



Research Article

Optimization Algorithm for AoI-Based UAV-Assisted Data Collection

Xiaoya Zhou ^{1,2} and Qi Zhu ^{1,2}

¹Nanjing University of Posts and Telecommunications, Jiangsu Key Laboratory of Wireless Communications, Nanjing 210000, China

²Nanjing University of Posts and Telecommunications, Engineering Research Center of Health Service System Based on Ubiquitous Wireless Networks, Ministry of Education, Nanjing 210000, China

Correspondence should be addressed to Qi Zhu; zhuqi@njupt.edu.cn

Received 31 January 2023; Revised 3 January 2024; Accepted 3 February 2024; Published 19 February 2024

Academic Editor: Diego Alexander Tibaduiza

Copyright © 2024 Xiaoya Zhou and Qi Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Regarding the issue of information freshness in systems that aid in data collection using unmanned aerial vehicles (UAVs), a data collection algorithm that is based on freshness and UAV assistance is proposed. Under the limitations of wireless sensor node communication distance and UAV parameters, the optimization problem of minimizing the average spatial correlation age of information (SCAoI) of all nodes in the area is set up. This problem is solved by optimizing the number of clusters, UAV flight trajectories, and the order of data collection from cluster member nodes. The maximum communication distance of the nodes is used as the cluster formation radius, and the maximum-minimum distance clustering algorithm is used to cluster the nodes in the region to obtain the minimum number of clusters. After it has been proven that the trajectory optimization problem in this study is NP-hard, the ant colony algorithm is applied to obtain the minimum flight time and the corresponding trajectory. By using the greedy algorithm to determine the member nodes in the sequence of data collection for a cluster, the instantaneous SCAoI of the UAV arriving at the cluster head is solved. Simulation results show that the proposed algorithm in this paper can effectively improve the freshness of data and reduce the average SCAoI of the system compared with the algorithm in the comparative literature, reducing the average SCAoI by about 61%.

1. Introduction

The continuous development of wireless communication technologies and the Internet of Things has given rise to many systems oriented towards real-time applications, such as smart homes, smart transportation, and smart health [1]. For such systems, it is essential that the information be current. The terminal device (e.g., sensors, surveillance cameras, smart wearables) needs to sense the data of the surrounding environment in real-time and monitor the system status to make accurate and reliable decisions and controls. If the terminal receives old data, it may affect system decisions and cause significant security risks. For information update applications, the significance of information freshness is rising quickly. To accurately describe the freshness of the information, academics have proposed the age of information

(AoI), and many scholars use it as a measure of information freshness. In application scenarios where information freshness is sensitive, the networking method of the system and the allocation of wireless communication resources can have a great impact on the information age of the data. This has become an open challenge with research implications. In this paper, the impact on the information age of the system is studied in the direction of the system networking method and the allocation of wireless communication resources.

During the update process, the AoI was utilized to record the freshness of state information in the queueing model [2], as the time between the generation of data and its reception at the receiving end. In [3], the age of information (AoI) is used as a measure of the freshness of information in the information update system. The impact of scheduling policies on the performance of AoI in single-server queues is investigated.

And the scheduling policy that can effectively improve AoI has been designed. In [4], they investigate a multisource preemptive queuing model as well as how to regulate the generation rate of each source under various preemption strategies to attain the highest level of information freshness overall. In [5], the AoI of a multisource queuing model with an FCFS (first-come, first-served) Poisson arrival strategy is examined. They establish precise formulas for the average AoI of the multisource M/M/1 and multisource M/G/1 queueing models. It is further demonstrated that, in time-sensitive control applications, reducing the average delay alone does not reduce the AoI. There is also a large literature on the use of AoI as a performance metric in caching networks. In [6, 7], the relationship between service delay and content freshness (as defined by AoI) in mobile edge caching networks is examined. The caches are placed near the users, which can successfully lessen the service latency of content delivery while reducing the latency and additional transmission resources required to update the cached content. To achieve a compromise between AoI and latency, a freshness-aware cache update strategy is developed. A cache refresh system is considered in [8]. Both the age of synchronization (AoS) and the age of information (AoI) are used to measure how recent the local cache is. The closed-form expressions for minimizing AoS and AoI are derived for larger and smaller refresh rates, respectively. In [9], a cache update algorithm based on content popularity and information freshness is proposed. The algorithm fully considers the mobility of users and the dynamics of popular content in time and space and introduces the age of information (AoI) to achieve dynamic updating of content. In addition, with the wide application of game theory and reinforcement learning in various fields, they have also been applied to solving AoI-related problems. A content resale problem is discussed in [10] which provides a hybrid multicast/unicast/D2D transmission architecture oriented towards age of information and cache assistance to increase the data transmission rate, reduce the burden of large data traffic, and improve system efficiency, where the problem is decomposed into two subproblems and the subproblems are solved through the Stackelberg game and auction framework, respectively. Reinforcement learning is used in the literature [11] to investigate the best update strategy for the age of information (AoI) and the urgency of information (UoI) of real-time status information based on resource constraints. Urgency of information (UoI) further includes context-aware weights indicating whether the monitored process is in an emergency. The simulation results show that the threshold of the optimal policy increases as the resource constraint is tightened.

UAVs, as a stable, mobile, and flexible flight device with low cost, can be used for image recognition or vision processing [12–14] and as an auxiliary device in wireless communication networks. In [15], a MEC network consisting of IoT devices, UAV base stations, edge clouds, and data centers with energy-efficient UAV support services is proposed, and a GreenUAV-CoCaCo algorithm is proposed to jointly optimize the communication, caching, and computation energy consumption of UAVs. In [16], cache-enabled UAVs are used to provide contextual messaging services to end devices. Unlike traditional network traffic, contextual

information changes over time, thus increasing the demand for AoI constraints. Cache replacement and content distribution strategies are designed to minimize the traffic on the ground network according to the requests of users and the dynamic changes of the content. In [17], maximizing the quality of service (QoS) based on the freshness of the data was studied while considering the range of the UAV. Modeling is used to convert the optimization problem into a semi-Markov decision process (SMDP), and a hierarchical deep Q network- (DQN-) based path planning algorithm is then suggested to learn the best course of action. In [18], an AI-based end-to-end framework is proposed to resolve the issue with UAV flight trajectory planning. After simulation, it is proved that the AI-based framework is like a commercial open-source solver in terms of accuracy but can be twice as efficient for scenarios with a large number of nodes. In [19], the concept of AoI is enlarged by including a brand-new metric called correlation-aware AoI (CAAoI) to assess the timeliness and degree of correlation of the data collected by UAVs from the ground. In [20], the topic of UAV-assisted data collection in wireless sensor networks is investigated, where the AoI of each wireless sensor node is used to gauge the freshness of information.

In the data collection scenarios of [18, 20], all use AoI as a measure of data freshness but ignore the data correlation due to data coming from the same collection device or multiple devices. In [19], it states that devices collecting the same type of information that are in proximity are highly correlated at the same moment, which can affect the diversity of the data. Therefore, in this paper, SCAoI is a performance index that is used to measure how recent the information is, which can balance the diversity and freshness of the data. The [19] investigates the AoI when the UAV performs data collection in a region where there are few wireless sensor nodes, but the number of nodes in real-world application scenarios will be much larger than the setup in the literature. The UAV in [19] travels above each node for data collection. Therefore, as the nodes multiply, the flight time also grows, which is not conducive to guaranteeing information freshness. In this paper, a SCAoI-based UAV-assisted data collection method is proposed. In this scenario, every node is first clustered, and the cluster head node gathers the information acquired by the nodes of the cluster. When the UAV flies over the cluster head node, the cluster head node transfers the data that it has temporarily stored to the UAV.

The following is a summary of the significant contributions in this paper.

- (1) The spatial correlation age of information is used to construct a model for UAV data collection that measures the freshness of information. All nodes are organized into clusters using the maximum-minimum distance clustering algorithm, and then, the member nodes use the time that the UAV spends moving between hovering positions to gather data in accordance with TDMA and upload it to the cluster head node. The UAV then flies over the cluster head node along the best route to receive the data that has been uploaded by the cluster head node

- (2) The closed expressions for the instantaneous SCAoI of data collected at the cluster head and the average SCAoI of all nodes are derived. Under the constraints of maximum communication distance between nodes and UAV flight parameters, the optimization problem of minimizing the average SCAoI is developed. The initial problem is divided into three optimization sub-problems, which must each be solved separately: cluster formation, trajectory, and data collection order of cluster members
- (3) First, the maximum communication distance of nodes is taken as the radius of cluster formation, and to produce the smallest possible number of clusters, all nodes are clustered using the maximum-minimum distance clustering algorithm. Then, it is shown that the trajectory optimization problem is NP-hard, and an ant colony algorithm is used to optimize the UAV's flight path to achieve the shortest possible flight time. Finally, the greedy algorithm is then employed to determine the optimal order for data collection at cluster nodes
- (4) Simulation results show that the proposed method can successfully reduce the average SCAoI. Comparing this method to how the UAV collects data at each node, the average SCAoI can be reduced by almost 61%, and the freshness of the collected data is effectively guaranteed

The remainder of the paper is structured as follows. Section 2 gives the data collection model with UAV assistance and the establishment of the optimization problem. Section 3 details the problem solution, the algorithm framework, and the detailed design process of the algorithm. Section 4 gives the simulation results. Finally, the entire study is concluded in Section 5. In addition, to improve the readability of this paper, the abbreviations and their meanings covered in this paper are summarized in Table 1.

2. System Model and Problem Building

2.1. System Model. Figure 1 depicts the system model used in this paper, with a UAV U , a data center DC, and N wireless sensor nodes, which are distributed at random in a rectangular area with a side length A . The sensor node $n_i \in \mathcal{N} = \{n_1, n_2, \dots, n_N\}$, with coordinates $L_i = (x_i, y_i)$, is used to collect information from the surrounding environment. The UAV serves as a data collection tool from the DC, collecting the data captured by the nodes in the region and returning it to the DC for processing. To facilitate the deployment of UAVs, all the nodes are divided into M clusters with cluster head nodes $ch_j \in \mathcal{CH} = \{ch_1, ch_2, \dots, ch_M\}$ and $\mathcal{CH} \subseteq \mathcal{N}$, and the binary variable $\gamma_{ij} = 1$ indicates that node n_i belongs to the cluster with ch_j as the cluster head, otherwise $\gamma_{ij} = 0$. The nodes in the same cluster as ch_j are denoted by the set $C_j = \{n_i | \gamma_{ij} = 1, n_i \in \mathcal{N}\}$, $1 \leq j \leq M$, and the number of nodes within C_j is denoted by N_j . The cluster head node serves two primary purposes. First, it gathers data from the member nodes, and

second, it establishes the hovering position of UAV. The UAV is transferred between hovering positions to upload the collected information to the cluster head node in a certain order according to TDMA. The UAV is circling close to where the cluster head node is located. While the UAV is hovering over the cluster head node, the cluster head node will transmit to it all the information gathered by the member nodes. Once the UAV has collected data from one cluster head node, it will move on to the next after receiving the data. The UAV moves according to its trajectory $V = [v_0, v_1, \dots, v_M, v_{M+1}]$. Since the UAV starts from DC and eventually returns to DC, so $v_0 = v_{M+1} = DC$. If the velocity v and height h remain unchanged, then the hovering position of the UAV can be expressed by the coordinates of the cluster head node. When the UAV moves from v_{l-1} to v_l , the flight time can be expressed as

$$f_{l-1,l} = \frac{\sqrt{(x_{l-1} - x_l)^2 + (y_{l-1} - y_l)^2}}{v}, l = 1, 2, \dots, M + 1, \quad (1)$$

where (x_{l-1}, y_{l-1}) and (x_l, y_l) are the coordinates of the cluster head node corresponding to v_{l-1} and v_l in the trajectory, respectively. Later in the paper, $f_{l-1,l}$ is denoted as f_l for brevity.

The following Figure 1 is an example of the process of collecting node information from DC by a UAV as an assistant device. Define the UAV trajectory as V , as shown in Figure 1, with $M = 6$. The UAV starts from DC and flies first to v_1 , the area where the cluster head node ch_5 is located, according to trajectory V . The nodes in C_5 use the time f_1 when the UAV flies from DC to v_1 for the order $G = [g_1, g_2, \dots, g_{N_5}]$ to upload the collected data to ch_5 by TDMA. After arriving at ch_5 , to receive the data sent by the cluster head node, the UAV hovers for a while and then flies through v_2, v_3, v_4, v_5 , and v_6 one by one and finally returns to DC, i.e., the final flight path of the UAV is $V = [DC, ch_5, ch_1, ch_4, ch_6, ch_2, ch_3, DC]$.

The UAV is used as an assistant device to collect information, and if it flies over each node to collect the information that node collects, it will take a long flight time because it needs to traverse each node, and the freshness of the information will be reduced as a result. Reducing the time taken for data to travel from the source to the destination is vital to ensuring the accuracy of the information. Three factors play a role in determining the time in this paper: the number of clusters M , the UAV flight trajectory V , and the order G in which the member nodes gather data. The number of clusters corresponds to the number of hovering locations along the flight path of the UAV. The amount of time spent in flight during the data-gathering procedure decreases as the cluster number decreases, which is better for preserving the accuracy of the data. The sensor nodes transmit the collected data in the form of time-stamped packets. Assuming that the maximum communication radius of a node is R , then the maximum cluster formation radius is R , which is the distance $d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \leq R$ between the cluster member node ch_j and the cluster head node n_i . The binary variable $\gamma_{ij} = 1$ is used to indicate that node n_i belongs to the cluster with ch_j as the cluster head, otherwise

TABLE 1: Table of abbreviations and full names.

Abbreviations	Full name	Abbreviations	Full name
UAV	Unmanned aerial vehicle	SCAoI	Special correlation age of information
AoI	Age of information	FCFS	First-come, first-served
AoS	Age of synchronization	UoI	Urgency of information
MEC	Mobile edge computing	TSP	Traveling salesman problem
QoS	Quality of service	SMDP	Semi-Markov decision process
DQN	Deep Q network	CAoI	Correlation-aware AoI
DC	Data center	TDMA	Time division multiple access
IoT	Internet of Things	NP-hard	Nondeterministic polynomial

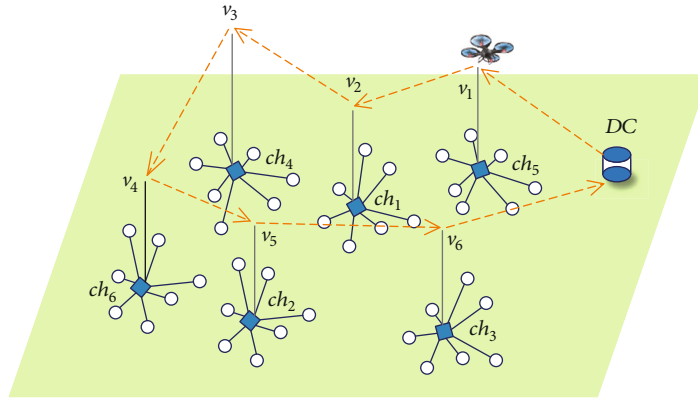


FIGURE 1: Model illustration of data collection with UAV assistance.

$\gamma_{ij} = 0$. For any node n_i , it can be classified into a certain cluster, and the clustering of all nodes can be represented as a vector $\mathbf{Y} = [\gamma_{ij}]$. When all nodes have completed clustering, the UAV must hover directly above the cluster head node to gather data. Therefore, it is important to optimize the path of UAV travel. A sensible flight path enables it to finish the data-gathering operation while spending the least amount of time feasible to make additional contributions to enhancing the freshness of information. In addition, the AoI proposed in this paper is aimed at gauging how recent a piece of information differs from the traditional definition by considering the effect of spatial correlation between data collection devices due to their proximity. This effect is more evident among nodes within the same cluster. Nodes within the same cluster are close to each other, which can make the similar data collected at the same moment highly correlated [21], and the diversity of the data is diminished, which is not conducive to data analysis. Consequently, we construct a data-gathering order for member nodes that can take into consideration the freshness of information and the correlation between collection devices.

2.2. Cluster Member Node and Cluster Head Node Communication Model. The communication between member nodes and cluster head nodes uses a point-to-point communication model on the ground with a transmission rate of

$$R_{\text{ch}} = B \log_2 \left(1 + \frac{g p_i}{P_n} \right), \quad (2)$$

where B is the bandwidth of system, p_i is the transmit power of node, P_n is the noise power, and $g = \beta_0 d^{-\alpha}$ is the channel gain, where β_0 is the reference channel gain of distance, d is the separation between the cluster head node and the member node, and α is a constant coefficient related to the environment. A member node must spend $t_{\text{ch}} = L/R_{\text{ch}}$ seconds before sending a packet with data quantity L to the cluster head node.

Most of the time, the nodes of the system are in a state of sleep, and when the UAV follows the established trajectory V , flying from v_l , $l \neq M$ to v_{l+1} , the nodes in the cluster corresponding to v_{l+1} are awakened and start collecting data in a certain order and sending it using TDMA to the cluster head node. In this process, ignoring the specific time required by the nodes to collect information, the time slot length of TDMA is the length of time used to send information to the cluster head node from the member nodes. The data gathered from each cluster member node must be sent to the cluster head nodes, so when the member nodes have not yet all finished uploading and the UAV has arrived above the cluster head, the UAV is required to extend the hovering time and wait for all nodes to finish the uploading task before the process of uploading data from the cluster head to the UAV.

2.3. Communication Model between Cluster Head Node and UAV. Data is uploaded to the UAV by the cluster head node using a probability-based approach for air-to-ground communication [22]. Considering that additive Gaussian white

noise is present, the data transmission rate of the cluster head node to the UAV flying above it can be expressed as

$$R_u = B \log_2 \left(1 + \frac{p_i}{\bar{\varphi}_i p_n} \right), \quad (3)$$

where $\bar{\varphi}_i$ is the typical route loss for both line-of-sight and non-line-of-sight transmission, expressed as

$$\bar{\varphi}_i = P_i(\text{LoS})\varphi_i(\text{LoS}) + [1 - P_i(\text{LoS})]\varphi_i(\text{NLoS}). \quad (4)$$

$P_i(\text{LoS})$ is the line-of-sight transmission likelihood, and $\varphi_i(\text{LoS})$ and $\varphi_i(\text{NLoS})$ are the non-line-of-sight transmission and the path loss of line-of-sight, respectively.

$$P_i(\text{LoS}) = \frac{1}{1 + \beta_1 \exp[-\beta_2(\theta_i - \beta_1)]}, \quad (5)$$

where β_1 and β_2 are constant parameters related to environmental factors, and θ_i is the angle formed between the node uploading data and the UAV, considering that the UAV flies exactly above the node to collect data, $\theta_i = \pi/2$. The path loss of line-of-sight and non-line-of-sight transmission is denoted as $\varphi_i(\text{LoS}) = PL_i + \zeta_{\text{LoS}}$ and $\varphi_i(\text{NLoS}) = PL_i + \zeta_{\text{NLoS}}$, respectively. Where PL_i is the free path loss, $PL_i = 20 \log_{10}(4\pi h f_c / c)$, h is the flight altitude of the UAV, f_c is the carrier frequency, c is the speed of light, and ζ_{LoS} and ζ_{NLoS} are the additional path loss, which takes a constant value. The cluster head node needs $t_u = L/R_u$ to send a packet with data quantity L to the UAV.

2.4. AoI Model. AoI is used to gauge how recent a piece of information is, which refers of the interval between the moment at which data is produced at the source and when it is received at the receiver. The AoI at ch_j at moment t can be defined as

$$\Delta_j(t) = t - u_k(t), \quad (6)$$

where $u_k(t)$ is the time at which ch_j received the most recent generation of data at point t , i.e., the timestamp.

If the packet carrying the timestamp u_k reaches the cluster head node at time τ_k , $\tau_k = u_k + l_k$ can be used to indicate the time of arrival at the cluster head node, and l_k is the time delay of communications with the cluster head node. When the cluster head node is receiving new data and its informational freshness has increased overall, i.e., its AoI decreases, and the process is shown in Figure 2. As a result, $r_1 = \tau_1 - l_1 = u_1$ can be used to indicate the drop in AoI that occurs when the first member node uploads data to the cluster head node, and the decrease in the second arrival can be expressed as $r_2 = \tau_2 - l_2 = u_2 - u_1$, and similarly $r_3 = \tau_3 - l_3 - r_1 - r_2 = u_3 - u_2$, as shown in Figure 2; the AoI at the moment t can be stated as

$$\Delta_j(\tau) = \tau - r_3 - r_2 - r_1 = \tau - \sum_{k=1}^3 (u_k - u_{k-1}). \quad (7)$$

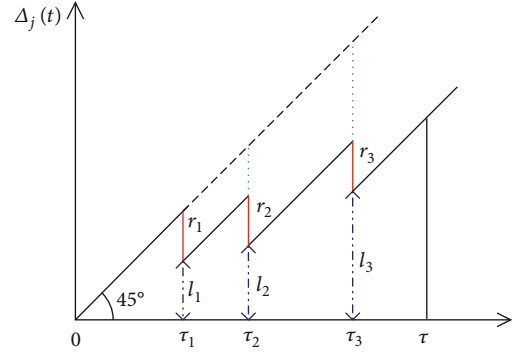


FIGURE 2: Illustration of the AoI of the cluster head node.

Equation (6) can be rewritten as

$$\Delta_j(t) = t - \sum_{k=1}^{U(t)} (u_k - u_{k-1}), \quad (8)$$

where $U(t)$ is the full count of data received by ch_j at time t and u_k is the timestamp of the k th upload data generation.

The gathered data shows a substantial association between the member nodes that are spatially adjacent to one another, so the effect of spatial correlation on the instantaneous AoI of cluster head nodes is considered. The instantaneous SCAoI is defined to characterize the instantaneous freshness of ch_j at time t [19], and the instantaneous SCAoI of ch_j at time t can be defined as

$$\Delta_j C(t) = t - \sum_{k=1}^{U(t)} (1 - \beta_k)(u_k - u_{k-1}), \quad (9)$$

$$\beta_k = \frac{1}{(d_k/\rho_s) + 1}, \quad (10)$$

where β_k is the correlation coefficient and $\beta_1 = 0$, the minimum distance d_k between the preceding $k-1$ member nodes and the k th member node that uploads data to the cluster head, and the constant ρ_s represents the strength of spatial correlation.

2.5. Problem Formation. The UAV starts at the DC and collects the information collected by all nodes in the area. Assuming that the UAV trajectory is V , with $v_j = \text{ch}_j$, $j \neq 0$, $M+1$, and $j \leq M$, taking C_j as an example, i.e., ch_j is the cluster head of this cluster, the detailed analysis of the cluster data generation process, after transmission, and finally arriving at DC. The AoI of the process can be considered the result of summing three parts of time. First, the member nodes in C_j use the flight time f_j of the UAV transfer from ch_{j-1} to ch_j to collect information in order G . All member nodes collect the data and need to transmit it to ch_j . Assuming that the UAV arrives at ch_j at time t_a^j , the instantaneous SCAoI of the cluster head node at that time, or $\Delta_j C(t_a^j)$, can be used to

represent the current freshness of cluster head nodes. Second, the temporary storage data of the cluster head node must be uploaded to the UAV. For the UAV to get the data, it must hover for a specific time, which is denoted as H and is the cumulative sum of the hovering times of the j th to M th clusters in the trajectory. Denoting the hovering time at the j th cluster head node by h_j , we have $H = \sum_{l=j}^M h_l$. The last part is the flight time, denoted by f_{j+1} from the j th flight to the $j+1$ st cluster, and as with the hover time, $F = \sum_{n=j+1}^{M+1} f_n$. For each cluster of data, they undergo the same process as in Figure 3. The UAV offloads the data collected from c to the DC at time t ; at which point, the instantaneous SCAoI of C_j is expressed as

$$\Delta_j C(t) = \Delta_j C(t_a^j) + H + F = \Delta_j C(t_a^j) + \sum_{l=j}^M h_l + \sum_{n=j+1}^{M+1} f_n. \quad (11)$$

Then, the AoI of all nodes in the system is shown in Figure 4, and the average SCAoI of all nodes can be expressed as

$$\begin{aligned} \overline{\Delta C(t)} &= \frac{1}{M} \sum_{j=1}^M \Delta_j C(t) = \frac{1}{M} (\Delta_1 C(t) + \Delta_2 C(t) + \dots + \Delta_M C(t)) \\ &= \frac{1}{M} \left(\Delta_1 C(t_a^1) + \sum_{l=1}^M h_l + \sum_{n=2}^{M+1} f_n + \Delta_2 C(t_a^2) + \sum_{l=2}^M h_l \right. \\ &\quad \left. + \sum_{n=3}^{M+1} f_n + \dots + \Delta_M C(t_a^M) + h_M + f_{M+1} \right) \\ &= \frac{1}{M} (\Delta_1 C(t_a^1) + h_1 + h_2 + \dots + h_M + f_2 + f_3 + \dots + f_{M+1} \\ &\quad + \Delta_2 C(t_a^2) + h_2 + h_3 + \dots + h_M + f_3 + f_4 + \dots + f_{M+1} \vdots \\ &\quad + \Delta_M C(t_a^M) + h_M + f_{M+1}) = \frac{1}{M} \sum_{j=1}^M \Delta_j C(t_a^j) \\ &\quad + \sum_{j=1}^M \frac{j}{M} h_j + \sum_{j=1}^M \frac{j}{M} f_{j+1}. \end{aligned} \quad (12)$$

By maximizing the number of clusters formed M , the flight trajectory V of the UAV, and the information collection order G of the nodes in the cluster, this research is aimed at reducing the average SCAoI of all the nodes in the system. The optimization issue is best described as

$$\begin{aligned} P : \min_{M, V, G} \overline{\Delta C(t)} \\ \text{s.t. (C1)} \quad d_{n, \text{ch}_j} \leq R \\ \quad \quad \quad \text{(C2)} \quad h_j = t_u N_j \\ \quad \quad \quad \text{(C3)} \quad f_{j+1} = \frac{d_{v_{j+1}}}{v}. \end{aligned} \quad (13)$$

The distance of a member node from the cluster head node cannot be greater than the maximum communication distance of nodes, according to constraint (C1). The number of nodes in the cluster and the rate of data transmission from the cluster head node to the UAV are both factors in constrain (C2) at equation (13) that affect the hovering duration of the UAV. Equation (C3) demonstrates the relationship between the flight time of the UAV and the separation between its two hovering places.

3. Problem Solving and Algorithm Design

3.1. Problem-Solving Framework. It is clear from equation (12) that the average SCAoI is the weighted sum of the instantaneous SCAoI, hover time, and flight time when the UAV reaches the head of each cluster, and the weighting factor is related to the number of clusters and trajectories. Due to the tight coupling between variables, the trajectory of the UAV and the clustering outcomes of nodes are tightly tied to the order in which they collected their data and cannot be solved directly, so the problem P is decomposed into three subproblems. Subproblem 1 is about the cluster formation optimization problem, subproblem 2 is the trajectory optimization problem, and subproblem 3 is the optimization problem of the order of data collection of the nodes within the cluster.

First, following grouping each node, it is possible to determine the set CH of cluster head nodes, the number M of cluster head nodes, the nodes C contained in each cluster, and the size C_{size} of each cluster. The best flying trajectory V for the UAV is then determined by using the coordinates of the cluster head node and the data center DC as inputs to the trajectory problem. Finally, based on the known situation of each cluster node C , the order of data collection from the cluster nodes G is determined. Algorithm 1 provides a full description of the process.

3.2. Distance-Based Clustering Method. Based on the above analysis, all nodes should be fairly divided into clusters, and the appropriate cluster head node should be chosen as the area where the UAV will hover to collect the data gathered by the member nodes. With an increase in hovering position, the flight time of the UAV rises, which is not good for keeping the information fresh and thus limits the number of clusters when the nodes are clustered. The more clusters there are, the longer the UAV must fly, and the fewer clusters there are, the better. Wireless sensor nodes have a finite communication range. They are unable to communicate with one another beyond this range. Therefore, when constructing the cluster, the distance between the cluster head node and member nodes must satisfy the requirement of the maximum communication radius of nodes. Subproblem 1 can be expressed as

$$\begin{aligned} \text{P1} : \min_M C_j \\ \text{s.t. (C1)} \quad d_{n, \text{ch}_j} \leq R, j = 1, 2, \dots, M. \end{aligned} \quad (14)$$

In this paper, a combination of maximum-minimum distance clustering and nearest-neighbor clustering is used

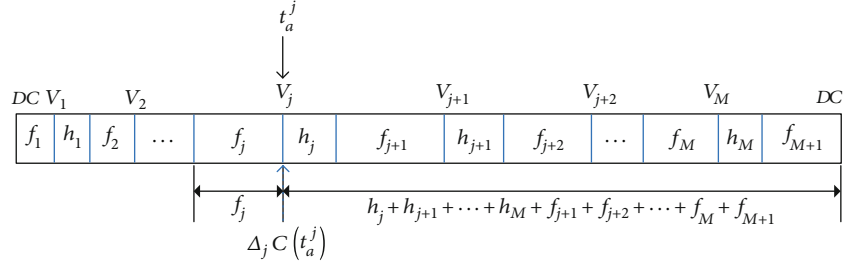
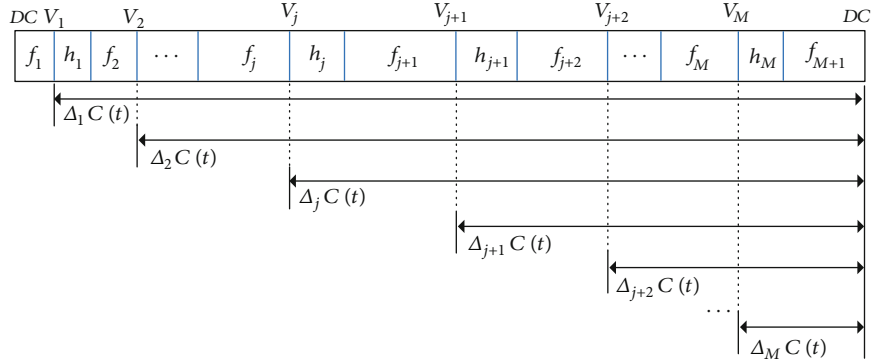
FIGURE 3: Illustration of the age composition of the instantaneous SCAoI of the data in C_j .

FIGURE 4: Illustration of the time spent collecting all data by UAV.

```

Input: coordinates  $L$  of  $N$  nodes in the region, cluster radius  $R$ , UAV flight
      height  $h$ , UAV flight speed  $v$  and other system parameters;
Output: Minimum value of the average SCAoI;
1: Initialization related parameters;
2: for  $i=1: N$  do
3:    $temp\_C(i) = \{cluster\ formation\ results\}$ ;
4:    $temp\_CH(i) = \{cluster\ head\ collection\}$ ;
5:    $temp\_C\_size(i) = \{cluster\ size\}$ ;
6: end for
7:  $M \leftarrow \min(size(CH))$ ,  $C \leftarrow temp\_C(M)$ ,  $CH \leftarrow temp\_CH(M)$ ,
    $C\_size \leftarrow temp\_C\_size(M)$ , obtain the result of subproblem 1;
8: Using the coordinates of  $CH$  and  $DC$ , optimize the UAV trajectory  $V$ ,
   and obtain the optimal solution of subproblem 2;
9: for  $j=1: C$  do
10:  for  $k=1: C\_size(j)$ 
11:   Optimize the order of data collection from member nodes within a cluster
    $G$ , obtain the optimal solution of subproblem 3;
12:  end for
13: end for
14: Solve the problem P.

```

ALGORITHM 1: Framework for solving problem P.

to solve P1. This algorithm uses the Euclidean distance between nodes as the main reference data to decide which cluster head nodes to choose. First, the initial cluster head node might be any node, and then, the next cluster head node is chosen from among the nodes with the greatest distance (which must be greater than the maximum communication radius) from the first cluster head node. Each remaining node's distance from the node that has emerged as the cluster head is determined. The present cluster head

node cannot divide all the nodes into clusters as needed if the maximum value of the minimum distance is greater than the cluster radius. A new cluster head node must be added, and the node corresponding to this maximum value is chosen as the new cluster head node. The cluster head nodes can then all be identified by calculating the greatest value of the minimum separation between the remaining nodes and the cluster head node until it is less than or equal to the cluster-forming radius. Finally, all nodes are clustered according to

Input: Coordinates L of N nodes in the region, cluster radius R ;
Output: The number of clusters M , the set of cluster heads CH , the number of nodes within each cluster C , the size of each cluster $C.size$;

- 1: for $i=1: N$ do
- 2: $ch_1 = n_1 \in N$, records $CH = \{ch_1\}$;
- 3: $ch_2 = n_j, n_j \in N$ and $n_j \notin CH$, n_j is the node with the largest distance from ch_1 among the remaining nodes, record $CH = \{ch_1, ch_2\}$;
- 4: if $\max \min (d_{n_j, CH}) > R, n_j \in N$ and $n_j \notin CH$ do
- 5: $ch_3 \leftarrow$ Nodes with maximum and minimum distance $> R$;
- 6: else Output CH ;
- 7: end if
- 8: end for
- 9: Complete clustering of the remaining nodes using the information from the cluster head with the fewest number.

ALGORITHM 2: Cluster formation algorithm based on maximum-minimum distance.

the idea of nearest-neighbor clustering, which is the principle of proximity. The specific algorithm steps are shown in Algorithm 2.

3.3. Ant Colony Algorithm-Based Trajectory Optimization.

Based on the clustering results, the UAV trajectory problem consisting of M cluster head nodes and a data center DC is solved. Since the flight trajectory of the UAV affects the calculation of the hovering time and flying time in equation (13), subproblem 2 can be written as

$$\begin{aligned}
 \text{P2 : } \min_V \sum_{i=1}^M \frac{i}{M} h_i + \sum_{i=1}^M \frac{i}{M} f_{j+1} \\
 \text{s.t. (C2) } h_j = t_u N_j \\
 \text{(C3) } f_{j+1} = \frac{d_{v_{j,j+1}}}{v}.
 \end{aligned} \tag{15}$$

First of all, prove that P2 is an NP-hard problem.

Proof. According to Algorithm 2, the number of clusters formed, the size of each cluster, the distance between cluster head nodes, and the hovering duration h of the UAV over each cluster head node can both be determined using the coordinates of cluster head nodes. P2 can be viewed as the shortest flight time that solves for DC as the starting point and travels through each cluster head node before arriving back at DC. The shortest time problem is equivalent to the shortest path problem during this operation because the flight speed of the UAV is constant. As in [23], if a certain typical NP-hard problem can be reduced to P2, then it is possible to show that P2 is identical to the NP-hard problem. The description of P2 is basically similar to the typical traveling salesman problem (TSP), which is to find a path that allows a traveler to visit each city once with the shortest total path length, provided that the city coordinates are known. After sorting, simplifying, and mapping each city into cluster head nodes with hover time h and flight time f , then P2 can be basically equivalent to a TSP problem, so P2 is also an NP-hard problem. \square

In this study, the NP-hard problem is solved by the ant colony algorithm because it is typically impossible to tackle NP-hard problems by addressing convex optimization problems. Since ants do not have vision, they cannot intuitively feel the distribution of food and can only rely on the pheromones left by their peers along the foraging process to identify the location of food. Pheromone is a biological hormone that will be volatilized over time after being excreted by ants. Therefore, when more pheromone is accumulated in a certain path, it means that there is more food in that path compared with other paths, which will attract more ants to go to that path to get food.

In this paper, each cluster head node is mapped to a city with hover time h and flight time f , which is the location where the ants need to find food in the ant colony algorithm. First, we initialize the parameters of the system, such as the number of ants, pheromone concentration, pheromone volatility factor, and the maximum number of iterations, so that all ants start to find the path from the coordinates where DC is located. By calculating the probability as part of a path search, we can choose which cluster head node the ants will visit next after they have visited all of the cluster head nodes. Each cluster head node has a corresponding hover time and flight time, so the value of equation (15) is calculated for the path of each ant, and the trajectory that minimizes the value of equation (15) in this iteration is recorded. Then, the pheromone concentration on the path in the system is updated, and the next path finding is performed. Once the maximum number of iterations has been reached, the trajectory with the smallest value of token (15) in all iterations is output. The following is a summary of the precise steps of the algorithm.

Step 1. Initialize the relevant parameters and place N_a ants in the system to make them all start their path exploration from the coordinates of the DC. Use the table allowed to record the nodes that have not been visited and the tube table to record the nodes that have been visited.

Step 2. Determine the next node to be visited and express the probability of the m th ant moving from ch_i to ch_j in round t with probability $P_{ij}^m(t)$.

$$P_{ij}^m(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \times [\eta_{ij}(t)]^\beta}{\sum_{m \in \text{allowed}_m} [\tau_{im}(t)]^\alpha \times [\eta_{im}(t)]^\beta} & \text{if } j \in \text{allowed}_m, \\ 0 & \text{other} \end{cases} \quad (16)$$

where the concentration of pheromones along the route between ch_i and ch_j is $\tau_{ij}(t)$. The length of the path connecting ch_i and ch_j is reciprocal to a heuristic function called $\eta_{ij}(t)$. The pheromone factor α indicates the extent to which the pheromone concentration has an impact on the path when determining the next node to be visited. The β is the heuristic function factor, and both α and β are constants. The table of allowed is used to record the nodes that have been visited by the m th ant, which can be regarded as a set consisting of a sequence of nodes that have been visited, complementary to the tube table, where the entire set is the collection of all data center and cluster head nodes. The next node to be visited is the one with the highest $P_{ij}^m(t)$ value.

Step 3. When all ants have finished visiting all nodes and return to DC, one round of trajectory planning is completed, and then, the value corresponding to the trajectory explored by each ant is calculated according to equation (15).

Step 4. Pheromone update. After a round of trajectory planning is finished, ants will have left behind pheromones along the route. The pheromones on the path between ch_i and ch_j are represented as follows:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{m=1}^{M_a} \Delta\tau_{ij}^m(t), \quad (17)$$

$$\Delta\tau_{ij}^m = \begin{cases} L_m^{-1} & \text{if the } m^{\text{th}} \text{ ant traverses } (i, j) \\ 0 & \text{other} \end{cases}, \quad (18)$$

where ρ is the pheromone volatilization coefficient. Equation (17) can be interpreted as the pheromone after the $t+1$ st round on the path being equal to the pheromone left on the path in the t th cycle plus the added pheromone. The added pheromone is the sum of the pheromones left on the path by all ants, and the size of the pheromone left by each ant is the reciprocal of its path length, as in equation (18).

Step 5. Algorithm iteration and end. When there have been fewer iterations than the maximum amount, the algorithm cycles back to Step 2 and increments the parameter that counts the iterations by 1; when the total number of iterations equals the number allowed, the algorithm iterates to the end and outputs the shortest trajectory.

The algorithmic procedure is described in Algorithm 3.

3.4. Greedy Algorithm-Based Data Collection Sequence. The clustering of system nodes and the flight path of the UAV can be calculated using the answers to P2 and P3 problems. Assuming that the UAV trajectory is V with $v_j = ch_j$, $j \neq 0$, $M+1$, and $j \leq M$, for cluster head ch_j , the moment when

the member nodes start data collection is the moment when ch_{j-1} finishes data transmission to the UAV and flies to ch_j , which can be regarded as the initial moment for collecting information in cluster C_j . As shown in Figure 3, the nodes in cluster C_j use the flight time f_j for data collection, and the moment when the UAV arrives at ch_j is denoted as f_j . According to equation (9), the instantaneous SCAoI of the data collected when the UAV arrives at ch_j as

$$\begin{aligned} \Delta_j C(t) &= \Delta_j C(t_a^j) = \Delta_j C(f_j) = f_j - \sum_{k=1}^{U(f_j)} (1-\beta_k)(u_k - u_{k-1}) \\ &= f_j - \sum_{k=1}^{N_j} (1-\beta_k)t_{ch}^{k-1} = \left| f_j - \sum_{k=1}^{N_j} t_{ch}^{k-1} \right| + \sum_{k=1}^{N_j} \beta_k t_{ch}^{k-1}, \end{aligned} \quad (19)$$

where t_{ch}^{k-1} is the time taken by the $k-1$ st node to send the information to the cluster head node when $k=1$ and $t_{ch}^{k-1} = 0$.

According to the definition of equation (9), the data-gathering process of cluster nodes in different orders affects the instantaneous SCAoI of the collected data when the UAV reaches ch_j . Therefore, it is necessary to optimize the nodes of cluster data collecting order, and subproblem 3 can be written as

$$\begin{aligned} P3 : \min_G \frac{1}{M} \sum_{j=1}^M \left(\left| f_j - \sum_{k=1}^{N_j} t_{ch}^{k-1} \right| + \sum_{k=1}^{N_j} \beta_k t_{ch}^{k-1} \right) \\ \text{s.t. (C3)} f_{j+1} = \frac{d_{v_{j+1}}}{v}. \end{aligned} \quad (20)$$

The correlation coefficient β_k of the k th uploaded node is related to the shortest distance between the previous $k-1$ already uploaded nodes. An exhaustive method is utilized to enumerate every collection order to arrive at the best value if we want to acquire the best instantaneous SCAoI. However, when the number of member nodes is large, this approach can obtain the exact optimal value, but it will consume a lot of computational resources and increase the complexity of the algorithm. Therefore, in this paper, we use a greedy algorithm to select the node that can ensure the smallest value of the objective function in the current state from the nodes that have not yet collected data and repeat the cycle until all nodes have finished uploading, so that we can obtain the data collection order that makes the instantaneous SCAoI at the cluster head node suboptimal. Algorithm 4 provides a description of the steps of the algorithm.

4. Simulation Results and Analysis

The proposed algorithm is simulated using MATLAB, and the computer processor used is a dual-core quad thread processor of Intel Core 8th generation. The simulation scenario is shown in Figure 1. The sensor nodes N are randomly

```

Input:  $M$  cluster head coordinates, coordinates of data center DC, maximum
iteration  $I_{\max}$  and other related parameters;
Output: Shortest path trajectory  $V$ , the optimal value of P2;
1: Initialization related parameters;
2: for  $i=1:I_{\max}$  do
3:   for  $j=1:N_a$  do
4:     Update  $tube, allowed$ 
5:     Determine the next visited cluster head node according to equation (16);
6:   end for
7:   Calculate the value corresponding to the path found by each ant according
to equation (15);
8:   Update the pheromone according to equation (17) (18);
9: end for
10: Choose the trajectory that minimizes equation (15), denoted as  $V$ , the
corresponding value is the optimal value of P2.

```

ALGORITHM 3: Trajectory optimization algorithm based on ant colony algorithm.

```

Input: Node coordinates  $L$  in each cluster, the set of cluster heads  $CH$ , the
node situation  $C$  in each cluster;
Output: The order of node uploads in each cluster, the value of the suboptimal
function of P3;
1: for  $i=1:M$  do
2:   for  $j=1:N_i$  do
3:      $table \leftarrow$  Record the nodes that have been uploaded;
4:      $list \leftarrow$  Record nodes that have not been uploaded yet;
5:     According to equation (20), the next node that can lower the objective
value of the function is found from the  $list$ .
6:   end for
7: end for

```

ALGORITHM 4: Greedy algorithm based on data collection order optimization.

distributed in a rectangular area with coordinates (0, 0), (0, 300), (300, 0), and (300, 300) as vertices, and the data center DC has coordinates (350, 150). The relevant parameters used for the simulation were set with reference to the literature [19, 20], and the specific values are shown in Table 2.

The algorithm in this study is compared to the algorithms in [19, 20] to evaluate the effectiveness of the algorithm presented in this research. Wireless sensor nodes are distributed in a rectangular area of 300 m \times 300 m, and a UAV is used to collect the data collected by the sensor nodes. The UAV always maintains a constant flight height and speed, and the simulation parameters are set as shown in Table 1 if no special instructions are given. In contrast to this paper, [19] aims to minimize the average information age of the nodes by optimizing the UAV trajectory, which requires the UAV for information collection to fly over each node and communicate directly with each node in the region. To further illustrate the superiority of the proposed algorithm in this paper, a comparison with the algorithm in [20] is also made. In [20], the same method of cluster formation is used to reduce the UAV hovering position to optimize the trajectory to further improve the freshness of the collected data, but it does not consider the influence of

TABLE 2: Simulation parameter table.

Parameter	Value
Maximum cluster formation ratio R (m)	50
System bandwidth B (KHz)	20
Transmit power of the node p_i (w)	0.1
Noise power p_n (dBm)	-110
Channel power gain at a reference distance of 1 m β_0 (dB)	-60
Path loss factor α	3.5
Constant parameters influenced by the environment β_1, β_2	10, 0.03
Carrier frequency f_c (GHz)	2
Extra path loss $\zeta_{LoS}, \zeta_{NLoS}$ (dB)	0, 20
UAV flight height h (m)	50
UAV flight speed v (m/s)	15
Data package size L (byte)	2560

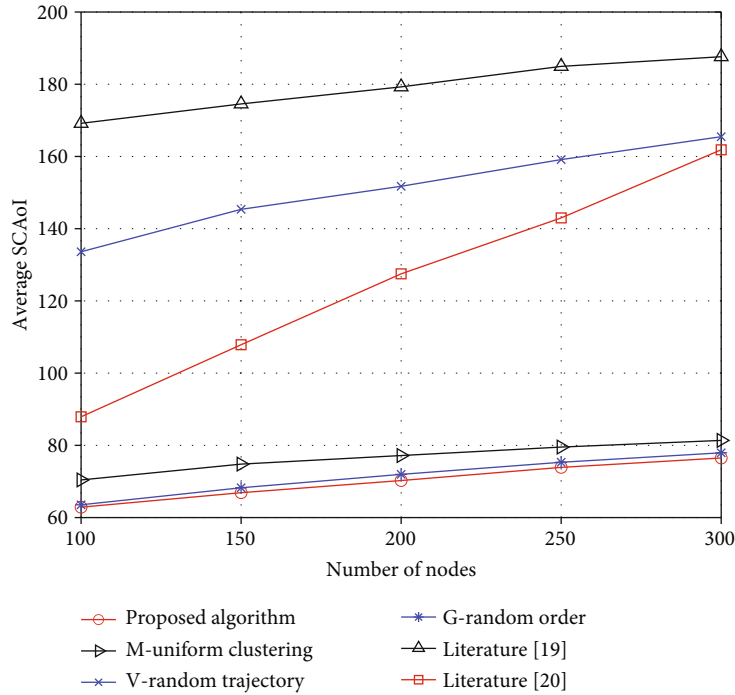


FIGURE 5: The variation of average SCAoI with the number of nodes in this paper and [19, 20].

the correlation between the nodes within the cluster on the collected data due to the location factor. The results of the simulation are shown in Figure 5.

Figure 5 gives the average SCAoI versus the number of nodes in the region, where the cluster formation radius $R = 50\text{m}$, flight height $h = 50\text{m}$, flight speed $v = 15\text{m/s}$, packet size $L = 2560$ byte, and degree of correlation $\rho = 10^2$. As nodes become more numerous, the average SCAoI of both this paper and [19] as well as [20] increases accordingly due to the increased time consumption of each process of information collection by the UAV. For the algorithm described in this work and [20], the increase in the number of nodes means that the UAV needs to collect more data, and therefore, the hovering time of the UAV increases. Additionally, there will be more clusters, which will lengthen the flight time of the UAV. In [19], when the number of nodes rises, the UAV must fly to every node to gather data, which takes up a lot of flight time. Therefore, both in this paper and in [19], the average SCAoI shows a rising trend as the number of nodes increases. When the number of nodes is the same, the average SCAoI in [19] is significantly larger than the algorithm in this paper and [20], which is in the middle. The primary reason is that, according to [19], the UAV visits each node to gather data, and the number of hovering positions is equal to the number of nodes. As a result, the average SCAoI and flight time of the UAV increase as the number of nodes increases, lowering the freshness of the data. In this paper, we effectively reduce the number of locations where UAVs need to hover by clustering, thus shortening the flight time of UAVs. Although [20] also uses clustering to reduce the hovering position of the UAV, the number of clusters cannot be effectively reduced, which also affects the optimization of the UAV trajectory and thus

the freshness of the data collected by the UAV. And as can be seen from the remaining three curves, the optimization of UAV trajectories contributes greatly to the improvement of information freshness compared to the clustering and data upload order. The reason is that with the same parameter settings, the cluster member nodes take tens of milliseconds to transmit a packet to the cluster head node, the cluster head node takes a few milliseconds to transmit a packet of the same size to the UAV, and the UAV takes a few seconds to fly to the next cluster head node. Such an order-of-magnitude relationship makes the optimization process far more effective for trajectories than for the other two variables. Overall, the algorithm proposed in this paper can improve the average SCAoI of the system by about 61%. From the point of view of the time complexity of the algorithm, the time complexity O_{AP} of the proposed algorithm in this paper and the time complexity O_{CL} of the algorithm in [19] can be denoted as $O_{\omega}(I_{\max} N_a M_{\text{ch}}^{\omega})$ and $\omega = [AP, CL]$, where M_{ch}^{AP} is the number of clusters of the proposed algorithm in this paper and the number of clusters in [19] is denoted as M_{ch}^{CL} . From the above analysis, it can be concluded that when the number of nodes N is the same, since the algorithm in [19] does not have a clustering step, each node can be considered a cluster head, i.e., $M_{\text{ch}}^{CL} = N$. And in this paper $M_{\text{ch}}^{AP} < N$, so it can be obtained as $M_{\text{ch}}^{AP} < M_{\text{ch}}^{CL}$, i.e., $O_{AP} < O_{CL}$. The time complexity of the proposed algorithm in this paper is also lower than in [19].

Figure 6 gives the variation of the average SCAoI as the number of nodes increases for different cluster formation radius, where the flight height $h = 50\text{m}$, flight speed $v = 15\text{m/s}$, packet size $L = 2560$ byte, and degree of correlation $\rho = 10^2$. As the cluster radius rises, it is evident from the graphic that the average SCAoI drops. This is because

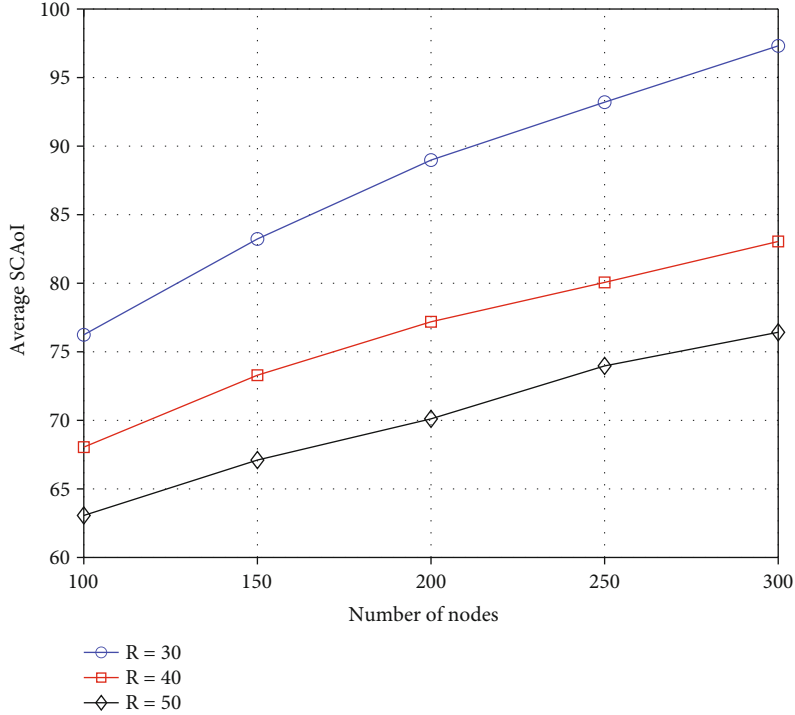


FIGURE 6: Variation of average SCAoI with the number of nodes for different cluster radius.

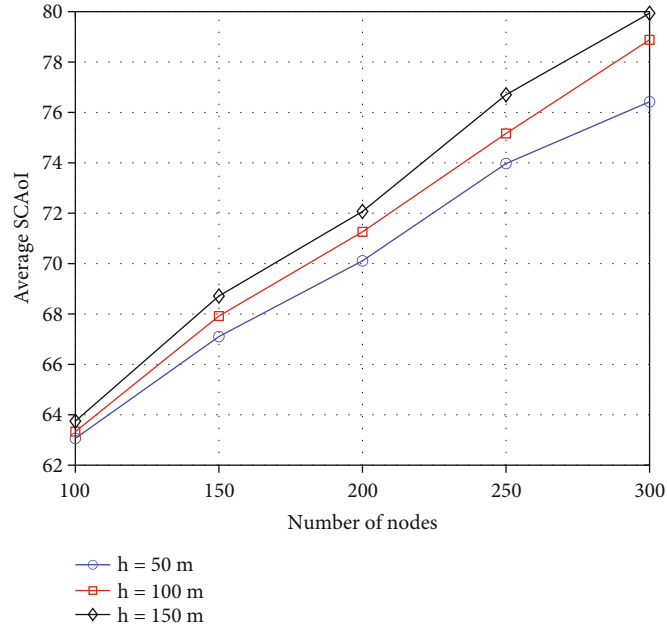


FIGURE 7: Variation of average SCAoI with the number of nodes at different flight altitudes.

when the number of nodes and node distribution are the same, the larger the cluster radius, the fewer clusters there will be. As a result, the hovering position of the UAV, flight time, and average SCAoI will all be reduced, while the freshness of information will also be increased.

The average SCAoI changes when there are more nodes under various UAV flight heights, as seen in Figure 7, where

the cluster formation radius $R = 50\text{m}$, flight speed $v = 15\text{m/s}$, packet size $L = 2560$ byte, and degree of correlation $\rho = 10^2$. When there are the same number of nodes, the varying UAV flight altitudes mostly influence how quickly data is transmitted from the cluster head node to the UAV. The transmitting power of the cluster head node is fixed, so as the flight altitude of the UAV rises, the data transmission

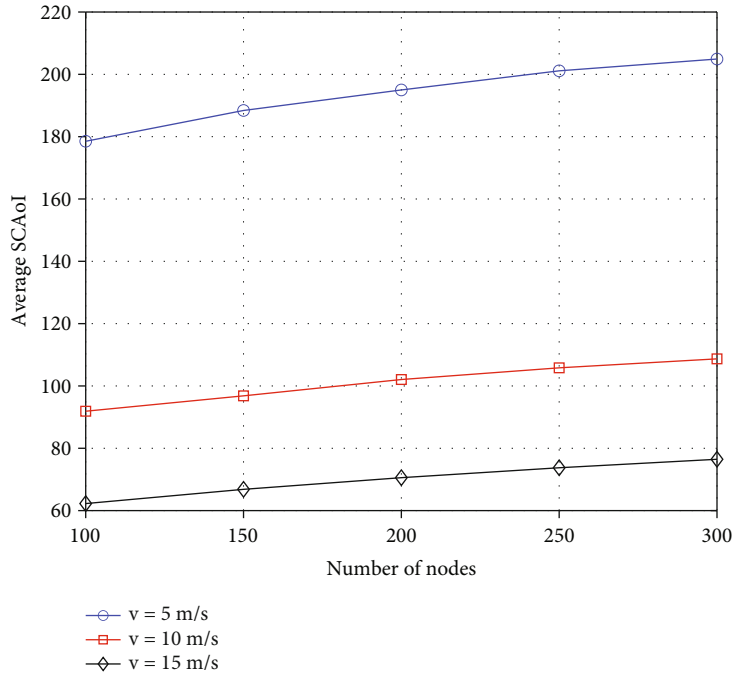


FIGURE 8: Variation of the average SCAoI with the number of nodes at different flight speeds.

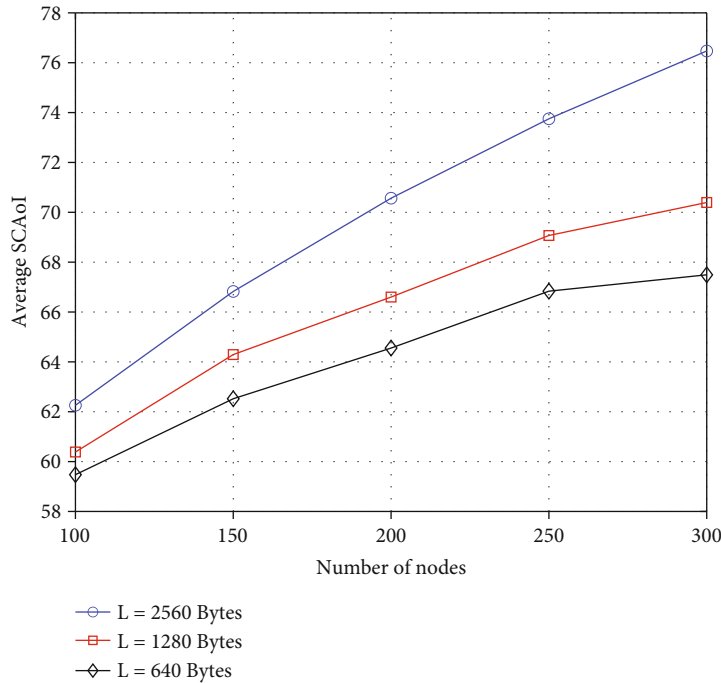


FIGURE 9: Variation of average SCAoI with the number of nodes for different data volumes.

rate between the cluster head node and the UAV declines. This is because from Eqs. (3)–(5) and the free path loss $PL_i = 20 \log_{10}(4\pi hf_c/c)$ in Section 2.3, it is known that the free path loss increases as the flight altitude h increases, causing the denominator part of the log function in Eq. (3) to increase, resulting in a decrease in R_u . As a result, more time must be spent transmitting the same number of data

packets, which lengthens the hovering time of the UAV. As a result, the average SCAoI will rise as the flight of UAV altitude rises, provided that there are the same number of nodes.

The average SCAoI changes when there are more nodes under different flight speeds, as seen in Figure 8, where the cluster formation radius $R = 50\text{m}$, flight height $h = 50\text{m}$,

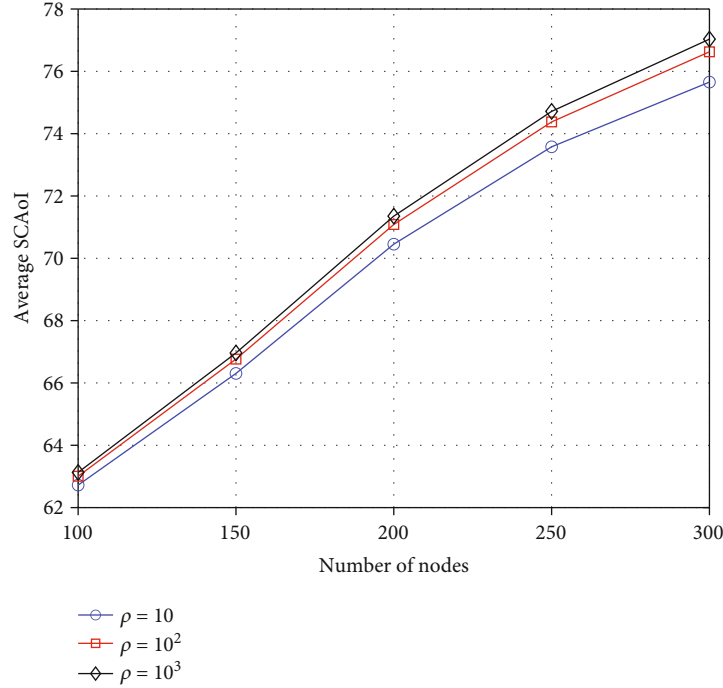


FIGURE 10: Variation of average SCAoI with increasing number of nodes for different degrees of correlation.

packet size $L = 2560$ byte, and degree of correlation $\rho = 10^2$. The shortest path length and flight trajectory of the UAV will be the same under the assumption that the number of nodes, distribution of nodes, and cluster formation are all constant. Accordingly, the shorter the flight time, the smaller the flight speed of the UAV will be, and the average SCAoI will decrease as flight speed increases.

The average SCAoI changes when there are more nodes under different data volumes, as seen in Figure 9, where the cluster formation radius $R = 50\text{m}$, flight height $h = 50\text{m}$, flight speed $v = 15\text{m/s}$, and degree of correlation $\rho = 10^2$. The effect of various data quantities on the average SCAoI is mostly apparent in two aspects when there are the same number of nodes. On the one hand, due to $R_u > R_{ch}$, the amount of data is the same when $t_u < t_{ch}$. When the amount of data increases, the increment of t_u is smaller than the increment of t_{ch} , and the hovering time of the UAV will increase as a result. On the other hand, the instantaneous SCAoI when the UAV reaches the cluster head node decreases as stated by equation (19); then, the average SCAoI also decreases, but its decrease is small. The cluster member nodes use TDMA to transmit data to the cluster head node, and the transmission time from the member nodes to the cluster head node is equal to the time slot length. The increment in UAV hovering time due to the increase in data volume is greater than the decrease in instantaneous SCAoI, so the average SCAoI increases with the growth in data volume.

The average SCAoI changes when there are more nodes under various correlation levels, as seen in Figure 10, where the cluster formation radius $R = 50\text{m}$, flight height $h = 50\text{m}$, flight speed $v = 15\text{m/s}$, packet size $L = 2560$ byte. When the number of nodes is the same, it is known from equation

(10) that when the correlation degree of the space is larger, its correlation coefficient β is also larger, the nodes' data collection has a higher correlation, and the corresponding average SCAoI will be larger, so the average SCAoI increases with the increase of the correlation degree.

5. Summary

In this paper, we study the problem of age-based optimization in information collection systems with a UAV. An optimization problem for the number of joint clusters, UAV flight trajectories, and data collection order of nodes within clusters is proposed. Minimize the average SCAoI of all nodes while ensuring the sensor node communication distance and UAV parameters. In order to solve the proposed problem, we decompose it into three subproblems. First, the maximum-minimum distance algorithm based on clustering is used to obtain the number of clusters and determine the cluster head node coordinates. Then, it is proved that the UAV trajectory problem in this paper is a typical NP-hard problem that can be solved by using the ant colony algorithm. The data collection order of the nodes in the cluster is solved by the greedy algorithm. The suboptimal solution of the proposed problem is obtained by solving the three optimization problems separately. Simulation results show that the algorithm proposed in this paper outperforms comparative literature algorithms in reducing the average SCAoI of nodes and improving the freshness of information. In the future, UAV caching will be the main research direction to consider the problem of freshness of user-requested content or the problem of joint caching optimization for multiple UAVs and users.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61971239 and 92067201) and the Jiangsu Provincial Key Research and Development Program (No. BE2022068-2).

References

- [1] X. Ke, H. Hu, B. Ai, Y. Zhang, and L. Pei, "Wireless network design for information freshness priority in 6G," *Journal of Internet of Things*, vol. 4, no. 1, pp. 80–91, 2020.
- [2] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: how often should one update?," in *2012 Proceedings IEEE INFOCOM*, Orlando, FL, USA, 2012.
- [3] Z. Liu, L. Huang, B. Li, and B. Ji, "Anti-aging scheduling in single-server queues: a systematic and comparative study," *Journal of Communications and Networks*, vol. 23, no. 2, pp. 91–105, 2021.
- [4] T. Zhang, S. Chen, and Z. Chen, "Internet of Things: the optimal generation rates under preemption strategy in a multi-source queuing system," *Entropy*, vol. 23, no. 8, 2021.
- [5] M. Moltafet, M. Leinonen, and M. Codreanu, "On the age of information in multi-source queueing models," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 5003–5017, 2020.
- [6] S. Zhang, L. Wang, H. Luo, X. Ma, and S. Zhou, "Age of information and delay tradeoff with freshness-aware mobile edge cache update," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019.
- [7] S. Zhang, L. Wang, H. Luo, X. Ma, and S. Zhou, "AoI-delay tradeoff in mobile edge caching with freshness-aware content refreshing," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5329–5342, 2021.
- [8] J. Zhong, R. D. Yates, and E. Soljanin, "Two freshness metrics for local cache refresh," in *2018 IEEE International Symposium on Information Theory (ISIT)*, Vail, CO, USA, 2018.
- [9] F. Jiang, X. Liang, C. Sun, and J. Wang, "Caching and update strategy based on content popularity and information freshness for fog radio access networks," *Journal of Electronics & Information Technology*, vol. 44, no. 9, pp. 3108–3116, 2022.
- [10] M. Zhou, J. Li, J. Yuan et al., "An architecture for aoI and cache hybrid multicast/unicast/D2D with cell-free massive MIMO systems," *IEEE Access*, vol. 11, pp. 43080–43088, 2023.
- [11] L. Wang, J. Sun, Y. Sun, S. Zhou, and Z. Niu, "A UoI-optimal updating policy for timely status information with resource constraint," *Entropy*, vol. 23, no. 8, 2021.
- [12] M. Khari, R. Dalal, A. Sharma, and B. Mehta, "Person identification in UAV shot videos by using machine learning," in *Multimodal Biometric Systems: Security and Applications*, pp. 45–60, CRC Press, 2021.
- [13] A. Rajagopal, A. Ramachandran, K. Shankar et al., "Fine-tuned residual network-based features with latent variable support vector machine-based optimal scene classification model for unmanned aerial vehicles," *IEEE Access*, vol. 8, pp. 118396–118404, 2020.
- [14] A. Rajagopal, G. Prasad Joshi, A. Ramachandran et al., "A deep learning model based on multi-objective particle swarm optimization for scene classification in unmanned aerial vehicles," *IEEE Access*, vol. 8, pp. 135383–135393, 2020.
- [15] G. Wu, Q. Liu, J. Xu, Y. Miao, and M. Pustišek, "Energy efficient task caching and offloading in UAV-enabled crowd management," *IEEE Sensors Journal*, vol. 22, no. 18, pp. 17565–17572, 2022.
- [16] L. Wang, S. Zhang, X. Li, and H. Luo, "Age-constrained dynamic content replacing and delivering for UAV-assisted context awareness," *China Communications*, vol. 19, no. 7, pp. 277–293, 2022.
- [17] Z. Qin, X. Zhang, X. Zhang, B. Lu, Z. Liu, and L. Guo, "The UAV trajectory optimization for data collection from time-constrained IoT Devices: a hierarchical deep Q-network approach," *Applied Sciences*, vol. 12, no. 5, 2022.
- [18] X. Zhou, W. Wang, N. U. Hassan, C. Yuen, and D. Niyato, "Age of information aware content resale mechanism with edge caching," *IEEE Transactions on Communications*, vol. 69, no. 8, pp. 5269–5282, 2021.
- [19] A. A. Al-Habob, O. A. Dobre, and H. V. Poor, "Age- and correlation-aware information gathering," *IEEE Wireless Communications Letters*, vol. 11, no. 2, pp. 273–277, 2022.
- [20] J. Liu, P. Tong, X. Wang, B. Bai, and H. Dai, "UAV-aided data collection for information freshness in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2368–2382, 2021.
- [21] S. Patten, B. Krishnamachari, and R. Govindan, "The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 4, no. 4, 2004.
- [22] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Letters On Wireless Communications*, vol. 3, no. 6, pp. 569–572, 2014.
- [23] N.-T. Nguyen and B. H. Liu, "The mobile sensor deployment problem and the target coverage problem in mobile wireless sensor networks are NP-hard," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1312–1315, 2019.