multiple IoT devices simultaneously. We then formulated the full and partial data collection maximization problems that allow the UAV to fully or partially collect data from IoT devices at each hovering location. We thirdly showed that both the problems are NP-hard, and instead devised efficient approximation and heuristic algorithms for them. We finally evaluated the performance of the proposed algorithms through experimental simulations. Simulation results demonstrate that the proposed algorithms are promising.

Acknowledgement
We appreciate the associate editor and the three anonymous referees for their constructive comments and invaluable suggestions, which help us improve the quality and presentation of the paper greatly. The work by Yuchen Li and Weifa Liang was supported by Australian Research Council under its Discovery Project Scheme with Grant No. DP200101985, and part of the work by Weifa Liang was conducted at Australian National University. The work by Wenzheng Xu was supported by the National Natural Science Foundation of China (NSFC) with grant number 61620330, Sichuan Science and Technology Program (Grant No. 2018GZDX0010 and 2017GZDXZ0003), and the National Key Research and Development Program of China (Grant No. 2017YFB0202403). The work by Xiaohua Jia was supported by the Research Grants Council of Hong Kong with Project No. CityU 11214316. The work by Yinlong Xu was supported by the NSFC with Grant No. 61772486, and the work by Haibin Kan was supported by the National Natural Science Foundation of China with Grants No. 61672166 and No. U19A2066, and the National Key Research & Development Plan with Grant No. 2019YFB2101703.

References


Zichuan Xu (M’17) received his PhD degree from the Australian National University in 2016, ME degree and BSc degree from Dalian University of Technology in China in 2011 and 2008, all in Computer Science. From 2016 to 2017, he was a Research Associate at Department of Electronic and Electrical Engineering, University College London, UK. He is currently an Associate Professor in School of Software at Dalian University of Technology. He is also a ‘Xinghai Scholar’ in Dalian University of Technology. His research interests include cloud computing, network function virtualization, software-defined networking, wireless sensor networks, routing protocol design for wireless networks, algorithmic game theory, and optimization problems.

Xiaohua Jia (A’10–SM’01–F’13) received the BSc and MEng degrees in 1984 and 1987, respectively, from the University of Science and Technology of China, and DSc in 1991 in information science from the University of Tokyo. He is currently a chair professor with the Department of Computer Science at City University of Hong Kong. He is currently a full professor with the School of Computer Science and Technology at USTC. Prior to that, he served the Department of Computer Science and Technology at USTC as an assistant professor, a lecturer, and an associate professor. His research interests include wireless networks, combinatorial optimization, design and analysis of parallel algorithms, etc. He received the Excellent Ph.D. Advisor Award of Chinese Academy of Sciences in 2006.

Yinlong Xu received his B.S. degree in Mathematics from Peking University in 1983, and M.S. and Ph.D. degrees in Computer Science from University of Science and Technology of China (USTC) in 1989 and 2004, respectively. He is currently a full professor with the School of Computer Science and Technology at USTC. Prior to that, he served the Department of Computer Science and Technology at USTC as an assistant professor, a lecturer, and an associate professor. His research interests include network coding, wireless networks, combinatorial optimization, design and analysis of parallel algorithms, etc. He received the Excellent Ph.D. Advisor Award of Chinese Academy of Sciences in 2006.

Wefia Liang (M’99–SM’01) received the PhD degree from the Australian National University in 1998, the ME degree from the University of Science and Technology of China in 1989, and the BSc degree from Wuhan University, China in 1984, all in Computer Science. He is currently a Professor at the Department of Computer Science, City University of Hong Kong, Hong Kong. He was a Professor at the Australian National University. His research interests include design and analysis of energy efficient routing protocols for wireless ad hoc and sensor networks, the Internet of Things, Mobile Edge Computing (MEC), Network Function Virtualization (NFV), Software-Defined Networking (SDN), design and analysis of parallel and distributed algorithms, approximation algorithms, combinatorial optimization, and graph theory. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, and he is a senior member of the IEEE.

Haibin Kan (M’13) received the Ph.D. degree from Fudan University, Shanghai, China, 1999. After receiving the Ph.D. degree, he became a faculty of Fudan University. From June 2002 to February 2006, he was with the Japan Advanced Institute of Science and Technology as an assistant professor. He went back Fudan University in February 2006, where he is currently a full professor. His research topics include coding theory, cryptography, and computation complexity.