Spears and shields: attacking and defending deep model co-inference in vehicular crowdsensing networks

Maoqiang Wu¹, Dongdong Ye¹, Chaorui Zhang² and Rong Yu¹*

Abstract

Vehicular CrowdSensing (VCS) network is one of the key scenarios for future 6G ubiquitous artificial intelligence. In a VCS network, vehicles are recruited for collecting urban data and performing deep model inference. Due to the limited computing power of vehicles, we deploy a device-edge co-inference paradigm to improve the inference efficiency in the VCS network. Specifically, the vehicular device and the edge server keep a part of the deep model separately, but work together to perform the inference through sharing intermediate results. Although vehicles keep the raw data locally, privacy issues still exist once attackers obtain the shared intermediate results and recover the raw data in some way. In this paper, we validate the possibility by conducting a systematic study on the privacy attack and defense in the co-inference of VCS network. The main contributions are threefold: (1) We take the road sign classification task as an example to demonstrate how an attacker reconstructs the raw data without any knowledge of deep models. (2) We propose a model-perturbation defense to defend against such attacks by injecting some random Laplace noise into the deep model. A theoretical analysis is given to show that the proposed defense mechanism achieves $\epsilon$-differential privacy. (3) We further propose a Stackelberg game-based incentive mechanism to attract the vehicles to participate in the co-inference by compensating their privacy loss in a satisfactory way. The simulation results show that our proposed defense mechanism can significantly reduce the effects of the attacks and the proposed incentive mechanism is very effective.

Keywords: Deep model co-inference, Differential privacy, Vehicular crowdsensing network, Stackelberg game

1 Introduction

With the development of the Internet of Vehicle (IoV), more and more vehicle-to-everything (V2X) communication technologies emerge, such as IEEE-based dedicated short-range communication (DSRC) technologies and 3GPP-based LTE technologies [1, 2]. These technologies support stable wireless communication between vehicles and roadside infrastructures [3, 4]. Meanwhile, artificial intelligence becomes more and more popular. In the future 6G vision, there is no doubt that deep neural models will appear everywhere including Vehicular CrowdSensing (VCS) networks, one of the key scenarios...
in the future 6G ubiquitous artificial intelligence. In a VCS network, Service Providers (SPs) always require vehicular devices to collect image data of urban regions as the input of deep models and carry out the model inference [5]. With the inference results, the SPs are able to make better decisions and provide higher quality services [6–8].

However, the existing device-only and edge-only inference paradigms are hard to support the deployment of deep model inference in VCS networks. On one hand, the vehicular device has to collect the streaming image data quickly and use them to perform the model inference when driving at a high velocity. On the other hand, a more complicated deep model consumes more computation resources and energy. The device-only inference paradigm that runs the model inference in vehicular devices is difficult to meet the two requirements due to the limited computation resources and battery capacity of the vehicular devices [9, 10]. Meanwhile, the edge-only inference paradigm allows the vehicular devices to upload their collected data and executes the model inference in edge servers, but brings about considerable communication costs because of the transmission of large-volume raw data [11–13]. Besides, privacy disclosure risk hinders the vehicles from sharing their raw data and being willing to join the VCS networks.

To solve the above disadvantages of model inference paradigms, the device-edge co-inference paradigm was proposed [14]. In this paradigm, a deep model is partitioned into two parts. One part is stored in the vehicular device, while the other part is kept by the edge server. The vehicular device runs the first part of the deep model and uploads the intermediate output. The edge server uses the intermediate data as the input of the rest of the deep model and obtains the final result [15]. Previous works focused on finding out an appropriate partitioning way that has a small-size intermediate output and puts the model layers of large computation load on the side of edge server [14, 16]. This can largely reduce communication costs and improve model inference efficiency. Besides, sharing intermediate model output instead of raw data alleviates the privacy disclosure issues to a certain extent [17].

Nonetheless, the device-edge co-inference paradigm still exists privacy issues. The attacker can reconstruct the raw data by obtaining and analyzing the intermediate model output [18]. Thus, designing defense mechanisms against privacy attacks is necessary. The work in [16, 19] chose a deeper layer as the partitioning point which outputs a smaller-size and less-information intermediate result. The work in [19, 20] used a drop-out mechanism to randomly set some pixel of input data or intermediate data into zero, which reduces the information carried in the intermediate output. The work in [19–21] injected randomly generated noise into input data or intermediate output, which perturbs the reconstruction performance. These defense mechanisms heavily rely on experimental experience and lack theoretical guidance.

In this paper, we introduce the device-edge co-inference paradigm into VCS networks. Through the collaboration of vehicular devices and edge servers, the execution efficiency of deep model inference applications in VCS networks is significantly improved. Besides, we use a black box reconstruction attack, which is able to recover the input raw data only based on the intermediate output, to validate the privacy vulnerability of the co-inference. We then design a model-perturbation defense mechanism against such attacks by adding randomly generated noise to perturb the intermediate output. A differential privacy (DP) theoretical analysis is provided to verify that the proposed mechanism can
guarantee $\epsilon$-DP. Compared with the common defense approach that directly adds noise into intermediate data [19, 21, 22], our proposed mechanism enables a lower privacy budget, i.e., a higher privacy protection level. We further design a Stackelberg game-based incentive mechanism that motivates vehicular devices to join the deep model inference and compensate for their economic loss from potential privacy leakage. The experimental results on the road sign classification dataset demonstrate that our proposed defense mechanism can significantly defend against the reconstruction attack and that the proposed incentive mechanism is effective.

In summary, the main contributions of this paper are as follows.

- We introduce the device-edge co-inference paradigm into VCS networks. The vehicular devices and edge servers work together to improve the efficiency of deep model inference and reduce the communication costs in VCS networks.
- We adopt a black-box reconstruction attack to recover the input image in the road sign classification task. This demonstrates the privacy vulnerability of the co-inference paradigm, which limits its deployment in VCS networks.
- We then propose a model perturbation mechanism that perturbs the model parameters to defend against the reconstruction attack. A DP theoretical analysis is provided as a theoretical guidance to alleviate privacy breaches in the co-inference of VCS networks.
- We further propose a Stackelberg game-based incentive mechanism. The mechanism quantifies the privacy loss of each vehicle by using DP properties and compensates them in a satisfactory way, thus attracting vehicles to join the co-inference in VCS networks.

The remainder of this paper is organized as follows. Section 2 introduces the co-inference paradigm for VCS networks and the reconstruction attack upon it. Section 3 describes the proposed model perturbation defense and the related analysis. Section 4 formulates the incentive mechanism design problem as a Stackelberg game. Section 5 provides a detailed description of game theory analysis. The simulation results and performance evaluation are shown in Sect. 6. Finally, the concluding remarks are made in Sect. 7.

### 2 Privacy vulnerability of co-inference

In this section, we first present the device-edge co-inference paradigm in VCS networks and then adopt the black-box reconstruction attack to demonstrate its privacy vulnerability.

#### 2.1 Co-inference in vehicular crowdsensing networks

Figure 1 gives an overview of a co-inference paradigm of VCS networks over one urban region. We describe the main entities as follows.

- **Vehicles** are running across the urban area and recruited by the SP to execute crowdsensing and deep model inference. Each vehicle is equipped with sensors and a
vehicular device. The sensors are used to collect data at a high rate while the vehicular device has the processing and storage resources to execute a portion of the deep model using the acquired data as input.

- **Edge Server** is rented by the SP to carry out the deep model inference. The edge server has significantly more computational capabilities and capacity than vehicular devices, allowing it to operate the more complex parts of the deep model. The deep model’s final calculation outputs assist the SP in making intelligent decisions.

- **Deep Model** is partitioned by the SP into two parts. The first part, which has a lower computation load, is kept in vehicular devices, while the remainder, which has a higher computation load, is kept in the edge server. The intermediate data, which is the result of the deep model of vehicular devices, is sent to the edge server. The intermediate data is used as input by the edge server to run its deep model section and obtain the final results.

The cooperation between vehicular devices and edge servers can reduce communication costs and deep model inference delay. The vehicular devices do not share raw data in the co-inference paradigm, but they are still vulnerable to privacy leakage, as illustrated by the following privacy attack.

### 2.2 Spears: black-box reconstruction attack

As shown in Fig. 2, the vehicular device stores the first part of layers $f_{01}$, while the edge server keeps the remainder $f_{02}$. The vehicular device inputs raw image data $x_0$ and obtains the intermediate output $v_0 = f_{01}(x_0)$. The attacker tries to be an eavesdropper in VCS networks and intercepts the vehicular device’s shared $v_0$. We consider a black-box setting that the attacker doesn’t know the structure and parameters of the deep model $f_{01}$. But it could query the model, i.e., use arbitrary data $X$ as input to run the model and observe the intermediate outputs $V = f_{01}(X)$. This assumption happens when the SP releases its APIs to other users. The black-box setting is more realistic than the white-box setting in which the structure and parameters of the deep model $f_{01}$ are accessible. It is harder for the attacker to reconstruct the image data under the black-box setting than under the white-box setting [18]. To this, the attacker can train an inverse model
\( g_\omega = f_{\theta_1}^{-1} \) to learn the inverse mapping from the intermediate output \( V \) to the original input \( X \).

The detailed attack process is shown in Algorithm 1 and includes three phases. In the observation phase, the attacker uses a set of samples \( X = \{x_1, \ldots, x_n\} \) to query \( f_{\theta_1} \) and gets \( V = \{f_{\theta_1}(x_1), \ldots, f_{\theta_1}(x_n)\} \). Here we consider that \( X \) follows the same distribution of \( x_0 \). In the learning phase, the attacker trains the inverse model \( g_\omega \) with \( V \) as inputs and \( X \) as targets. The loss function is given as

\[
l(\omega; X) = \frac{1}{n} \sum_{i=1}^{n} \left\| g_\omega(f_{\theta_1}(x_i)) - x_i \right\|_2^2.
\]  

(1)

Note that the structure of \( g_\omega \) needs not to be related to that of \( f_{\theta_1} \). In our experiment, we use a totally different structure. In the reconstruction phase, the attacker inputs \( v_0 \) into the trained inverse model and obtains the recovered image \( x'_0 = g_\omega(v_0) \).
3 Differential privacy method

In this section, we first introduce preliminaries on DP method and then describe the proposed model perturbation defense mechanism. A theoretical analysis is given to show that the proposed defense mechanism provides $\varepsilon$-DP.

3.1 Preliminaries on differential privacy

DP is a statistical framework for measuring the privacy risk. The definition of $\varepsilon$-DP is as follows [23].

**Definition 1** ($\varepsilon$-DP) Given two neighboring inputs $X$ and $X'$ which differ in at least one sample, a randomized mechanism $f$ provides $\varepsilon$-DP if

$$\Pr[f(X) \in S] \leq e^\varepsilon \Pr[f(X') \in S].$$

(2)

According to the above definition, given any neighboring inputs $X$ and $X'$ into the mechanism $f$, the probability of their outputs being in the same range $S$ is characterized by $\varepsilon$. The parameter $\varepsilon$ denotes the privacy budget. A smaller $\varepsilon$ leads to a better privacy protection for the vehicular device. That is to say, given any output, the attacker cannot tell if it is generated by inputting $X$ or $X'$.

A common defense approach is to introduce randomly generated noise of some specific probability distribution into the output $f(\cdot)$ [24]. One probability distribution used widely in DP is Laplace distribution denoted by $\text{Lap}(0, \sigma)$, where $0$ is the mean and $\sigma$ is the scale. The Laplace Mechanism [20] is defined by

$$M_f(X) = f(X) + \text{lap}(0, \frac{\Delta f}{\varepsilon}).$$

(3)

It provides $\varepsilon$-DP when the added noise is sampled from $\text{Lap}(0, \sigma)$ with $\sigma \geq \frac{\Delta f}{\varepsilon}$. Here $\Delta f$ is the global sensitivity indicating that the maximum difference between the outputs $||f(X) - f(X')||_1$ with any pair of inputs $X$ and $X'$. 

---

**Algorithm 1** Black-box Reconstruction Attack Algorithm

*Input:* input data $X = x_1, x_2, \ldots, x_n$ of the same distribution from target data $x_0$, output $v_0$ of target data, batch size $B$, epoch number $E$, learning rate $\eta$

*Output:* recovered data $x'_0$

1: **Observation Phase**
2: query the model by input data $V = f_{E1}(X)$
3:
4: **Learning Phase**
5: initialize parameters of inverse network $\omega_0$
6: for each epoch $1 \leq t \leq E$ do
7: $\beta \leftarrow \{X, V\}$ into batches of size $B$
8: for each batch $b \in \beta$ do
9: $\omega_{t+1} \leftarrow \omega_t - \eta \nabla E(\omega_t; b)$
10: achieve inverse network $g_\omega$
11:
12: **Reconstruction Phase**
13: recover the target data $x'_0 = g_\omega(v_0)$
14:
15: return $x'_0$
3.2 Shields: model perturbation defense

Instead of adding noise directly into the intermediate output $f_{\theta_1}(X)$ [19, 21, 22], we introduce noise into the deep model parameters $\theta_1$ as shown in Fig. 2. This can avoid drastic change of the intermediate output and reduce the negative effect on the following inference. The challenge is that it is difficult to calculate the sensitivity. Hence, we limit the maximum value of the parameters within a fixed bound $G$ to calculate the sensitivity. The clipping operation is carried out during the deep model training [23, 24]. Here, the parameters $\theta_1$ is bounded by $\theta_1/\max\left(1, \frac{||\theta_1||_{\infty}}{G}\right)$. It means that the value of the parameter is not larger than G. Thus, the sensitivity can be approximately calculated as

$$\Delta = \max_{X, X'} ||\theta_1 - \tilde{\theta}_1||_1 \leq 2G,$$

$$f_{\theta_1}(X) \in S,$$

$$f_{\tilde{\theta}_1}(X') \in S.$$  \hspace{1cm} (4)

Next, we add the noise randomly sampled from the Laplace distribution $\text{Lap}(0, \frac{2G}{\epsilon})$ into the bounded parameters. The process of defense mechanism is shown in Algorithm 2. We show that Algorithm 2 gives the $\epsilon$-DP guarantee in Theorem 1.

**Theorem 1** Given the sensitive data $X$ and the deep model $f_{\theta_1}$, Algorithm 2 satisfies $\epsilon$-DP when its injected Gaussian noise $\text{Lap}(0, \sigma)$ is chosen by $\sigma = \frac{2G}{\epsilon}$.

**Proof**

Given any adjacent inputs $X$ and $X'$,

$$\frac{\text{Pr}[f_{\theta_1 + \text{Lap}(0, \frac{2G}{\epsilon})}(x) = S]}{\text{Pr}[f_{\tilde{\theta}_1 + \text{Lap}(0, \frac{2G}{\epsilon})}(x') = S]} = e^{-\frac{\epsilon}{2G}||\theta_1||}$$

$$e^{-\frac{\epsilon}{2G}||\tilde{\theta}_1||} = e^{\frac{\epsilon}{2G}||\theta_1 - \tilde{\theta}_1||} \leq e^{\epsilon}.$$  \hspace{1cm} (5)

According to Definition 1, we have $\epsilon = \frac{2G}{\epsilon}$ and Algorithm 2 satisfies $\epsilon$-DP. The proof is now completed. $\square$

4 Incentive mechanism for co-inference

In this section, we first describe the utility functions of the vehicular devices and the edge server. Then, we formulate the incentive mechanism design problem as a two-stage Stackelberg game problem. We theoretically prove that the game has a unique equilibrium.
4.1 Utility of vehicle and edge server

We consider that there is a set of vehicles $N$ running across the urban area and being recruited by the SP for collecting data of targets. Each vehicle $i$ is running at a constant speed $v_i \in [20, 60]$ km/h. According to [5], a slower vehicle could stay in an area longer and capture more image data. According to [25], the quality of the captured image data from a slower vehicle is higher, i.e., the image data has less vagueness occurred by the shake of sensors. In other words, the collected data of a vehicle with lower $v_i$ has more considerable quality and quantity. After that, the vehicles input the collected data to run the deep model on their vehicular devices.

As aforementioned, the model perturbation defense mechanism provides a privacy protection for vehicular devices. In practice, there are a lot of deep model inference scenarios that need to protect the privacy of vehicular devices. For example, in the scenarios of recognizing target recognition, such as vehicle license plate, pedestrian, road sign, etc., the images contain sensitive information that may expose the drivers’ driving habits or the vehicles’ moving path. It may cause economic loss to the drivers if without the privacy protection. In this paper we consider the deep model inference for the road sign classification scenario and conduct experiments to measure the inference performance under the defense mechanism. The result in Fig. 5 shows that with a larger privacy budget $\epsilon$, the inference accuracy increases. We fit the inference accuracy curve as

$$A_i = a \log(b \epsilon_i + 1),$$

which is used to measure a vehicular device’s inference performance with its chosen $\epsilon_i$.

We can see that the SP expects the vehicles to choose a higher $\epsilon_i$ for a higher inference accuracy. Thus, the SP would design a reward $R$ to compensate the privacy loss of vehicles. Given the reward from the SP, each vehicle’s profit is related to its contribution characterized by $\epsilon_i$ and $v_i$. Similar to [5, 26], the profit of $i$ is denoted as $R(\frac{\epsilon_i}{v_i} / \sum_{i \in I} \frac{\epsilon_i}{v_i})$. The cost of $i$ is defined as its potential privacy loss. Base on the DP analysis, a lower $\epsilon$ means a higher privacy protection level. If the privacy is breached, the economical loss of $i$ is denoted as $c_i + \frac{e}{v_i}$, where $e$ is the expense on executing crowdsensing tasks under driving speed $v_i$, $e$ is the unit expense, and $c_i$ is the estimated value of collected data by $i$. Thus, the utility function of $i$ is given as

$$U_i(\epsilon_i) = \frac{\frac{\epsilon_i}{v_i} R - \epsilon_i(c_i + \frac{e}{v_i})}{\sum_{i \in I} \frac{\epsilon_i}{v_i}}.$$
The SP needs to aggregate the inference results from all the vehicles to alleviate individual error from crowdsensing. Here we consider that the aggregated inference performance is the weighted sum of all the vehicles, which is given as

$$\bar{A} = a \sum_{i \in N} \frac{1}{v_i} \log(b\epsilon_i + 1), \quad (8)$$

where $\frac{1}{v_i}$ is the weight. The SP puts higher weight on the slower vehicles since they usually collect data with higher quality and quantity for model inference [5, 25]. Thus, the utility function of the SP is given as

$$U_S(R) = \lambda \bar{A} - R = \lambda a \sum_{i \in N} \frac{1}{v_i} \log(b\epsilon_i + 1) - R, \quad (9)$$

where $\lambda$ is the conversion factor from inference performance to profits, and $R$ is the reward that the SP offers to the vehicles.

**4.2 Game formulation**

A Stackelberg game is a decision-making tool that contains a leader player and several follower players [26]. Each player is rational and only wants to maximize its utility. The follower players can observe the decision made by the leader player and choose their strategies accordingly [27]. A non-cooperative game is a decision-making tool that contains several rational players competing with each other [28]. They make the decisions at the same time.

In this paper, the problem is how the SP designs the reward $R$ to compensate the privacy loss of vehicles while each vehicle chooses the privacy budget $\epsilon_i$ to complete co-inference tasks of VCS networks under privacy protection. We formulate the problem as a Stackelberg game, where the SP is the leader player, while the vehicles are the follower players. Each vehicle has to decide its optimal response $\epsilon_i^*$ given $R$ and other vehicles’ privacy strategies. Mathematically, the problem is written as

**Problem1:**

$$\begin{align*}
\max_{\epsilon_i} & \quad U_i(\epsilon_i) \\
\text{s.t.} & \quad C1: \quad \epsilon_{\min} \leq \epsilon_i \leq \epsilon_{\max}.
\end{align*} \quad (10)$$

The value of $\epsilon_i$ affects both the inference accuracy and the privacy protection level. A higher $\epsilon_i$ brings to a higher inference accuracy but a higher risk of privacy leakage. Here, $\epsilon_{\min}$ ensures that the inference accuracy under model perturbation defense is acceptable and $\epsilon_{\max}$ ensures that the least privacy protection level requirement is satisfied. The SP can command the expected inference performance by controlling $R$ and aims to find the optimal reward $R^*$ to balance the gained profit from deep model inference and expense on rewarding. Mathematically, the problem is written as **Problem2:**

$$\max_{R} U_S(R). \quad (11)$$
The Stackelberg game is made up of Problem 1 and 2. The objective of this game is to find a Stackelberg Equilibrium (SE) point from which the SP and the vehicles have no motivation to deviate. The definition of SE is as follows [28].

**Definition 2** Let $\epsilon^*_i$ be the optimal solution for Problem 1 and $R^*$ is the optimal solution for Problem 2. The point $(R^*, \epsilon^*)$ is an SE for the proposed Stackelberg game, if it satisfies

\[
U_S(R^*, \epsilon^*) \geq U_S(R, \epsilon^*), \\
U_i(\epsilon^*_i, R^*) \geq U_i(\epsilon_i, R^*), \forall i,
\]

where $\epsilon^*$ with entry $\epsilon^*_i$ is the set of best responses of the vehicles.

The vehicles compete with each other on the reward and thus form a non-cooperative subgame. There may exist a Nash Equilibrium (NE) point where no vehicle can enhance its utility by changing its strategy unilaterally. The definition of NE is as follows [28].

**Definition 3** Let $(\epsilon^*_i, \epsilon^*_j)$ be the solution for Problem 1, where $\epsilon^*_j$ is the set of the best responses of the vehicles except $i$. The point $(\epsilon^*_i, \epsilon^*_j)$ is a NE point for the proposed non-cooperative subgame if it satisfies

\[
U_i(\epsilon^*_i, \epsilon^*_j) \geq U_i(\epsilon_i, \epsilon^*_j), \forall i.
\]

5 Game theory method

In this section, we use the backward induction method of game theory to analyze the two games and find the NE and the SE.

5.1 Subgame nash equilibrium

We use the backward induction method to analyze the existence and uniqueness of the NE in the subgame.

**Theorem 2** There exists a NE point in the non-cooperative subgame among vehicles.

1 Proof

The strategy space of each vehicle is non-empty, convex, and compact. From Eq. (7), $U_i$ is continuous with respect to $\epsilon$ in $[\epsilon_{\min}, \epsilon_{\max}]$. We take the first and second derivatives of $U_i$ with respect to $\epsilon_i$ and obtain

\[
\frac{\partial U_i}{\partial \epsilon_i} = \frac{R \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{\nu_j}}{\left(\sum_{i \in N} \frac{\epsilon_j}{\nu_j}\right)^2} - \left(\frac{c_i + \frac{\epsilon}{\nu_i}}{\nu_i}\right),
\]

\[
\frac{\partial^2 U_i}{\partial \epsilon_i^2} = -\frac{2R^2 \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{\nu_j}}{\left(\sum_{i \in N} \frac{\epsilon_j}{\nu_j}\right)^3} < 0.
\]
We prove that $U_i$ is strictly concave with respect to $\epsilon_i$. Thus, the NE point exists. The proof is now completed. □

Let $\frac{\partial U_i}{\partial \epsilon_i} = 0$ and we get the best response function of $i$ as

$$
\epsilon_i^* = \begin{cases} 
\epsilon_{\text{min}}, & R < \frac{R}{v_i} \\
\sqrt{\frac{R v_i}{k_i}} - a_i v_i, & R \leq R < \frac{R}{v_i} \\
\epsilon_{\text{max}}, & R \geq \frac{R}{v_i},
\end{cases}
$$

(15)

where $a_i = \sum_{j \in N \setminus \{i\}} \epsilon_j \frac{v_j}{v_i}$, $k_i = \epsilon_i + \frac{\epsilon_i}{v_i}$, $R = \frac{k_i \epsilon_i (a_i^{-1} \epsilon_{\text{min}})^2}{a_i}$, and $\overline{R} = \frac{k_i \epsilon_i (a_i^{-1} \epsilon_{\text{max}})^2}{a_i}$.

**Theorem 3** At the NE point for the non-cooperative subgame among the vehicles, the best response of $i$ has a closed-form expression given by

$$
\epsilon_i^* = \frac{R v_i (|N| - 1)}{\sum_{i \in N} v_i k_i} \left(1 - \frac{v_i k_i (|N| - 1)}{\sum_{i \in N} v_i k_i}\right).
$$

(16)

**1 Proof**

According to Eq. (15), we have

$$
\sum_{j \in N \setminus \{i\}} \epsilon_j \frac{v_j}{v_i} = \frac{v_i k_i}{R} \left(\sum_{i \in N} \epsilon_i \frac{v_i}{v_i}\right)^2.
$$

(17)

By computing the summation of this expression for all the vehicles, we obtain

$$
\sum_{i \in N} \epsilon_i \frac{v_i}{v_i} = \frac{R (|N| - 1)}{\sum_{i \in N} v_i k_i}.
$$

(18)

We substitute Eq. (18) into Eq. (17) and get

$$
\frac{R (|N| - 1)}{\sum_{i \in N} v_i k_i} - \frac{\epsilon_i}{v_i} = \frac{v_i k_i}{R} \left(\sum_{i \in N} v_i k_i\right)^2.
$$

(19)

which can be rewritten as Eq. (16). The proof is now completed. □

**Theorem 4** The NE for the non-cooperative subgame is unique if the following condition is satisfied.

$$
\sum_{i \in N} v_i k_i > 2 v_i k_i (|N| - 1).
$$

(20)

**1 Proof**

According to Eqs. (17) and (18), we have

$$
\sum_{j \in N \setminus \{i\}} \epsilon_j \frac{v_j}{v_i} = v_i k_i R \left(\frac{(|N| - 1)}{\sum_{i \in N} v_i k_i}\right)^2.
$$

(21)
Given $R$ offered by the SP and privacy strategies $\epsilon_{-i}$ offered by other vehicles, the best response function in Eq. (15) is denoted as $\epsilon_i^* = B_i(\epsilon_{-i}, R)$. The NE is unique if $B(\epsilon, R) = (B_1, B_2, \ldots, B_N)$ can be proved to be the standard function which meets the following conditions [5, 29].

- **Positivity**: $B(\epsilon, R) > 0$,
- **Monotonicity**: For all $\epsilon$ and $\epsilon'$, $B(\epsilon, R) \geq B(\epsilon', R)$ if $\epsilon \geq \epsilon'$,
- **Scalability**: For all $\mu > 1$, $\mu B(\epsilon, R) > B(\mu \epsilon, R)$.

We first analyze the positivity. According to Eq. (20), we have $\sum_{i \in N} \frac{\epsilon_i}{v_j} < \frac{R}{v_j k_i}$, and thus conclude that

$$\sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} < \frac{R}{4v_j k_i}.$$  \tag{22}

We further conclude that $\sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} \leq \sqrt{\frac{R}{v_j k_i}} \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j}$. Thus, we have

$$B_i(\epsilon_{-i}, R) = v_i \left( \frac{R}{v_j k_i} \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} - \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} \right) > 0,$$  \tag{23}

which satisfies the positivity condition.

We then analyze the monotonicity. Taking the first derivative of $B_i(\epsilon_{-i}, R)$ with respect to $\epsilon_j$, $j \in N \setminus \{i\}$, we have

$$\frac{\partial B_i(\epsilon_{-i}, R)}{\partial \epsilon_j} = \frac{v_i}{v_j} \left( \frac{1}{2} \left( \frac{1}{v_j k_i} \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} - 1 \right) \right).$$  \tag{24}

According to Eq. (22) that $\sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} < \frac{R}{4v_j k_i}$, we have $\frac{1}{2} \left( \frac{1}{v_j k_i} \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j} - 1 \right) > 0$. Thus, the monotonicity condition is satisfied.

Finally we analyze the scalability. We have

$$\mu B_i(\epsilon_{-i}, R) - B_i(\mu \epsilon_{-i}, R) = v_i (\mu - \sqrt{\mu}) \sqrt{\frac{R}{v_j k_i} \sum_{j \in N \setminus \{i\}} \frac{\epsilon_j}{v_j}} \geq 0.$$  \tag{25}

Therefore, $\mu B_i(\epsilon_{-i}, R) \geq B_i(\mu \epsilon_{-i}, R)$ is always satisfied for $\mu > 1$. The scalability condition is satisfied. $B(\epsilon_{-i}, R)$ meets the three conditions and is a standard function. Thus, uniqueness of the NE is proved. The proof is now completed.

$\blacksquare$

Generally, we can obtain the NE point by using the best response dynamics [29]. Problem 1 is resolved and then we analyze the SE in the following.
5.2 Stackelberg equilibrium
We substitute $\epsilon^*_i$ into the objective function of Problem 2 and have

$$U_S = \lambda a \sum_{i \in N} \frac{1}{v_i} \log(h_i R + 1) - R,$$

where $h_i = \frac{b v_i (|N| - 1)}{\sum_{i \in N} v_i k_i} \left(1 - \frac{v_i (|N| - 1)}{\sum_{i \in N} v_i k_i}\right)$.

**Theorem 5** There exists a unique SE for the proposed Stackelberg game among the SP and the vehicles.

1 Proof
The strategy space of the SP is non-empty, convex, and compact. $U_S$ is continuous with respect to $R$ in $[0, +\infty]$. We take the second derivatives of Eq. (26) with respect to $R$ and get

$$\frac{\partial^2 U_S}{\partial R^2} = -\lambda a \sum_{i \in N} \frac{h_i^2}{v_i(h_i R + 1)^2} < 0.$$  

(27)

Thus, $U_S$ is strictly concave with respect to $R$ and the SP has a unique optimal strategy $R^*$ in maximizing its utility. According to Theorem 4, given any reward from the SP, the vehicles always choose a unique set of best responses $\epsilon^*$ to reach the NE. Therefore, when the SP chooses $R^*$, all players determine their optimal strategies. This satisfies the condition in Definition 2 that there exists a unique SE point. The proof is now completed. $\Box$

The objective function of Problem 2 is a concave function and can be solved by using the existing typical convex optimal algorithms (e.g., dual decomposition algorithm [30] ). If the SP has global information, such as $c_i$, he can find out $R^*$ in a centralized manner. However, to protect the privacy of each vehicle, [31] inspires us to design a distributed algorithm that performs the optimization without any private information. The proposed incentive mechanism is carried out cyclically. At each cycle, the SP and the vehicles reach an agreement by Algorithm 3. Under the agreement, the vehicles finish the co-inference tasks by choosing a privacy budget and obtaining the responding rewards. In Algorithm 3, the SP updates the reward value by using a gradient-assisted searching algorithm, i.e., Eq. (28), and offers it to the vehicles. Each vehicle receives the reward value, determines its privacy budget based on Eq. (15), and returns the strategy to the SP. The iterations continue until the difference of the updated reward value is less than a preset threshold. Note that the communication delay is negligible due to the small size of shared information. The frequency of update, i.e., the number of iterations to reach convergence, depends on the learning rate and the threshold. When executing the algorithm, the vehicles conduct wireless communication with an access point (AP). Each vehicular node uploads its strategy information, i.e., privacy budget $\epsilon_i$, to the nearest AP and other vehicular nodes can query this strategy information with negligible delay.
6 Results and discussion

In this section, we conduct the experiments to evaluate the performance of the black-box reconstruction attack and the proposed model perturbation defense mechanism. We also conduct the simulations to evaluate the performance of the proposed incentive mechanism.
6.1 Attack and defense evaluation

6.1.1 Experimental setup

We conduct experiments on the GTSRB dataset for road sign recognition that consists of 39208 samples for training and 12630 samples for testing. We adopt a Convolution Neuron Network (CNN) as deep model with 6 convolution layers and 2 fully connected layers. Each convolution layer has 64 channels and the kernel size is 3. There is a max-pooling layer after every two convolution layers. The model is partitioned at the 2nd, 4th, and 6th convolution layers. We use ADAM as our optimizer and set the learning rate as 0.001. The adopted inverse model consists of two deconvolution layers and one ReLU layer between them. Each deconvolution layer has 64 channels and the kernel size is 3.

6.1.2 Measurement metrics

We use three metrics to measure the attack and defense performance. Mean-Square Error (MSE) measures pixel-wise similarity. Peak Signal-to-Noise Ratio (PSNR) quantifies the pixel-level reconstruction quality of the images. Structural Similarity Index (SSIM) reflects the human perceptual similarity of two images according to their luminance, contrast, and structure. It ranges from [0, 1], where 1 denotes the most similar.
6.1.3 Attack performance

Figure 3 and Table 1 show the recovered performance via black-box reconstruction attack. As shown in Fig. 3, when the deep model is split in a shallower layer, the reconstructed images have high fidelity. When the deep model is split in a deeper layer, the reconstructed images lose some details and become blurry. Even if the split point is in the 6th layer, the details of road signs can still be clearly identified. Table 1 shows that the reconstructed images have higher MSE, PSNR, and lower SSIM, when the deep model is split in a deeper layer. Thus, with a deeper split layer, the black-box reconstruction attack becomes harder.

6.1.4 Defense performance

Figure 4 and Table 2 show the recovered performance under model perturbation defense mechanism with different privacy budget $\epsilon$. We set the split point in the 4th layer and randomly sample noise with privacy budget $\epsilon$ as 5, 10, 500, respectively. As shown in Fig. 4, with a lower $\epsilon$, the reconstructed images become blurrier and lose more details. When $\epsilon = 5$, the details of road signs are hard to be identified. Table 2 shows that the
recovered images under defense with lower $\epsilon$ have higher MSE, PSNR, and lower SSIM. The deep model inference accuracy decreases when $\epsilon$ becomes smaller. The reason is that the injected noise also perturbs the inference results. Generally, the model perturbation defense mechanism reduces the quality of image reconstruction while slightly decreasing the inference performance. These results offer an intuitive guide for the DP and the vehicles for balancing inference performance and privacy protection.

### 6.1.5 Inference performance

For better characterizing the influence caused by the model perturbation mechanism on the inference performance, we set different privacy budgets to observe the inference accuracy depression. Figure 5 shows that the inference accuracy drops with the decrease of $\epsilon$. We also fit the curve based on the observed results.

### 6.2 Incentive mechanism performance

#### 6.2.1 Simulation setup

We consider that there are 5–30 vehicles being recruited for executing VCS and co-inference. The driving speed is randomly chosen in the range of $[5, 15]$ m/s. The profit coefficient is $\lambda \in [1000, 1250]$. The expected value of privacy is $c_i \in [5, 10]$ and the expense for joining the crowdsensing task is $e = 10$.

#### 6.2.2 Performance comparison

Figure 6 shows the performance comparison among the centralized approach, the distributed approach, and the linear approach. The centralized approach assumes that the SP knows the estimated value of collected data of each vehicle so that the SP can use a convex algorithm to directly calculate $R^*$ in a centralized manner. Our proposed distributed algorithm allows the SP to approach the SE point in a distributed manner without the need for any private information. The linear approach also considers that the SP has no knowledge of the vehicles’ private information but the given rewards are linear to the privacy budget of vehicles. As shown in Fig. 6, with the centralized approach, the SP obtains the highest utility. The reason is that the SP knows the estimated value of collected data so that it can directly find out the optimal solution. By using the linear approach, the SP obtains the lowest utility, while the vehicle obtains the highest utilities. The reason is that in the linear approach, the vehicle’s reward is linear to its own

![Figure 7](image-url)
privacy budget, without relation to other vehicles’ strategies. The performance of our proposed distributed algorithm is much better than linear approach but slightly worse than that of the centralized approach with an average distance of 0.05%. In general, our distributed algorithm enables the SP to obtain the highest utility when it has no knowledge of the vehicles’ private information. In addition, the result also shows that the utilities of both the SP $U_S$ and the vehicle increase with the growing profit coefficient $\lambda$. It is because a higher $\lambda$ allows the SP to gain more considerable profit from deep model inference so that the SP provides a higher reward.

### 6.2.3 The impact of privacy value

Figure 7 shows the performance of incentive mechanism with respect to the estimated value of privacy $c_i$. When the evaluated privacy value is higher, the utilities of both the SP and the vehicle decrease. The reason is that when a vehicle estimates a higher value for its privacy, it will choose a lower $\epsilon$ to protect its privacy. The profit of the SP becomes lower accordingly.
6.2.4 The impact of vehicles’ size

Figure 8 shows the performance of incentive mechanism with respect to the number of vehicles. When the number of vehicles grows, $U_S$ increases, while $U_i$ decreases. The reason is that the more vehicles can collect more data to perform deep model inference for the SP. But the increasing number of vehicles brings about more strict competition among them.

6.2.5 The impact of velocity

Figure 9 shows the performance of incentive mechanism with respect to the velocity of vehicles. When the velocity of vehicles becomes higher, the utilities of both the SP and the vehicle decrease. The reason is that when the vehicles drive at a high speed, their weight in the deep model inference decreases. The SP obtains a lower profit and gives lower rewards to the vehicles.

7 Conclusion

In this paper, we adopted the device-edge co-inference paradigm to improve the inference efficiency in VCS networks and studied its privacy preservation. We evaluated the black-box reconstruction attack, which recovers the input data of the vehicular devices, and proposed a model perturbation defense mechanism based on DP theory against the attack. We designed a Stackelberg game-based incentive mechanism that encourages the vehicular devices to participate in the co-inference by compensating their privacy loss. Experimental results demonstrated the effectiveness of our proposed defense mechanism and incentive mechanism.

Abbreviations
VCS: Vehicular CrowdSensing; SP: Service Provider; DP: Differential Privacy.

Acknowledgements
Not applicable

Authors' contributions
MW and DY designed the incentive mechanism and conducted the simulations; MW designed the defense mechanisms, performed the privacy attack and defense experiments, and then wrote the manuscript. All authors discussed the results and revised the manuscript. The authors read and approved the final manuscript.

Funding
The work is supported in part by National Key R&D Program of China (No. 2020YFB1807802, 2020YFB1807800), National Natural Science Foundation of China (No. 61971148), Guangxi Natural Science Foundation, China (No. 2018GXNSFDA281013), and Foundation for Science and Technology Project of Guilin City (No. 20190214-3).

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
1 School of Automation, Guangdong University of Technology, Guangzhou, China. 2 Department of Information Engineering, Chinese University of Hong Kong, Hong Kong, China.

Received: 20 August 2021   Accepted: 3 November 2021
Published online: 17 November 2021
References

1. H. Zhou, W. Xu, J. Chen, W. Wang, Evolutionary v2x technologies toward the internet of vehicles: challenges and opportunities. Proc. IEEE 108(2), 308–323 (2020)
2. X. Liu, X. Zhang, Noma-based resource allocation for cluster-based cognitive industrial internet of things. IEEE Trans. Ind. Inf. 16(8), 5379–5388 (2019)
3. X. Liu, X. Zhang, Rate and energy efficiency improvements for 5G-based IoT with simultaneous transfer. IEEE Internet Things J. 6(4), 5971–5980 (2018)
4. X. Liu, X. Zhang, M. Jia, L. Fan, W. Lu, X. Zhai, 5G-based green broadband communication system design with simultaneous wireless information and power transfer. Phys. Commun. 28, 130–137 (2018)
5. M. Wu, X. Huang, B. Tan, R. Yu, Hybrid sensor network with edge computing for AI applications of connected vehicles. J. Internet Technol. 21(5), 1503–1516 (2020)
6. X. Huang, P. Li, R. Yu, Y. Wu, K. Xie, S. Xie, Fedparking: a federated learning based parking space estimation with parked vehicle assisted edge computing. IEEE Trans. Veh. Technol. 70(9), 9355–9368 (2021)
7. L. He, K. He, Towards optimally efficient search with deep learning for large-scale MIMO systems. IEEE Trans. Commun. PP(99), 1–12 (2022)
8. S. Tang, L. Chen, Computational intelligence and deep learning for next-generation edge-enabled industrial IoT. IEEE Trans. Netw. Sci. Eng. PP(99), 1–12 (2022)
9. X. Huang, R. Yu, D. Ye, L. Shu, S. Xie, Efficient workload allocation and user-centric utility maximization for task scheduling in collaborative vehicular edge computing. IEEE Trans. Veh. Technol. 70(4), 3773–3787 (2021)
10. L. Chen, Physical-layer security on mobile edge computing for emerging cyber physical systems. Comput. Commun. PP(99), 1–12 (2022)

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Maoqiang Wu received his M.S. degree in control science and engineering from Guangdong University of Technology, in 2017. He is currently pursuing the Ph.D. degree in Guangdong University of Technology. His current research interests include edge computing, federated learning, and artificial intelligence.
Technology. His research interests include mobile crowdsourcing and data privacy in edge intelligence.

Dongdong Ye received his M.S. degree in 2018 from Guangdong University of Technology, where he is currently working toward the Ph.D. degree. His research interests include game theory, resource management in wireless communications and networking.

Chaorui Zhang received B.Eng. degree from South China University of Technology, Guangzhou, China, in 2012, and Ph.D. degree in Information Engineering from The Chinese University of Hong Kong, Hong Kong, in 2018. Now he is in the Hong Kong Research Center of Huawei. His research interests include combinatorial optimization, large-scale graph computing, and AI techniques in solving networking problems, such as high performance graph computing, smart grid, power systems, computer networking, and IoT.

Rong Yu received his B.S. degree in communication engineering from the Beijing University of Posts and Telecommunications, China, in 2002, and Ph.D. degree in electronic engineering from Tsinghua University, China, in 2007. After that, he was with the School of Electronic and Information Engineering, South China University of Technology. In 2010, he joined the School of Automation, Guangdong University of Technology, where he is currently a professor. He is the author or coauthor of over 120 international journals and conference papers. He is the co-inventor of over 80 patents in China. He was a member of the Home Networking Standard Committee, China, where he led the standardization work of three standards. His research interests include wireless networking and mobile computing such as edge computing, deep learning, blockchain, connected vehicles, smart grids, and Internet of Things.