Hybrid Sensor Network with Edge Computing for AI Applications of Connected Vehicles

Maoqiang Wu[1,2], Xumin Huang[1,2], Beihai Tan[1,3], Rong Yu[1,3]

[1] School of Automation, Guangdong University of Technology, China
[2] Guangdong Key Laboratory of IoT Information Technology, China
[3] Guangdong-HongKong-Macao joint Laboratory for Smart Discrete Manufacturing, China
maoqiang.wu@vip.163.com, huangxu_min@163.com, bhtan@gdut.edu.cn, yurong@gdut.edu.cn

Abstract

For supporting Artificial Intelligence (AI) applications of connected vehicles, sensing data is collected by Service Providers (SPs) as input to AI models to execute model inference. Based on the inference results, SPs respond to users' requests. To ensure the quality of service (QoS), enhancing the sensing quality of data collection and shortening the latency of inference execution are two crucial issues. To address these issues, we propose the integration of hybrid sensor network and edge computing. Hybrid sensor network enables the cooperation of dynamic vehicular nodes and static sensor nodes for improving sensing quality. Edge computing fulfills local processing of sensing data in edge servers to improve the overall performance of services. After that, we study the problem for SP-side assigning sensing tasks and corresponding rewards between vehicular nodes and sensor nodes. A three-party Stackelberg game is leveraged to design the task assignment scheme, which allows the three parties to reach a deal with optimal pricing strategies and sensing strategies. We also develop a resource allocation scheme which enables SPs to optimally allocate computation resource of edge servers for minimizing the delay of inference execution. Numerical results indicate that the proposed task assignment scheme based on hybrid sensor network outperforms the schemes based on pure vehicular nodes or sensor nodes. The designed resource allocation scheme achieves the convergence of 4.2 times faster than that of the greedy algorithm.

Keywords: AI applications of connected vehicles, Hybrid sensor network, Edge computing, Three-party Stackelberg game

1 Introduction

Artificial Intelligence (AI) applications have been widely implemented for connected vehicles. Service Providers (SPs) collect image data of target objects in urban areas as input to AI models (e.g., Deep Neural Network model) and execute model inference. According to the inference results, SPs respond to users' requests.

For supporting AI applications of connected vehicles, data collection of transportation system is necessary. To extend the spatial coverage of observing areas, we pay attention to the utilization of hybrid sensor network, which consists of static wireless sensor network (WSN) and dynamic vehicular crowdsensing network (VCN). Here, WSN could perform regular sensing tasks by deploying static sensor nodes in cities [1]. But it is hard to use a large-scale WSN due to the enormous economic expense. VCN employs moving vehicles to sense their surrounding environment and achieves higher coverage due to the natural mobility of vehicles [2]. Nevertheless, the number of vehicles for gathering data cannot be guaranteed sometimes, for example, at night. Therefore, we consider the integration of these two types of networks for obtaining sufficient data sources.

In addition, edge computing is exploited to improve the overall performance of AI applications. First, the massive collected data should be locally stored and analyzed for reducing bandwidth consumption instead of being transferred to a remote cloud. Second, distributed data management has excellent advantages in coping with large-scale data with higher efficiency, such as load balancing, decreasing response time, and fault tolerance. Last but not least, the gathered data generally has the feature of local relevance. This means that the data had better be processed proximal to users for the better quality of service (QoS) [3]. Based on the advantages, we consider that an SP can rent edge servers to perform the AI model inference locally for low latency, reduced energy consumption, and enriched location awareness.

In summary, we integrate hybrid sensor network with edge computing into a new system for provisioning AI applications of connected vehicles. More specifically, the hybrid sensor network combines fixed sensors and moving vehicles on the roads to acquire

*Corresponding Author: Rong Yu; E-mail: yurong@gdut.edu.cn
DOI: 10.3966/160792642020092105023

sufficient image data in a specific interesting area. Then SPs schedule dedicated edge servers for local inference by taking the gathered image data as input data of AI models. As a consequence, this also gives rise to two essential optimization problems consisting of a task assignment problem in data collection and a resource allocation problem in model inference.

Next, we focus on solving the problems above for optimizing SP-side decision making in the system implementation. In the system, vehicular nodes and sensor nodes carry out sensing tasks published by an SP to get rewards. The SP, vehicular nodes, and sensor nodes are three independent parties and aim to maximize their utilities. To solve the task assignment problem, we leverage a three-party Stackelberg game to reach a satisfying deal about data collection. In the game, the SP acts as a leader to decide the optimal pricing strategies, while the vehicular nodes and sensor nodes become followers to determine their sensing strategies accordingly. On the other hand, as various data from different nodes is uploaded to a nearby edge server for local inference, we formulate the resource allocation problem as a nonconvex problem and present a suboptimal algorithm for minimizing the inference latency. Finally, simulation results are provided to demonstrate the efficiency and effectiveness of the schemes above.

The main contributions of this paper are summarized as follows.

(1) We propose a new system by jointly using a hybrid sensor network and edge computing for provisioning AI applications of connected vehicles. We describe the system implementation with crucial entities, which mainly include users in AI applications, vehicular nodes and sensor nodes for data collection, and an SP renting edge servers for local inference.

(2) We elaborately design the task assignment scheme by using a three-party Stackelberg game. In the scheme, the SP achieves the maximum sensing satisfaction with the minimal monetary cost, while both the vehicular nodes and sensor nodes get the optimal revenues.

(3) We develop the resource allocation scheme based on an iteration algorithm so that the SP is able to obtain the inference results promptly. The proposed algorithm enables the convergence rate of 4.2 times faster than that of the greedy algorithm.

The rest of this paper is organized as follows. We summarize the related work in Section 2. In Section 3, we describe the system model of the proposed system. We formulate the task assignment problem in Section 4 and provide the optimal solution in Section 5. In section 6, we study the resource allocation problem and propose a suboptimal solution. Simulation results are presented in Section 7. Then we conclude the paper in Section 8.

2 Related Work

Much work has been done for studying data collection and edge computing in supporting AI applications of connected vehicles. We shall respectively introduce existing work in these two aspects.

2.1 Data Collection for Vehicular Service

Data collection schemes can be divided into WSN and VCN. WSN is widely used for the vehicular environment [1]. The static sensor nodes deployed along the road shoot passing vehicles by cameras, and the data supports the analysis and prediction of vehicle mobility and traffic congestion. However, WSN is difficult for large-scale deployment due to the tremendous cost.

VCN utilizes vehicular sensors for acquiring various data. Due to the natural mobility of vehicles, VCN achieves more substantial spatial coverage with less deployment expense when compared with WSN. When participating in data collection and transmission, vehicles consume a considerable amount of resource and have a risk of privacy disclosure. Thus, some work focused on incentivizing vehicles for executing sensing tasks. In [4-5], recruited mechanisms were studied for maximizing the spatial coverage with limited incentive budget. A study in [6] aimed at minimizing the incentive cost when satisfying the requirement of spatial coverage. In [7-8], game theory-based incentive mechanisms were proposed, where the SP chooses its pricing strategy and vehicles decide their sensing strategies accordingly. Nevertheless, it is difficult to recruit sufficient vehicles to collect data sometimes, for example, at night.

To ensure the spatial coverage, we have proposed a hybrid sensor network in our previous work [7]. The hybrid sensor network allows the cooperation of VCN and WSN. Furthermore, in this paper, we pay attention to a detailed application scenario where the proposed hybrid sensor network is utilized to perform image data collection tasks. Meanwhile, we identify the difference between the two kinds of nodes for sensing in the network. The mathematic model with specific parameters is also given to adapt to the practical application.

2.2 Edge Computing for Vehicular Crowd-sensing

Management and analysis of the collected data take essential parts in vehicular services and consume computation and storage resources. Edge/fog computing is an attractive paradigm that provides distributed resources in the network edge [9-10]. Instead of uploading the collected data to a remote cloud, some work suggested localized management and processing the data [11]. In [12-13], systems were proposed to locally analyze the collected data and make decisions in fog servers to reduce bandwidth consumption and centralized burden. In [14], a framework was developed to integrate deep learning-based data validation and edge computing-based local processing. The resource allocation of edge/fog servers is crucial for reducing the energy consumption and processing latency [15-16]. Some work has proposed the resource allocation method, such as reinforcement learning (RL) methods [17-18]. However, RL methods spend time cost to learning optimal action selection. In this paper, we use the edge servers to perform AI model inference. In order to respond to users more quickly, we propose a resource allocation scheme without learning process to minimize the inference latency.

3 System Model

In this section, we introduce the proposed system and describe the procedure for provisioning AI applications.

3.1 System Entities

As shown in Figure 1, we consider the system over one urban area. The main entities are summarized as follows.

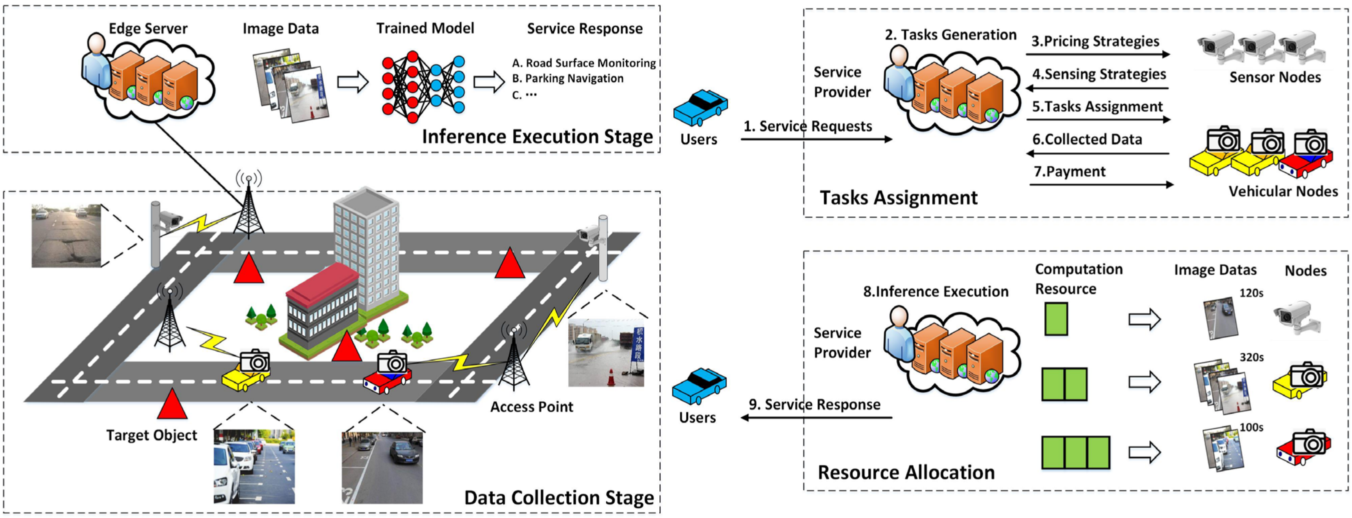

Figure 1. The system integrating hybrid sensor network and edge computing for AI applications of connected vehicles

Users send requests to the SP for AI applications. The required services have the feature of local relevance and delay-sensitive. Take parking navigation service as an example. Users want to obtain the locations of unoccupied parking spaces before they arrive at the destination.

SP collects image data as input of AI model by incentivizing the hybrid sensor network and executes AI inference by renting geographically distributed edge servers. According to the inference results, the SP responds to users' requests timely. For example, the SP gathers image data of roadside and recognizes unoccupied parking spaces for supporting parking navigation service.

Hybrid sensor network consists of a group of dynamic vehicular nodes and multiple static sensor nodes. Vehicular nodes are vehicles passing through the urban region and equipped with onboard cameras. Due to the mobility feature, vehicular nodes can take images covering multiple angles of the target objects. But the motion of vehicles also brings image blur problem. Sensor nodes are fixed camera sensors deployed over the urban region and have their own working duties, such as street monitoring. When the number of vehicular nodes is insufficient, sensor nodes are recruited for compensation.

Edge servers are rented by the SP and have sufficient computation and storage resources for supplying local data management and AI inference execution. Compared with transmitting the data to cloud data centers, localized inference execution in edge servers reduces the burden of data transmission and the latency of service response.

3.2 System Procedure

The procedure of the proposed system includes four stages. In the service request stage, users send requests to the SP for AI applications. The service requests include the location of target regions and tolerant response delay. Then SP generates sensing tasks in the edge server of the target region. The sensing tasks imply required data size, location of target objects, tolerant latency for completing the tasks.

In the data collection stage, the SP needs to assign sensing tasks and corresponding rewards. The SP desires to assign sensing tasks to achieve higher sensing quality by considering coverage ability and motion blur. Meanwhile, both vehicular nodes and sensor nodes want to gain more rewards. A tasks assignment scheme is needed for the three parties to reach a deal with satisfying pricing strategies and sensing strategies. After that, vehicular nodes and sensor nodes carry out the assigned sensing tasks and gain the agreed rewards.

In the inference execution stage, the SP exploits the edge server to carry out model inference based on the collected data. For the data uploaded by each node, the SP generates a corresponding inference task. All the inference tasks are executed in a parallel computing manner. The latency of performing an inference task depends on its available computation resource. In order to shorten the latency of inference execution, a resource allocation scheme is needed to optimally allocate the computation resource of the edge server to each inference task.

In the service response stage, the SP responds to users based on the inference results. The accuracy of inference results depends on the sensing quality of the nodes. The response delay is related to the time consumption of executing sensing tasks and inference tasks. Thus, the task assignment scheme and the resource allocation scheme are crucial for improving the QoS.

4 Task Assignment for Data Collection

We consider the AI application scenario of the hybrid sensor network which consists one SP, dynamic vehicular nodes, and static sensor nodes. The task assignment problem is formulated as a three-party Stackelberg game.

4.1 Sensing Tasks and Sensing Quality

The generated sensing tasks specify required data size D, a set of target objects I, and the maximum tolerant delay time T for carrying out the sensing tasks. The information of target object $i \in I$ is a tuple $\left(l_i, \vec{l_i}, \theta\right)$, where l_i is the location of i, $\vec{l_i} = \left\{\overrightarrow{l_i^1}, \overrightarrow{l_i^2}, \cdots \overrightarrow{l_i^K}\right\}$ is the different K aspects of i, $\overrightarrow{l_i^k}$ is a vector indicated by an angle in $\left[0, 2\pi\right]$ with 0 degree is the one pointing to the right (east on the map), θ is the effective angle. The effective angle range to $\overrightarrow{l_i^k}$ is $\left[\overrightarrow{l_i^k} - \theta, \overrightarrow{l_i^k} + \theta\right]$. The effective angle ranges of one target object's different aspects are assumed to be not overlapped.

There exists a set of nodes S (including vehicular nodes and sensor nodes) equipped with cameras for carrying out the sensing tasks. The camera information of node $s \in S$ is a tuple $\left(d_s, \varphi_s, \vec{l_s}\right)$, where d_s is the effective range of the camera, beyond which people can hardly distinguish anything in the photo, φ_s is the field-of-view, which implies how wide the camera can catch, and $\vec{l_s}$ is the orientation of the camera. When s takes a photo P_{si} of i in a location, $\vec{l_s}$ points to $\vec{l_i}$. The picture P_{si} covers i if the range of P_{si} includes i. The picture P_{si} covers $\overrightarrow{l_i^k}$ if the angle between $\overrightarrow{l_i^k}$ and $\vec{l_s}$ is smaller than θ. According to [19-20], the sensing quality of a node depends on its coverage ability for the target objects, i.e., the number of the target objects' aspects which the photos taken by the nodes can cover. We use an indicator α_{si}^k to imply whether s can take photos covering $\overrightarrow{l_i^k}$. If the aspect is covered, $\alpha_{si}^k = 1$, otherwise, $\alpha_{si}^k = 0$. Therefore, the coverage ability of s is defined as $\sum_{i \in I} \sum_{k \in K} \alpha_{si}^k$. According to [21], the sensing quality also relate to the motion of the camera. The speed fluctuation of vehicular nodes (learned from the accelerometer sensor in vehicles) leads to increasing image blur. So, the sensing quality is defined as follows.

Definition 1 (Sensing Quality): Sensing quality measurement is based on coverage ability and motion blur. Formally, it can be defined as:

$$q_s = \frac{\sum_{i \in I} \sum_{k \in K} \alpha_{si}^k}{1 + y_s}, \quad \textbf{(1)}$$

where y_s is the average acceleration of s, which can be observed from the historical driving manner. If s is a static sensor node, $y_s = 0$.

The sensing satisfaction of the SP at one node depends on not only the node's sensing quality but also the node's collected data quantity and time consumption of performing the sensing task [3]. The image data has an explicit lifetime of utility. For example, an image of unoccupied parking space may be valid for 10 minutes, and after that, the parking space may be occupied. Thus, the sensing satisfaction of the SP is defined as follows.

Definition 2 (Sensing Satisfaction): The sensing satisfaction of the SP at s is based on the node's sensing quality, collected data size, and the time consumption of executing the sensing task. Formally, the sensing satisfaction at s can be defined as:

$$Q_s = \delta q_s f_s^{se} t_s^{se} \left(T - t_s^{se} - t_s^{tr}\right), \quad \textbf{(2)}$$

where δ is a parameter, f_s^{se} is the sensing frequency of s, t_s^{se} is the time consumed by s for sensing, t_s^{tr} is the time consumed for uploading data. $f_s^{se} t_s^{se}$ denotes the data size of s and $T - t_s^{se} - t_s^{tr}$ denotes how much time that s finishes the sensing task earlier than the deadline. The sensing satisfaction of the SP at all nodes is $Q = \sum_{s \in S} Q_s$.

4.2 Utilities of Three Parties

We consider a set of vehicular nodes $M \subset S$ recruited to execute sensing tasks. The moving path of vehicular node $m \in M$ is denoted as a set of location coordinate L_m, and the path length is denoted as $|L_m|$. Vehicular node m is driven at average speed v_m along the path and utilizes the onboard camera to take photos of target objects. The residence time of m in the urban region is $t_m^{se} = |L_m| / v_m$, which is also the time consumed for sensing. The sensing strategy of m is its adjustable sensing frequency f_m^{se}, and its collected data size is $f_m^{se} t_m^{se}$. The collected image data is considered to be similar and redundant if the time and locations of the photos are too close [20]. To ensure the temporal-spatial diversity of the collected data, we set $f_m^{se} \le \beta v_m$, where β is a parameter. When m moves out of the urban region and cannot cover the target objects again, it uploads the collected data to a nearby access point (AP). The channel transmission rate is expressed as

$$r_m = B \log_2 \left(1 + \frac{P_0 h_m \left(d_m \right)^{-r}}{w_0} \right), \tag{3}$$

where B is the bandwidth of the leased transmission channel between m and the AP, P_0 is the transmission power, h_m is the channel gain, w_0 is the power level of white noise, and d_m is the distance between m and the AP. Here we assume that all the vehicular nodes have the same bandwidth and the same transmission power. The time consumed for uploading data is $t_m^{tr} = f_m^{se} t_m^{se} / r_m$. To ensure the sensing task execution satisfies the delay constraint, we have $t_m^{se} + t_m^{tr} \le T$. Each vehicular node is selfish and competing for a reward. The profit of m is the reward gained from the SP. As similar as that in [22-23], the earned reward is proportional to the contributed data size and denoted as $R\left(f_m^{se} t_m^{se} / \sum_{m \in M} f_m^{se} t_m^{se} \right)$, where R is the reward to all the vehicular nodes. The cost of m for sensing is $c_m^{se} f_m^{se} t_m^{se}$, where c_m^{se} is the sensing cost of unit data. The cost of m for transmission is $c^{tr} f_m^{se} t_m^{se} / r_m$, where c^{tr} is the communication cost of unit time for occupying the bandwidth. Therefore, the utility of m is

$$U_m\left(f_m^{se} \right) = \frac{f_m^{se} t_m^{se}}{\sum_{m \in M} f_m^{se} t_m^{se}} R - \left(c_m^{se} + \frac{c^{tr}}{r_m} \right) f_m^{se} t_m^{se}. \tag{4}$$

We consider a set of sensor nodes $N \subset S$ recruited to perform sensing tasks, where $M \cup N = S$ and $M \cap N = \varnothing$. The sensor nodes are assumed to set the uniform sensing frequency f^{se}, which ensures the collected data has time diversity and not redundant. Thus, the sensing strategy of sensor node $n \in N$ is its sensing time consumption t_n^{se}, and the collected data size is $f^{se} t_n^{se}$. After taking all the photos, n uploads the collected image data to a nearby AP. As similar as that of vehicular nodes, the channel transmission rate of n is $r_n = B \log_2 \left(1 + P_0 h_n \left(d_n \right)^{-r} / w_0 \right)$, where h_n is the channel gain, and d_n is the distance between n and the AP. The time consumed for uploading data is $t_n^{tr} = f^{se} t_n^{se} / r_n$. To ensure that the sensor node can finish the sensing task within the tolerant delay, we have $t_n^{se} + t_n^{tr} \le T$. When multiple vehicular nodes and sensor nodes upload the sensing data simultaneously, they use different spectrum bands via OFDMA technology, same as IEEE standard 802.11p/D3.0 for vehicular networks [24-25]. There is no mutual interference among different spectrum bands. The profit of n is the reward attained from the SP. Different from the vehicular nodes, the sensor nodes are administrated by the government and have no competition for the reward. The reward received by n is denoted as $p_n f_n^{se} t_m^{se}$, where p_n is the price offered by the SP for unit data. The cost of n for sensing is $c_n^{se} f^{se} t_n^{se}$ and the cost for transmission is $c^{tr} f_n^{se} t_n^{se} / r_n$. Each sensor node has its own working duties, such as traffic monitoring. During the period of occupation by the SP for sensing, the sensor nodes have the risk of disturbing their own work and causing economic loss. The risk is denoted as $w_n \left(t_n^{se} \right)^2$, where w_n is a risk factor. Therefore, the utility of n is

$$U_n\left(t_n^{se} \right) = \left(p_n - c_n^{se} - \frac{c^{tr}}{r_n} \right) f^{se} t_n^{se} - w_n \left(t_n^{se} \right)^2. \tag{5}$$

The profit of the SP is its expected sensing satisfaction at all the nodes. The sensing qualities of vehicular nodes and sensor nodes are different based on their spatial coverage and motion blur. Therefore, the sensing satisfaction function in Eqn. (2) can be rewritten as

$$\begin{aligned} Q = &\delta \sum_{m \in M} q_m f_m^{se} t_m^{se} \left(T - t_m^{se} - \frac{f_m^{se} t_m^{se}}{r_m} \right) \\ &+ \delta \sum_{n \in N} q_n f^{se} t_n^{se} \left(T - t_n^{se} - \frac{f_n^{se} t_n^{se}}{r_n} \right), \end{aligned} \tag{6}$$

where q_m and q_n are the sensing quality of each vehicular node and each static node, respectively. The cost of the SP is the paid reward. Therefore, the utility of the SP is

$$U_0(R,\mathbf{p}) = Q - R - \sum_{n\in N} p_n f^{se} t_n^{se}, \quad (7)$$

where $\mathbf{p} = \{p_1, \cdots, p_N\}$ with entry p_n denotes the price that the SP offers to n.

4.3 Game Formulation

A Stackelberg game is a strategic game which consists of a leader and multiple followers competing with each other [22-23]. The leader chooses its strategy first, and the followers decide their strategies subsequently [26]. Here the problem is how the SP assigns the sensing tasks and the rewards to maximize its sensing satisfaction. We formulate the problem as a three-party Stackelberg game, where the SP is the leader while the vehicular and the sensor nodes are two groups of followers.

Under the Stackelberg game formulation, the sensing strategy of m is f_m^{se}, which depends on R. Each vehicular node has to determine its optimal f_m^{se*} given R and the sensing strategies of the other vehicular nodes. Mathematically, the problem can be written as **Problem 1:**

$$\begin{aligned} &\max_{f_m^{se}} U_m\left(f_m^{se}\right) \\ &\text{s.t.C1:} 0 \le f_m^{se} \le f_m^{\max} = \min\left\{\beta v_m, r_m\left(\frac{T}{t_m^{se}} - 1\right)\right\} \end{aligned} \quad (8)$$

where $f_m^{\max}$ in C1 is the maximum sensing frequency, which ensures the spatial-temporal diversity of collected data and completing task within the tolerant latency.

In the proposed game, the sensing strategy of n is t_n^{se}, which depends on p_n. Each sensor node has to decide its optimal t_n^{se*} given p_n. Mathematically, the problem can be written as **Problem 2:**

$$\begin{aligned} &\max_{t_n^{se}} U_n\left(t_n^{se}\right) \\ &\text{s.t.C2:} 0 \le t_n^{se} \le t_n^{\max} = \frac{r_n T}{f^{se} + r_n}, \end{aligned} \quad (9)$$

where $t_n^{\max}$ in C2 is the maximum sensing time ensuring the task can be accomplished within the tolerant latency.

Clearly, the SP can command the total collected data size by controlling R and p_n. However, setting high reward and price also increases the operating cost of the SP. Therefore, the SP needs to find the optimal R^* and $\mathbf{p}^*$ to maximize its utility. Mathematically, the problem can be written as **Problem 3:**

$$\begin{aligned} &\max_{R,\mathbf{p}} U_0(R,\mathbf{p}) \\ &\text{s.t.C3:} \sum_{m\in M} f_m^{se} t_m^{se} + \sum_{n\in N} f^{se} t_n^{se} = D. \end{aligned} \quad (10)$$

C3 ensures the assigned tasks collect enough data.

Problem 1, Problem2, and Problem 3 together form the three-party Stackelberg game. The objective of this game is to find a Stackelberg Equilibrium (SE) point from which the SP (leader), the vehicular nodes (followers), and the sensor nodes (followers) have no motivation to deviate.

Definition 3 (Stackelberg Equilibrium): Let f_m^{se*} be the solution for Problem 1, t_n^{se*} be the solution for Problem 2, and $(R^*, \mathbf{p}^*)$ be the solution for Problem 3. The point $(R^*, \mathbf{p}^*, \mathbf{f}^{\mathbf{se}*}, \mathbf{t}^{\mathbf{se}*})$ is an SE for the proposed three-party Stackelberg game if, for any $(R, \mathbf{p}, \mathbf{f}^{\mathbf{se}}, \mathbf{t}^{\mathbf{se}})$, the following conditions are satisfied:

$$\begin{aligned} &U_0\left(R^*, \mathbf{p}^*, \mathbf{f}^{\mathbf{se}*}, \mathbf{t}^{\mathbf{se}*}\right) \ge U_0\left(R, \mathbf{p}, \mathbf{f}^{\mathbf{se}*}, \mathbf{t}^{\mathbf{se}*}\right), \\ &U_m\left(f_m^{se*}, R^*\right) \ge U_m\left(f_m^{se}, R^*\right), \forall m, \\ &U_n\left(t_n^{se*}, p_n^*\right) \ge U_n\left(t_n^{se}, p_n^*\right), \forall n, \end{aligned} \quad (11)$$

where $\mathbf{f}^{\mathbf{se}}$ with entry f_m^{se} denotes the sensing frequency that vehicular node m chooses, and $\mathbf{t}^{\mathbf{se}}$ with entry t_n^{se} denotes the sensing time that sensor node n determines.

The vehicular nodes compete with each other. Thus, a non-cooperative subgame is formulated among the vehicular nodes. There may exist a Nash Equilibrium (NE) point where anyone cannot enhance its utility by changing its strategy unilaterally.

Definition 4 (Nash Equilibrium): Let $\left(f_m^{se*}, \mathbf{f}_{-m}^{se*}\right)$ be the solution for Problem 1. The point $\left(f_m^{se*}, \mathbf{f}_{-m}^{se*}\right)$ is a NE for the non-cooperative subgame if, for any $\left(f_m^{se}, \mathbf{f}_{-m}^{se*}\right)$, the following conditions are satisfied:

$$U_m\left(f_m^{se*}, \mathbf{f}_{-m}^{se*}\right) \ge U_m\left(f_m^{se}, \mathbf{f}_{-m}^{se*}\right), \forall m. \quad (12)$$

5 Game Analysis

In this section, the backward induction method is used to analyze the game and find the NE and the SE.

5.1 Subgame Nash Equilibrium

We analyze the existence and uniqueness of the NE.

Theorem 1: A Nash equilibrium exists in the subgame among vehicular nodes.

Proof: The strategy space of each vehicular node is defined to be $\left[0, f_m^{\max}\right]$, which is a non-empty, convex, compact subset of the Euclidean space. From Eqn. (4), U_m is continuous in $\left[0, f_m^{\max}\right]$. We take the first and second derivatives of U_m with respect to f_m^{se} and get

$$\begin{aligned} \frac{\partial U_m}{\partial f_m^{se}} &= \frac{R t_m^{se} \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se}}{\left(\sum_{m \in M} f_m^{se} t_m^{se}\right)^2} - \left(c_m^{se} + \frac{c^{tr}}{r_m}\right) t_m^{se}, \\ \frac{\partial^2 U_m}{\partial f_m^{se2}} &= \frac{-2R\left(t_m^{se}\right)^2 \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se}}{\left(\sum_{m \in M} f_m^{se} t_m^{se}\right)^3} < 0. \end{aligned} \tag{13}$$

U_m is strictly concave with respect to f_m^{se}. Therefore, the Nash equilibrium exists. The proof is now completed.

Lemma 1: The best response function of vehicular node m is

$$f_m^{se*} = \begin{cases} 0, R < c_m a_m \\ \sqrt{\dfrac{R a_m}{a_m t_m^{se2}}} - \dfrac{a_m}{t_m^{se}}, c_m a_m \le R < \dfrac{c_m\left(a_m + f_m^{\max} t_m^{se}\right)^2}{a_m} \\ f_m^{\max}, R \ge \dfrac{c_m\left(a_m + f_m^{\max} t_m^{se}\right)^2}{a_m} \end{cases} \tag{14}$$

where $a_m = \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se}$ and $c_m = c_m^{se} + c^{tr} / r_m$.

Proof: Eqn. (13) implies that the first derivative of U_m strictly decreases on f_m^{se}. Let

$$\begin{aligned} \lim_{f_m^{se} \to 0} \frac{\partial U_m}{\partial f_m^{se}} &= \frac{R t_m^{se}}{a_m} - c_m t_m^{se}, \\ \lim_{f_m^{se} \to f_m^{\max}} \frac{\partial U_m}{\partial f_m^{se}} &= \frac{R t_m^{se} a_m}{\left(a_m + f_m^{\max} t_m^{se}\right)^2} - c_m t_m^{se}. \end{aligned} \tag{15}$$

We consider the best response strategy f_m^{se*} to maximize U_m in three cases. In case 1 that $R_m \le c_m a_m$, we have $\lim_{f_m^{se} \to 0} \frac{\partial U_m}{\partial f_m^{se}} < 0.$ In means that U_m is a decreasing function with respect to f_m^{se}, and the best response strategy is $f_m^{se*} = 0$. In case 2 that $R \ge \frac{c_m\left(a_m + f_m^{\max} t_m^{se}\right)^2}{a_m}$, we have $\lim_{f_m^{se} \to f_m^{\max}} \frac{\partial U_m}{\partial f_m^{se}} \ge 0.$ In means that U_m is an increasing function with respect to f_m^{se}, and the best response strategy is $f_m^{se*} = f_m^{\max}$. In case 3 that $c_m a_m \le R < \frac{c_m\left(a_m + f_m^{\max} t_m^{se}\right)^2}{a_m}$, we have $\lim_{f_m^{se} \to 0} \frac{\partial U_m}{\partial f_m^{se}} \ge 0$ and $\lim_{f_m^{se} \to f_m^{\max}} \frac{\partial U_m}{\partial f_m^{se}} < 0$. It means that U_m firstly increases and then decreases with respect to f_m^{se}. Thus U_m is a strictly concave function with respect to f_m^{se}. By solving $\frac{\partial U_m}{\partial f_m^{se}} = 0,$ we obtain $f_m^{se*} \sqrt{\frac{R a_m}{a_m t_m^{se2}}} - \frac{a_m}{t_m^{se}}$. The proof is now completed.

Theorem 2: The uniqueness of the Nash equilibrium in the non-cooperative subgame is guaranteed if the following condition

$$\sum_{m \in M} c_m > 2 c_m \left(|M| - 1\right) \tag{16}$$

is satisfied. $|M|$ is the number of vehicular nodes.

Proof: From Theorem 1, we know that there exists a NE in the subgame. Given R and sensing strategies $\mathbf{f}_{-m}^{se*}$ of other vehicular nodes, the best response function of m is defined as $f_m^{se*} = \mathrm{B}\left(\mathbf{f}_{-m}^{se}, R\right)$, which is given in Eqn. (14). The uniqueness of the NE can be proved by showing that the best response function is the standard function, which needs to satisfy the following conditions [26].

Positivity: $\mathrm{B}\left(\mathbf{f}_{-m}^{se}, R\right) > 0$;

Monotonicity: For all $\mathbf{f}_{-m}^{se}$ and $\mathbf{f}_{-m}^{se\prime}$, if $\mathbf{f}_{-m}^{se} \ge \mathbf{f}_{-m}^{se\prime}$, then $\mathrm{B}\left(\mathbf{f}_{-m}^{se}, R\right) \ge \mathrm{B}\left(\mathbf{f}_{-m}^{se\prime}, R\right)$;

Scalability: For all $\mu > 1$, $\mu \mathrm{B}\left(\mathbf{f}_{-m}^{se}, R\right) \ge \mathrm{B}\left(\mu \mathbf{f}_{-m}^{se}, R\right)$.

Firstly, for the positivity, under the condition in (16), we can get (from lemma 1) $\sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se} < \frac{R}{4c_m} < \frac{R}{c_m}$ and conclude that $\sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se} < \sqrt{\frac{R}{c_m} \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se}}$. Thus, we have

$$\frac{1}{t_m^{se}} \left(\sqrt{\frac{R}{c_m} \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se}} - \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se} \right) > 0, \tag{17}$$

which satisfies the positivity condition. Secondly, taking the first derivative of $\mathrm{B}\left(\mathbf{f}_{-m}^{se}, R\right)$ with respect to f_j^{se}, $j \in M \backslash \{m\}$, we have

$$\frac{\partial \mathrm{B}\left(\mathbf{f}_{-m}^{se}, R\right)}{\partial f_j^{se}} = \frac{t_j^{se}}{t_m^{se}} \left(\frac{1}{2} \sqrt{\frac{R}{c_m \sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se}}} - 1 \right). \tag{18}$$

Under the condition $\sum_{j \in M \backslash \{m\}} f_j^{se} t_j^{se} < \frac{R}{4c_m},$ we have

$\frac{1}{2}\sqrt{\frac{R}{c_m}\sum_{j\in M\setminus\{m\}} f_j^{se}t_j^{se}}-1>0$. Thus, when $\mathbf{f}_{-m}^{se}\geq\mathbf{f}_{-m}^{se\prime}$, we have $\mathrm{B}\left(\mathbf{f}_{-m}^{se},R\right)\geq\mathrm{B}\left(\mathbf{f}_{-m}^{se\prime},R\right)$. The monotonicity is proved. Finally, for $\mu>1$, we have

$$\mu\mathrm{B}\left(\mathbf{f}_{-m}^{se},R\right)-\mathrm{B}\left(\mu\mathbf{f}_{-m}^{se},R\right)=\frac{\mu-\sqrt{\mu}}{t_m^{se}}\sqrt{\frac{R\sum_{j\in M\setminus\{m\}} f_j^{se}t_j^{se}}{c_m}}. \quad \textbf{(19)}$$

Thus, $\mu\mathrm{B}\left(\mathbf{f}_{-m}^{se},R\right)\geq\mathrm{B}\left(\mu\,\mathbf{f}_{-m}^{se},R\right)$ is always satisfied for $\mu>1$. The scalability is proved. The best response function meets the three conditions and is a standard function. The NE is unique. The proof is now completed.

Theorem 3: For the non-cooperative subgame among the vehicular nodes, the unique Nash equilibrium for vehicular node *m* has a closed-form expression give by

$$f_m^{se*}=\frac{R\left(|M|-1\right)}{t_m^{se}\sum_{m\in M}c_m}\left(1-c_m\frac{\left(|M|-1\right)}{\sum_{m\in M}c_m}\right), \quad \textbf{(20)}$$

when the condition in (16) holds.

Proof: Based on Eqn. (14), we have

$$\sum_{j\in M\setminus\{m\}} f_j^{se}t_j^{se}=\frac{c_m}{R}\left(\sum_{m\in M} f_m^{se}t_m^{se}\right)^2. \quad \textbf{(21)}$$

Then we calculate the summation of this expression for all vehicular nodes and get

$$\sum_{m\in M} f_m^{se}t_m^{se}=\frac{R\left(|M|-1\right)}{\sum_{m\in M}c_m}. \quad \textbf{(22)}$$

Substituting Eqn. (22) into Eqn. (21), we have

$$\frac{R\left(|M|-1\right)}{\sum_{m\in M}c_m}-f_m^{se}t_m^{se}=\frac{c_m}{R}\left(\frac{R\left(|M|-1\right)}{\sum_{m\in M}c_m}\right)^2, \quad \textbf{(23)}$$

which can be rewritten as Eqn. (20). The proof is now completed.

Lemma 2: Under the condition in Eqn. (16), the condition

$$\sum_{j\in M\setminus\{m\}} f_j^{se}t_j^{se}<\frac{R}{4c_m} \quad \textbf{(24)}$$

is satisfied.

Proof: Based on Eqn. (21) and Eqn. (22), we have

$$\sum_{j\in M\setminus\{m\}} f_j^{se}t_j^{se}=Rc_m\left(\frac{|M|-1}{\sum_{m\in M}c_m}\right)^2. \quad \textbf{(25)}$$

Based on (16), we have $\frac{|M|-1}{\sum_{m\in M}c_m}<\frac{1}{2c_m}$, and thus conclude that $\sum_{j\in M\setminus\{m\}} f_j^{se}t_j^{se}<\frac{R}{4c_m}$. It means that if (16) holds, the condition in (24) is satisfied. The proof is now completed.

Generally, we can get the NE by utilizing the best response dynamics [26]. Problem 1 is solved and we analyze the SE in the following.

5.2 Stackelberg Equilibrium

Firstly, we analyze the best response of a sensor node. The strategy space of each sensor node is defined to be $\left[0,t_n^{\max}\right]$, which is a non-empty, convex, compact subset of the Euclidean space. From Eqn. (5), U_n is continuous in $\left[0,t_n^{\max}\right]$. Let $c_n=c_m^{se}+c^{tr}/r_n$, we calculate the first and second derivatives of U_n with respect to t_n^{se} and get

$$\begin{aligned}&\frac{\partial U_n}{\partial t_n^{se}}=p_n-c_n-2w_nt_n^{se},\\&\frac{\partial^2 U_n}{\partial t_n^{se2}}=-2w_n<0.\end{aligned} \quad \textbf{(26)}$$

U_n is strictly concave. By using $\frac{\partial U_n}{\partial t_n^{se}}=0$, we obtain

$$t_n^{se*}=\begin{cases}0, p_n<c_n\\ \frac{p_n-c_n}{2w_n}, c_n\leq p_n<2w_nt_n^{\max}+c_n\\ t_n^{\max}, p_n\geq 2w_nt_n^{\max}+c_n\end{cases} \quad \textbf{(27)}$$

Given p_n, Eqn. (27) is the solution for Problem 2.

We now study the best strategy of the SP. By substituting f_m^{se*} and t_n^{se*} into the objective function of Problem 3, which leads to **Problem 3a:**

$$\begin{aligned}&\max_{R,\mathbf{p}} U_0=aR-bR^2+\sum_{n\in N}e_np_n-\sum_{n\in N}g_np_n^2-z\\&\text{s.t.C4:}R\sum_{m\in M}h\left(1-c_mh\right)+\sum_{n\in N}\frac{f^{se}\left(p_n-c_n\right)}{2w_n}=D\\&\text{C5:}0\leq R\leq R^{\max}=\arg\max_m\frac{c_m\left(\sum_{m\in M}f_m^{\max}t_m^{se}\right)^2}{\sum_{j\in M\setminus\{m\}}f_j^{\max}t_j^{se}}\\&\text{C6:}c_n\leq p_n\leq p_n^{\max}=2w_nt_n^{\max}+c_n\end{aligned} \quad \textbf{(28)}$$

where $h=\frac{|M|-1}{\sum_{m\in M} c_m}$, $a=\sum_{m\in M}\delta q_m h\left(T-t_m^{se}\right)\left(1-c_m h\right)-1$, $b=\sum_{m\in M}\frac{\delta q_m h^2}{r_m}\left(1-c_m h\right)^2$, $e_n=\frac{q_n f^{se}}{2w_n}\left(\delta T+\frac{\delta c_n u}{w_n}+\frac{c_n}{q_n}\right)$, $g_n=\frac{f^{se}}{2w_n}\left(\frac{\delta g_n u}{2w_n}+1\right)$, $z=\sum_{n\in N}\frac{\delta q_n f^{se} c_n}{2w_n}\left(T+\frac{c_n u}{2w_n}\right)$, $u=1+f^{se}/r_n$. C4 ensures that the assigned sensing tasks collect enough amount of data. C5 and C6 limit the ranges of R and p_n, respectively. The proof is as follows. Based on Eqn. (27), optimal price p_n^* is bounded by c_n and $p_n^{\max}$. This is because that when $p_n=c_n$, sensor node n will not carry out the sensing task. When $p_n=p_n^{\max}$, n will sense data at maximum sensing time. Offering a price less than c_n or larger than $p_n^{\max}$ will not increase the SP's profit, but will enhance the SP's cost. Based on Eqn. (14), $f_m^{se*}=f_m^{\max}$ when $R\geq\frac{c_m\left(a_m+f_m^{\max}t_m^{se}\right)^2}{a_m}$. If $R=R^{\max}$, each vehicular node will sense data at its maximum sensing frequency. Offering a reward larger than $R^{\max}$ will not enhance the SP's profit, but will increase the SP's cost. Then we analyze the best response of the SP.

Lemma 3: The optimized strategies R^* and $\mathbf{p}^*$ of the SP is unique.

Proof: The strategy spaces of the SP for R and p_n are defined to be $\left[0,R^{\max}\right]$ and $\left[c_n,p_n^{\max}\right]$, respectively. Each strategy space is a non-empty, convex, compact subset of the Euclidean space. From Eqn. (28), U_0 is continuous with respect to R and p_n in the strategy spaces. We take the second derivatives of Eqn. (28) with respect to R and p_n respectively, and get $\frac{\partial U_0}{\partial R}=-2b<0$, $\frac{\partial U_0}{\partial p_n}=-2g_n<0$. We can observe that all the constraints are linear. So, Problem 3a is a convex problem with multiple variables. The SP has the unique optimized strategies R^* and $\mathbf{p}^*$, which can be directly found out by using the existing typical convex optimal algorithms (e.g., dual decomposition algorithm [27]).

Theorem 4: There exists a unique Stackelberg equilibrium in the proposed three-party game.

Proof: With given reward R, each vehicular node acts as a follower and always chooses a unique best response f_m^{se*} to reach the NE among all the vehicular nodes. Given price p_n, each sensor node acts as a follower and always has its unique best response t_n^{se*} due to the concave character of the utility function U_n. No matter what strategy the SP chooses, each vehicular node, as well as each sensor node, always has its unique best response. Based on Lemma 3, given the strategy chosen by each sensing node, the utility function of the SP is strictly concave with respect to R and p_n, respectively. Hence, the SP would be able to find a unique R^* and a unique $\mathbf{p}^*$ to maximize its utility. The condition (11) is satisfied. Therefore, there exists a unique SE. The proof is now completed.

We note that it is able to solve the task assignment problem in a centralized fashion if the SP has global information, such as c_m^{se}, c_n^{se}, and w_n. However, in order to protect the privacy of each vehicular node and each sensor node, we design a distributed algorithm, where the optimization can be performed by the three parties without the need for any private information. In the proposed algorithm, wireless communication is adopted between nodes and APs, and APs communicate with each other by wire connection. In each round, the SP offers reward and prices to the nodes via APs. Each sensor node chooses its strategy according to the offer price. Each vehicular node decides its strategy based on the offered reward and the strategies from other vehicular nodes. The process iterates until enough sensing data would be collected.

Algorithm 1 Distributed Algorithm to reach the SE

Input: The sensing quality of the vehicular nodes, $\{q_m\}$; The residence time of the vehicular nodes, $\{t_m^{se}\}$ The sensing quality of the sensor nodes, $\{q_n\}$; The sensing frequency of the sensor nodes, f^{se}; The required data size D; The preset step, $\{u_0,u_1,\cdots,u_N\}$;

Output: Optimal reward, R^*; Optimal prices, $\mathbf{p}^*$;

1: Initialization R, $\mathbf{p}$;
2: **repeat**
3: **for** each vehicular node $m\in M$ **do**
4: Vehicular node m adjusts its sensing strategy f_m based on Eqn. (14);
5: **end for**
6: **for** each sensor node $n\in N$ **do**
7: Sensor node n adjusts its sensing strategy t_n based on Eqn. (26)
8: **end for**
9: $R=R+u_0R$
10: $p_n=p_n+u_np_n$
11: **until** $\sum_{m\in M} f_m^{se}t_m^{se}+\sum_{n\in N} f^{se}t_n^{se}>=D$
12: $R^*=R$ and $\mathbf{p}^*=\mathbf{p}$;

6 Resource Allocation for Inference

In this section, we study the resource allocation problem. The problem focuses on how the SP allocates the computation resources of the edge server to minimize the delay of getting inference results. The problem is nonconvex and nonsmooth, and we propose an iteration algorithm to find a locally optimized solution.

6.1 Problem Formulation

When carrying out the sensing task, each node has its sensing data size $x_s = f_s^{se} t_s^{se}$ and time consumption $t_s = t_s^{se} + t_s^{tr}$. When receiving the uploaded data, the SP utilizes the computation resource of the edge server for executing inference. We define the inference based on the data uploaded by node s as inference task s. The inference tasks are executed with a parallel computing manner. In order to shorten the latency of inference execution, the SP needs to allocate the computation resources to the inference tasks optimally.

Following [28], we model the computation recourse, i.e., the number of cycles, needed for inference task s as αx_s, where $\alpha > 0$ is the factor depends on computation complexity of the deep learning model for inference. The computation recourse allocated to inference task s is denoted as f_s^{in}, which implies CPU's computational speed. Thus, the time consumed for inference execution is $\alpha x_s / f_s^{in}$. The total time consumed to carry out sensing task s and corresponding inference task is $t_s + \alpha x_s / f_s^{in}$. Since the inference tasks are executed with a parallel computing manner, the time consumed to obtain all the inference results is $\max\left\{t_s + \alpha x_s / f_s^{in}\right\}$, $\forall s \in S$. Following [28], we model the power consumption of CPU as $\varepsilon\left(f_s^{in}\right)^3$, where ε is a coefficient which depends on chip architecture. As f_s^{in} is cycles per second, the energy consumption per cycle is $\varepsilon\left(f_s^{in}\right)^2$. Thus, the energy consumption for inference task s is $\varepsilon\alpha x_s\left(f_s^{in}\right)^2$. The problem focuses on how the SP allocates the computation resources of the edge server to minimize the delay of obtaining all the inference results. Meanwhile, the stringent requirement for energy consumption needs to be satisfied. Mathematically, the problem can be written as **Problem 4:**

$$\begin{aligned} &\min_{\mathbf{f}^{in}} \max_{s} \left\{ t_s + \frac{\alpha x_s}{f_s^{in}} \right\} \\ &\text{s.t.C7:} \sum_{s \in S} f_s^{in} \le F \\ &\quad \text{C8:} \sum_{s \in S} \varepsilon \alpha x_s \left(f_s^{in}\right)^2 \le E_{\max} \\ &\quad \text{C9:} 0 < f_s^{in} < F \end{aligned} \tag{29}$$

C7 constrains the total amount of the computation resources in the edge server. C8 reflects the energy constraint. C9 specifies the domain of computation resource allocated to inference task s. Due to the nonconvexity and nonsmooth of objective function, P4 is a nonconvex and nonsmooth problem.

6.2 Locally Optimal Solution

Lemma 4: In the optimal solution $\mathbf{f}^{in*}$ for Problem 4, the following condition

$$t_1 + \frac{\alpha x_1}{f_1^{in*}} = t_2 + \frac{\alpha x_2}{f_2^{in*}} = \cdots = t_S + \frac{\alpha x_S}{f_S^{in*}} \tag{30}$$

is satisfied, which means that all the inference results are obtained at the same delay.

Proof: Suppose that the condition is not satisfied, which means that for the optimal solution $\mathbf{f}^{in*}$, there exists at least one inference task s finished at the maximum delay. To minimize the maximum delay, the SP has to allocate more computation resource f_s^{in} to the inference task, which violates the given optimal solution f_s^{in*}. The proof is now completed.

We propose a suboptimal algorithm to find a locally optimal solution without learning process. Based on Lemma 4, we introduce a new variable λ to rewrite the original Problem 4 as **Problem 4a:**

$$\begin{aligned} &\min_{\mathbf{f}^{in}} \lambda \\ &\text{s.t.C7,C8,C9} \\ &\quad \text{C10:} t_s + \frac{\alpha x_s}{f_s^{in}} = \lambda \end{aligned} \tag{31}$$

Given any λ, there exists a unique $\mathbf{f}^{in}$ to satisfy C10. If $\mathbf{f}^{in}$ satisfies C7, C8, C9, Problem 4a is feasible for $\left(\lambda, \mathbf{f}^{in}\right)$. Therefore, we could solve Problem 4a by Algorithm 2. Specifically, we first give a feasible solution $\left(\lambda, \mathbf{f}^{in}\right)$, then we decrease λ with iteration. This process is continued until the solution does not satisfy the constraints. Algorithm 2 yields a non-increasing objective, which is bounded by a value larger than zero and converges to the stationary point.

Algorithm 2 Iteration Algorithm

Input: The set of time consumed for sensing tasks, $\mathbf{t} = \{\mathbf{t_1}, \cdots, \mathbf{t_S}\}$; The set of uploaded data, $\mathbf{x} = \{\mathbf{x_1}, \cdots, \mathbf{x_S}\}$; The preset step, η; The computation recourse limitation F and the energy consumption limitation $E_{\max}$;

Output: Optimal time consumption, λ^*; Optimal computation resource allocation, $\mathbf{f^{in*}}$;

1: Initialize with a feasible λ;
2: Calculate the allocated computation resource $f_s^{in} = \frac{\alpha x_s}{\lambda - t_s}$;
3: **while** $\sum_{s \in S} f_s^{in} \le F$ and $\sum_{s \in S} \varepsilon \alpha x_s \left(f_s^{in}\right)^2 \le E_{\max}$ **do**
4: $\quad\lambda = \lambda - \eta$;
5: $\quad f_s^{in} = \frac{\alpha x_s}{\lambda - t_s}$;
6: **end while**
7: $\lambda^* = \lambda$ and $f_s^{in*} = f_s^{in}$;

7 Numerical Results

We evaluate the performance of the proposed task assignment scheme and resource allocation scheme by extensive simulations. Vehicular nodes' behavior is formulated on a real dataset from San Francisco Yellow Cabs [29], which includes moving traces of 536 urban taxis over one month. We randomly take a metropolitan area of 5KM$\times$5KM in the map of the dataset for observation. We randomly distribute 10 target objects near the roads, and each has 4 effective angles. Sensor nodes are deployed along the road based on two types of distributions: even distribution and uneven distribution. We set that the cameras of all nodes have the same field-of-view $\varphi_s = 45^\circ$ and effective range $d_s = 50m$ [19]. We consider that the collected data is used for image classification, which can support the vehicular services. Taking parking navigation service as example, the image data is used for identifying whether a parking space is occupied. We consider the model similar with the face recognition model in [30], where the data size is 420KB and the total number of CPU cycles is 1000Megacycles. Thus, $\alpha = 2$ mega cycle/KB . The computation resource of the edge server is F=16GHz. Similar setting is made in [31]. The other parameters are set according to Table 1, which is similar to the parameters set in [24, 28].

Table 1. Parametet setting in the simulation

Parameter	Setting
A vehicular node: v_m, t_m, q_m	[11.1, 16.7]m/s, [120, 240]s, [10, 13]
A sensor node: f_n, q_n, w_n	0.1MB/s, [4, 6], [0.001, 0.005]
Communication parameters B, P, h, N_0, r	20MHz, 1W, 10^{-3}W, 10^{-9}W, 2.5
Cost parameters Cse, Ctr	0.1RMB/MB, 0.1RMB/s
The edge server F, α, E_{max}, ε	16 GHz, 2 mega cycle/KB, 1J, 10^{-26}

7.1 Performance Evaluation of Task Assignment

Figure 2 shows the utility of the SP U_0 with respect to reward R and average price $\overline{p}$. Without the constraint of required data size D , U_0 is a convex function with respect to R and $\overline{p}$. With the constraint $D = 100\text{M}$, R and $\overline{p}$ are linear relations, which is the plane in the figure. The unique SE is on the tangent line of the plane and the surface. No matter what D is, there exists a unique SE.

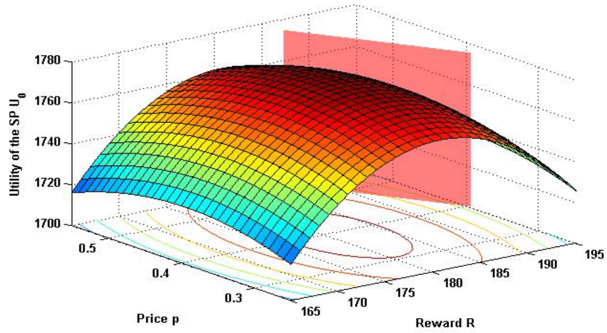

Figure 2. U_0 with respect to R and $\overline{p}$

Considering the case of parking navigation service, an image of unoccupied parking space is generally valid for no more than 10 minutes. Here we set the tolerant delay $T = 480s$, the number of vehicular nodes $M = 20$, the number of sensor nodes $N = 10$. We compare the performance of our proposed task assignment scheme used for hybrid sensor network (HSN), pure VCN, and pure WSN. We also compare the performance of the centralized scheme and the distributed scheme. With the centralized scheme, the SP is available to know the private information of each vehicular node and each sensor node, such as c_m^{se} , c_n^{se} , w_n . As shown in Figure 3, the utility of the SP by using the distributed scheme reduces only 0.02% than that by using the centralized scheme. The SP gets higher utility by using HSN than by using pure VCN and pure WSN. It is because that HSN provides the improved sensing quality and reduced delay of obtaining the data. The SP achieves growing utility with respect to increasing D when using HSN or pure VCN. But when the SP utilizes pure WSN, the SP's utility firstly increases then decreases with respect to growing D . It is because the sensor nodes consume more time for collecting the required data and thus diminish the sensing satisfaction.

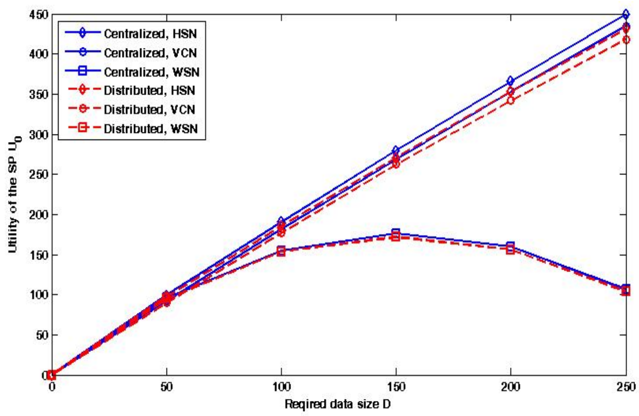

Figure 3. U_0 comparison using different task assignment schemes

Figure 4 shows with more required data, the SP prefers to obtain more data from the vehicular nodes. It is because the vehicular nodes increase the sensing frequency to get more data. Meanwhile, the sensor nodes need to increase sensing time to obtain more data, which also increases the delay. The SP has higher sensing satisfaction at the vehicular nodes than that at the sensor nodes. With the higher tolerant delay T, the SP also chooses to get more data from the vehicular nodes. It is because that the vehicular nodes can save more time for finishing the assigned sensing tasks and gain higher sensing satisfaction.

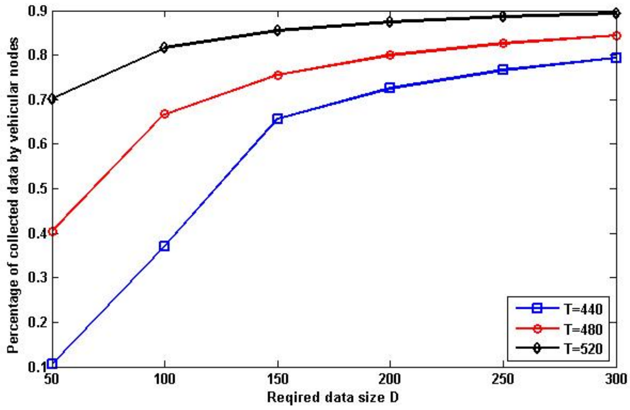

Figure 4. Percentage of collected data by vehiular node with respect to D

Figure 5 indicates that the utility of the SP increases when the number of vehicular nodes M increases. It is because that the growing number of vehicular nodes promote their internal competition and also enhance their sensing quality. The sensing satisfaction increases accordingly. Similarly, the increasing number of sensor nodes N lead to the enhancement of SP's utility. The SP prefers to obtain more sensing data from vehicular nodes when the distribution of sensor nodes is uneven. It is because that the sensor nodes of even distribution have higher coverage ability than that of uneven distribution.

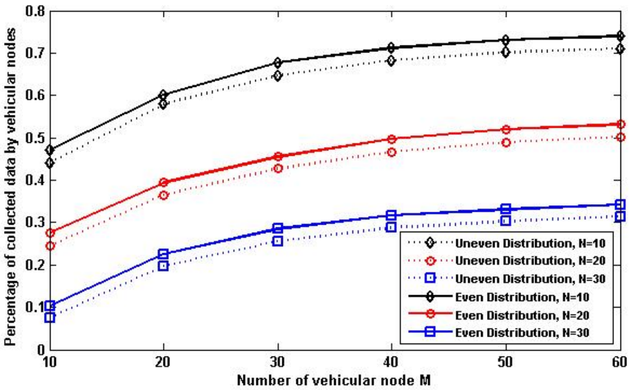

Figure 5. U_0 with respect to M

Figure 6 shows that with more vehicular nodes, the SP prefers to get more data from the vehicular nodes. It is because increasing M enhances the sensing quality of the vehicular nodes. Meanwhile, the competition among the vehicular nodes reduces the incentive cost of the SP. Similarly, the SP chooses to obtain more data from the sensor nodes if more sensor nodes. The SP get less utility when the distribution of sensor nodes is uneven. It is because the sensor nodes of uneven distribution have lower coverage ability than that of even distribution.

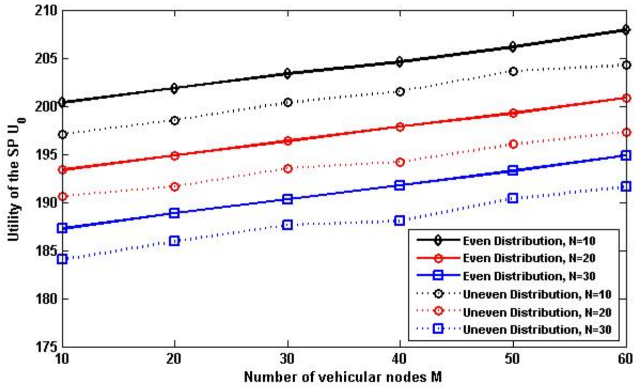

Figure 6. Percentage of collected data by vehiular node with respect to M

7.2 Performance Evaluation of Resource Allocation

We set $T = 480\text{s}$, $D = 100\text{M}$, $M = 20$, $N = 10$. Based on the simulation, the maximum delay of a node performing the assigned sensing task is 305s. We compare the performance of our proposed iteration algorithm and the greedy algorithm for resource allocation. The greedy algorithm allocates more amount of computation resource f_s^{in} to the inference task of maximum delay in each iteration. As shown in Figure 7, the convergence rate of our proposed algorithm is 4.2 times faster than that of the greedy algorithm.

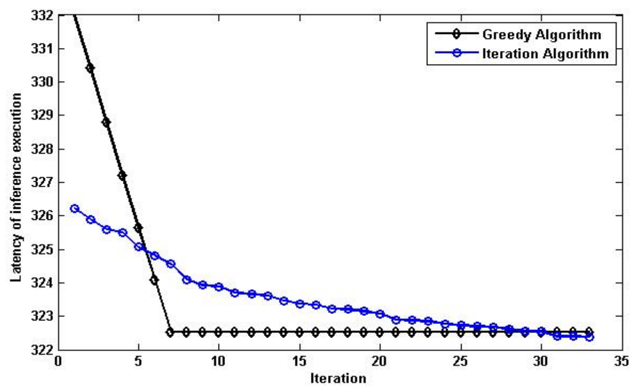

Figure 7. Convergence of the algorithms

Figure 8 shows the delay of getting the inference results with respect to M. With the increasing M, the delay decreases. It is because that with the higher M, the nodes can finish the sensing tasks earlier. The delay increases with the growing required data size D. It is because that when D grows, the nodes need to collect more data and consume more time.

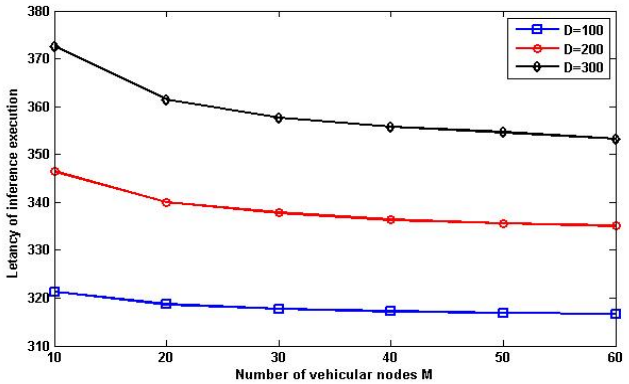

Figure 8. Latency of inference execution with respect to M

8 Conclusion

In this paper, we propose a new system to facilitate AI applications of connected vehicles. The system utilizes hybrid sensor network to increase the sensing quality and exploits edge servers to reduce the service delay. We formulate a three-party game to design the task assignment scheme. In the SE, the SP, the vehicular nodes, and the sensor nodes reach a deal with satisfying pricing strategies and sensing strategies. We also present a resource allocation scheme to help the SP for minimizing the delay of performing AI inference. Numerical results demonstrate that our proposed task assignment scheme is effective for massive data collection and our proposed resource allocation scheme is efficient for fast inference execution.

Acknowledgments

The work is supported in part by program of NSFC under Grant No. 61971148, the Science and Technology Program of Guangdong Province under Grant No. 2015B010129001, Natural Science Foundation of Guangxi Province under Grant No. 2018GXNSFDA281013, Foundation for Science and Technology Project of Guilin City under Grant No. 20190214-3, and Key Science and Technology Project of Guangxi under Grant No. AA18242021.

References

[1] W. Alasmary, H. Sadeghi, S. Valaee, Strategic Sensing in Vehicular Networks Using Known Mobility Information, *IEEE Transactions on Vehicular Technology*, Vol. 67, No. 3, pp. 1932-1945, March, 2017.

[2] H. Qiu, J. Chen, S. Jain, Y. Jiang, M. McCartney, G. Kar, F. Bai, D. K. Grimm, M. Gruteser, R. Govindan, Towards Robust Vehicular Context Sensing, *IEEE Transactions on Vehicular Technology*, Vol. 67, No. 3, pp. 1909-1922, March, 2017.

[3] J. Ni, A. Zhang, X. Lin, X. S. Shen, Security, Privacy, and Fairness in Fog-based Vehicular Crowdsensing, *IEEE Communications Magazine*, Vol. 55, No. 6, pp. 146-152, June, 2017.

[4] S. Abdelhamid, H. S. Hassanein, G. Takahara, Reputation-aware, Trajectory-based Recruitment of Smart Vehicles for Public Sensing, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 5, pp. 1387-1400, May, 2017.

[5] F. Campioni, S. Choudhury, K. Salomaa, S. G. Akl, Improved Recruitment Algorithms for Vehicular Crowdsensing Networks, *IEEE Transactions on Vehicular Technology*, Vol. 68, No. 2, pp. 1198-1207, February, 2018.

[6] X. Wang, W. Wu, and D. Qi, Mobility-aware Participant Recruitment for Vehicle-based Mobile Crowdsensing, *IEEE Transactions on Vehicular Technology*, Vol. 67, No. 5, pp. 4415-4426, May, 2018.

[7] M. Wu, D. Ye, S. Tang, R. Yu, Collaborative Vehicle Sensing in Bus Networks: A Stackelberg Game Approach, *IEEE/CIC International Conference on Communications in China (ICCC)*, Chengdu, China, 2016, pp. 1-6.

[8] L. Xiao, T. Chen, C. Xie, H. Dai, H. V. Poor, Mobile Crowdsensing Games in Vehicular Networks, *IEEE Transactions on Vehicular Technology*, Vol. 67, No. 2, pp. 1535-1545, February, 2018.

[9] J. Liu, J. Wan, Q. Wang, P. Deng, K. Zhou, Y. Qiao, A Survey on Position-based Routing for Vehicular Ad Hoc Networks, *Telecommunication Systems*, Vol. 62, No. 1, pp. 15-30, May, 2016.

[10] J. Liu, J. Wan, B. Zeng, Q. Wang, H. Song, M. Qiu, A Scalable and Quick-response Software Defined Vehicular Network Assisted by Mobile Edge Computing, *IEEE Communications Magazine*, Vol. 55, No. 7, pp. 94-100, July, 2017.

[11] S. Bhandari, H. Zhao, H. Kim, J. Cioffi, An Optimal Cache Resource Allocation in Fog Radio Access Networks, *Journal of Internet Technology*, Vol. No. 7, pp. 2063-2069, December, 2019.

[12] Y. Lai, F. Yang, J. Su, Q. Zhou, T. Wang, L. Zhang, Y. Xu, Fog-based Two-phase Event Monitoring and Data Gathering in Vehicular Sensor Networks, *Sensors*, Vol. 18, No. 1, pp. 82, December, 2017.

[13] S. Basudan, X. Lin, K. Sankaranarayanan, A Privacy-preserving Vehicular Crowdsensing-based Road Surface Condition Monitoring System Using Fog Computing, *IEEE Internet of Things Journal*, Vol. 4, No. 3, pp. 772-782, June, 2017.

[14] Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz, J. Rodriguez, Robust Mobile Crowd Sensing: When Deep Learning Meets Edge Computing, *IEEE Network*, Vol. 32, No. 4, pp. 54-60, August, 2018.

[15] M. Chen, Y. Hao, Task Offloading for Mobile Edge

Computing in Software Defined Ultra-Dense Network, *IEEE Journal on Selected Areas in Communications*, Vol. 36, No. 3, pp. 587-597, March, 2018.

[16] M. Chen, Y. Hao, L. Hu, K. Huang, V. Lau, Green and Mobility-aware Caching in 5G Networks, *IEEE Trans. Wireless Communications*, Vol. 16, No. 12, pp. 8347-8361, December, 2017.

[17] Y. Wang, K. Wang, H. Huang, T. Miyazaki, S. Guo, Traffic and Computation Co-offloading with Reinforcement Learning in Fog Computing for Industrial Applications, *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 2, pp. 976-986, 2018.

[18] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, S. Guo, Green Resource Allocation based on Deep Reinforcement Learning in Content-centric IoT, *IEEE Transactions on Emerging Topics in Computing*, February, 2018.

[19] Y. Wu, Y. Wang, W. Hu, G. Cao, Smartphoto: A Resource-aware Crowdsourcing Approach for Image Sensing with Smartphones, *IEEE Transactions on Mobile Computing*, Vol. 15, No. 5, pp. 1249-1263, May, 2016.

[20] B. Guo, Q. Han, H. Chen, L. Shangguan, Z. Zhou, Z. Yu, The Emergence of Visual Crowdsensing: Challenges and Opportunities, *IEEE Communications Surveys & Tutorials*, Vol. 19, No. 4, pp. 2526-2543, July, 2017.

[21] B. Guo, H. Chen, Z. Yu, X. Xie, S. Huangfu, D. Zhang, Fliermeet: A Mobile Crowdsensing System for Cross-space Public Information Reposting, Tagging, and Sharing, *IEEE Transactions on Mobile Computing*, Vol. 14, No. 10, pp. 2020-2033, October, 2015.

[22] D. Yang, G. Xue, X. Fang, J. Tang, Incentive Mechanisms for Crowdsensing: Crowdsourcing with Smartphones, *IEEE/ACM Transactions on Networking (TON)*, Vol. 24, No. 3, pp. 1732-1744, June, 2016.

[23] X. Kang, S. Sun, J. Yang, *Incentive Mechanisms for Motivating Mobile Data Offloading in Heterogeneous Networks: A Salary-plus-bonus Approach*, https://arxiv.org/abs/1802.02954, 2018.

[24] C. Yang, Y. Liu, X. Chen, W. Zhong, S. Xie, Efficient Mobility-aware Task Offloading for Vehicular Edge Computing Networks, *IEEE Access*, Vol. 7, pp. 26652-26664, February, 2019.

[25] Q. Wang, D. Gao, W. Zhu, Cloud-enabled Soft-ware-defined Vehicular Networks: Architecture, Applications and Challenges, *Journal of Internet Technology*, Vol. 20, No. 6, pp. 1819-1828, November, 2019.

[26] Z. Xiong, S. Feng, D. Niyato, P. Wang, Z. Han, *Edge Computing Resource Management and Pricing for Mobile Blockchain*, https://arxiv.org/abs/1710.01567, 2017.

[27] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.

[28] Y. Wang, M. Sheng, X. Wang, L. Wang, J. Li, Mobile-edge Computing: Partial Computation Offloading Using Dynamic Voltage Scaling, *IEEE Transactions on Communications*, Vol. 64, No. 10, pp. 4268-4282, October, 2016.

[29] M. G. Piorkowski, N. Sarafijanovic-Djukic, *Crawdad Dataset epfl/mobility (v. 2009/02/24)*, https://crawdad.org/epfl/mobility/20090224, 2009.

[30] T, Soyata, R. Muraleedharan, C. Funai, M. Kwon, W. Heinzelman, Cloud-Vision: Real-time Face Recognition Using A Mobile-Cloudlet-Cloud Acceleration Architecture, *2012 IEEE Symposium on Computers and Communications (ISCC)*, Cappadocia, Turkey, 2012, pp. 59-66.

[31] X. Chen, Decentralized Computation Offloading Game for Mobile Cloud Computing, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 26, No. 4, pp. 974-983, April, 2015.

Biographies

Maoqiang Wu is now a Ph.D. student in Guangdong University of Technology, China. He received the M.S. degree from Guangdong University of Technology in 2017. His researching interests mainly focus on artificial intelligence, mobile crowdsensing, and connected vehicles.

Xumin Huang received the M.S. degree from Guangdong University of Technology, China, in 2016, and the Ph.D. degree at the same school in 2019. His research interests mainly focus on wireless communication and networking.

Beihai Tan received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2007. He is currently a Faculty Member for Future Information Network and Data Laboratory, Guangdong University of Technology, Guangzhou. His current research interests include artificial intelligence, machine learning and their applications.

Rong Yu received his B.S. degree from Beijing University of Posts and Telecommunications, China, in 2002 and Ph.D. degree from Tsinghua University, China, in 2007, respectively. He is currently a Professor at Guangdong University of Technology. His research interests include edge computing, connected vehicles, smart grid and Internet of Things.