future internet

MDPI

Article

CPCROK: A Communication-Efficient and Privacy-Preserving Scheme for Low-Density Vehicular Ad Hoc Networks

Junchao Wang [1], Honglin Li [2], Yan Sun [3], Chris Phillips [3], Alexios Mylonas [1,*] and Dimitris Gritzalis [4]

[1] Cybersecurity and Computing Systems Research Lab, Department of Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK; j.wang32@herts.ac.uk
[2] School of Computer Science, University of St Andrews, St Andrews KY16 9AJ, UK; hl222@st-andrews.ac.uk
[3] School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK; yan.sun@qmul.ac.uk (Y.S.); chris.i.phillips@qmul.ac.uk (C.P.)
[4] Department of Informatics, Athens University of Economics & Business, 10434 Athens, Greece; dgrit@aueb.gr
* Correspondence: a.mylonas@herts.ac.uk

Abstract: The mix-zone method is effective in preserving real-time vehicle identity and location privacy in Vehicular Ad Hoc Networks (VANETs). However, it has limitations in low-vehicle-density scenarios, where adversaries can still identify the real trajectories of the victim vehicle. To address this issue, researchers often generate numerous fake beacons to deceive attackers, but this increases transmission overhead significantly. Therefore, we propose the Communication-Efficient Pseudonym-Changing Scheme within the Restricted Online Knowledge Scheme (CPCROK) to protect vehicle privacy without causing significant communication overhead in low-density VANETs by generating highly authentic fake beacons to form a single fabricated trajectory. Specifically, the CPCROK consists of three main modules: firstly, a special Kalman filter module that provides real-time, coarse-grained vehicle trajectory estimates to reduce the need for real-time vehicle state information; secondly, a Recurrent Neural Network (RNN) module that enhances predictions within the mix zone by incorporating offline data engineering and considering online vehicle steering angles; and finally, a trajectory generation module that collaborates with the first two to generate highly convincing fake trajectories outside the mix zone. The experimental results confirm that CPCROK effectively reduces the attack success rate by over 90%, outperforming the plain mix-zone scheme and beating other fake beacon schemes by more than 60%. Additionally, CPCROK effectively minimizes transmission overhead by 67%, all while ensuring a high level of protection.

Keywords: beacon; mix zone; privacy; pseudonym changing; RNN; transmission overhead; VANET

Academic Editors: Stefano Rinaldi and Alan Oliveira De Sá

Received: 17 February 2025
Revised: 26 March 2025
Accepted: 3 April 2025
Published: 9 April 2025

Citation: Wang, J.; Li, H.; Sun, Y.; Phillips, C.; Mylonas, A.; Gritzalis, D. CPCROK: A Communication-Efficient and Privacy-Preserving Scheme for Low-Density Vehicular Ad Hoc Networks. *Future Internet* **2025**, *17*, 165. https://doi.org/10.3390/fi17040165

1. Introduction

The Vehicular Ad Hoc Network (VANET), as a branch of Intelligent Transportation Systems (ITS), is a dynamic wireless network that seamlessly connects vehicles and roadside units (RSUs) to provide a wide range of traffic-related and entertainment applications. The VANET is designed to enhance traffic safety and efficiency by increasing the awareness of vehicles about their surrounding traffic [1]. The key aspect of VANET applications lies in communication, enabling vehicles to exchange information and interact with other vehicles (V2V), road infrastructure (V2I), and everything else (V2X) [2–5]. This communication facilitates enhanced awareness of surrounding traffic conditions, thereby significantly improving traffic safety and efficiency [1,6,7].

Figure 1 illustrates the VANET scenario, wherein it is formed by a short-haul wireless network connecting RSUs, vehicles, and a backbone network. RSUs are established by trusted authorities for dedicated short-range communication (DSRC). Vehicles are equipped with On-board Units (OBUs) that possess unique IDs, primarily utilized for authentication purposes by the authority. This authentication process enables vehicles to establish safe connections with one another, either directly or by leveraging RSUs for re-transmission. In VANETs, vehicles broadcast safety messages to report accidents or congestion. Additionally, to support various functionalities within the VANET, vehicles are required to periodically broadcast beacons, known as Cooperative Awareness Messages (CAMs), to apprise others of their state information [8].

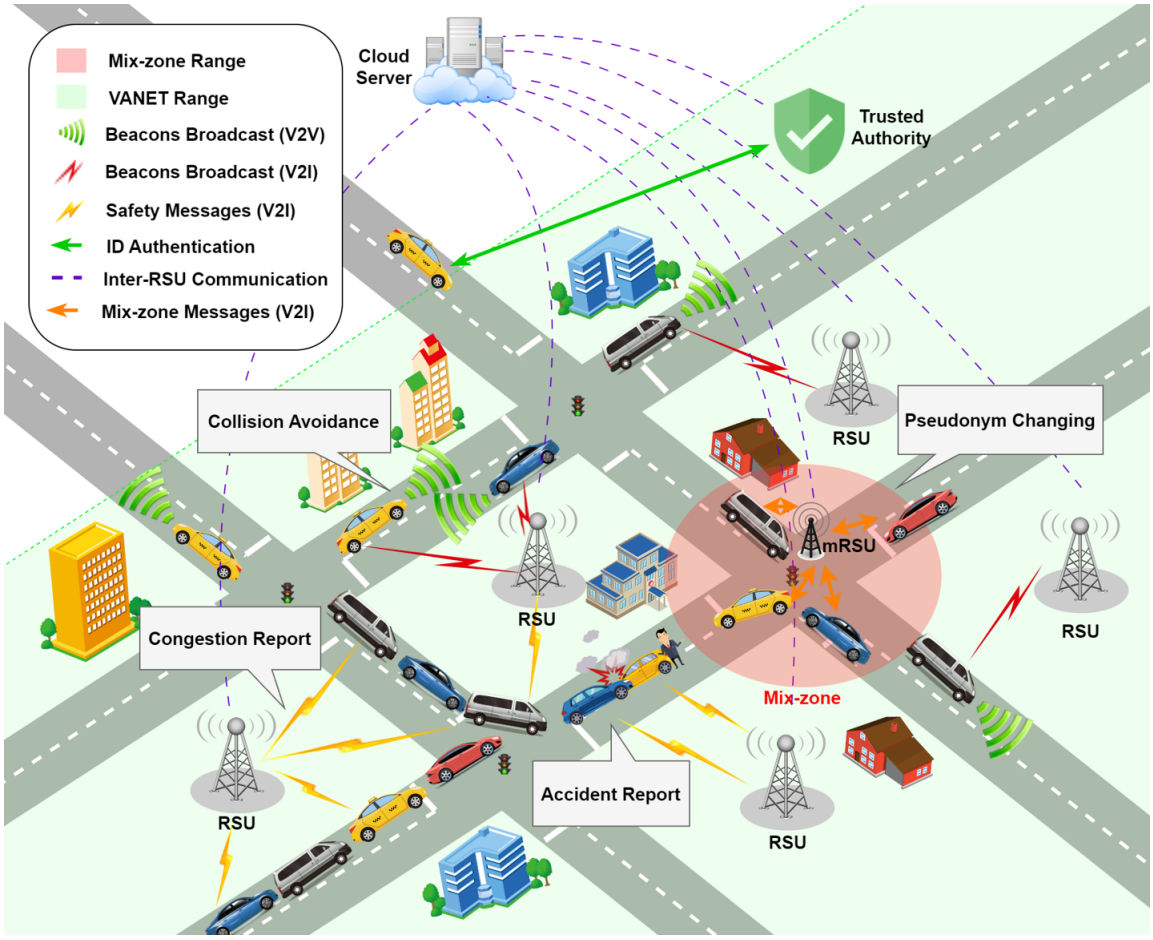


Figure 1. The diagram of the VANET scenario.

Despite the enhanced traffic safety and efficiency brought by periodically broadcasting beacons, the lack of security protection of vehicles in VANETs suffers from the problem of low data trustworthiness [9,10]. There is growing concern about the potential threat to confidentiality, particularly regarding privacy issues [11,12]. Due to the inherent openness of wireless communication in VANETs [13], attackers can intercept and eavesdrop on CAM beacons and exploit the state information they contain, such as vehicle ID, speed, GPS position, timestamp, and acceleration, to uncover a vehicle's trajectory. Consequently, this could result in severe consequences, such as unauthorized vehicle tracking, potential misuse of personal information, and even the facilitation of targeted attacks on individual vehicles or groups of vehicles.

Therefore, VANETs have gained considerable research attention in addressing privacy concerns arising from beacon broadcasting while preserving the core functionality of the VANET [14]. Using pseudonyms instead of real IDs in authentication processes is

widely recognized as an effective measure to safeguard vehicle privacy [15]. The European standard ETSI TS 102 941 [16] exemplifies the responsibility of the authority in issuing pseudonym certificates, which should be updated within a time frame of five minutes [17].

Despite the implementation of the pseudonym, the threat of pseudonym-linking attacks persists, enabling attackers to associate previous and new pseudonyms [18]. These attacks can take on a syntactic level, involving correlation of previous and current pseudonyms when a vehicle changes its pseudonym individually, or a semantic level, wherein the vehicle behavior and trajectory are estimated based on the state information of broadcast beacons to unveil identity and undermine pseudonym changes [19].

To combat pseudonym-linking attacks, various mix-zone-related approaches have been proposed [19–21]. As shown in Figure 1, vehicles collectively replace their pseudonyms upon leaving the mix zone by communicating with the mix-zone Roadside Unit (mRSU) [20]. Anonymity algorithms like k-anonymity [22] are employed to make the pseudonyms used within the mix zone unlinkable. This makes it challenging for attackers to associate a vehicle's trace inside and outside the mix zone, as well as the new pseudonym, with the vehicle's real ID or previous pseudonyms based solely on beacons [23]. However, most research assumes a high-vehicle-density scenario where vehicles have enough neighbors to maintain anonymity. Yet the efficacy of the mix zone is compromised when there are insufficient vehicles, allowing attackers to employ beacon-based semantic attacks and disclose vehicle trajectories despite the full anonymity of beacons [24]. A vehicle may not be able to travel to a high-density area where privacy protection remains a priority for an extended period. Therefore, changing pseudonyms in low-density areas is a critical issue that must be addressed to ensure privacy.

Several studies [25–27] have addressed the problem of low vehicle density by implementing fake beacons generated by the RSU, commonly referred to as chaff or decoy beacons. The purpose of these beacons is to mimic real vehicles and artificially augment the vehicle density. The inclusion of supplementary fake beacons poses a challenge for adversaries attempting to accurately link inbound and outbound traces within the mix zone. While these approaches effectively mitigate the susceptibility to pseudonym-linking attacks in scenarios characterized by low vehicle density, they also introduce new challenges.

It is crucial to generate accurate fake beacon trajectories for the pseudonym-changing scheme. These trajectories should mimic genuine vehicle behavior in terms of location and speed, but originate from different exits to confuse attackers. However, generating such trajectories can be challenging due to the requirement for timely updates on the vehicle's behavior to accurately predict its trajectory. Inaccurate predictions may potentially expose vehicle trajectories to attackers through semantic analysis.

On the other hand, most fake-beacon-based pseudonym-changing schemes involve generating multiple fake trajectories to effectively deceive attackers [22,24,26]. However, generating fake trajectories to mimic real vehicles requires the fake beacons to be processed by RSUs similarly to genuine ones. While this approach reduces the risk of vehicle privacy leakage, it also introduces excessive network overhead to the VANET, threatening the environment awareness functionality of the VANETs [28,29].

Inspired by the capability of neural network models to assist in vehicle trajectory prediction, which can be further applied to the protection of vehicle privacy, we propose the Communication-Efficient Pseudonym-Changing Scheme Within the Restricted Online Knowledge Scheme (CPCROK) to protect vehicle privacy in low-density VANETs and effectively address the shortcomings of the aforementioned pseudonym-changing scheme. CPCROK outperforms the state-of-the-art safeguards (the plain mix zone, other fake beacon schemes) in reducing the attacker's success rate while minimizing transmission overhead.

The contributions of this work are summarized as follows:

1. We propose CPCROK to protect vehicle privacy in VANET scenarios with minimal additional network overhead, even in low-vehicle-density and limited real-time vehicle information scenarios. CPCROK achieves this through a modular design, consisting of three modules: the IKF-CGTGM, an improved Kalman filter method to capture nonlinear vehicle movements, providing real-time coarse-grained trajectory knowledge for the next phase; the RNN-FGTGM, a lightweight RNN model to capture intricate vehicle motion patterns and adjust predicted trajectories based on previous IKF results; and the FBG module, an iterator to generate a single, highly convincing fake trajectory outside the mix zone to deceive attackers. CPCROK minimizes the need to generate a large number of fake beacons by introducing an accurate hierarchical vehicle trajectory prediction approach and reducing the reliance on real-time vehicle state information.
2. CPCROK outperforms the mix-zone scheme by over 90% in vehicle privacy protection. Compared to other fake beacon strategies, CPCROK shows an improvement of over 50%. Additionally, it reduces the transmission overhead of generating fake beacons by 67%, achieving a commendable level of protection.

The rest of this paper is organized as follows: Section 2 presents the state of the art in pseudonym-changing schemes. Section 3 introduces the system and the threat models. Section 4 details the CPCROK scheme. Section 5 evaluates CPCROK's performance with regards to privacy protection offered and transmission overhead compared with mix zone and other fake beacon schemes. Finally, Section 6 concludes the paper and suggests further work.

2. Related Work

To overcome pseudonym-linking attacks, especially semantic attacks, the mix zone has been proved as a widely acknowledged and the most effective technology [20,30–36]. The mix zone allows a group of vehicles to enter an area within a close time frame, change pseudonyms, and exit through different exits to achieve anonymity.

Freudiger et al. [30] first introduced a mix zone in the VANET named CMIX (Cryptographic MIX), where beacons are encrypted. The pseudonym-changing process is concealed, preventing continuous vehicle tracking by eavesdroppers. In addition, some studies proposed maintaining silence within the mix zone [31,32]. However, this can lead to reduced awareness of the vehicle's surroundings, giving rise to other issues such as the inability to promptly prevent collisions [33]. Conversely, while encryption might lead to increased computational overhead, the inherent low latency of VANETs ensures the timely delivery of beacons. To enhance the privacy protection of CMIX, other encryption-based mix-zone schemes focus on supplementary techniques like game theory [34] and statistical approaches [35,36].

Research on mix-zone deployment locations and methods aims to strike a balance between maximizing privacy protection and minimizing the impact on network performance and resource utilization. Instead of junctions, some studies have concentrated on repositioning mix zones to areas where vehicles come to a full stop. For instance, Lu et al. [37] recommended establishing mix zones at social locations like shopping mall parking lots, while [38] proposed sites like gas stations and toll booths. However, this approach limits the opportunities for encountering a mix zone, which subsequently diminishes privacy due to the scarcity of suitable locations.

Another alternative line of study suggests using dynamic mix zones since the VANET is decentralized and self-organized. A cluster of vehicles initiates a dynamic mix zone, enabling them to change or swap pseudonyms based on their awareness of the surrounding

traffic through encrypted V2V transmissions [39]. This occurs, for instance, when vehicles share a similar direction and velocity [40] or possess simultaneous intent [41]. However, establishing the optimal timing to trigger dynamic mix zones in scenarios with low vehicle density remains a challenging task.

The methods mentioned above either overlook low-vehicle-density scenarios or exhibit inadequate performance in such situations. Recently, some approaches based on generating fake beacons have been proposed to address privacy concerns in low-vehicle-density scenarios. The usage of fake states is widespread in Mobile Ad Hoc Networks (MANETs) [42], where user mobility is uncertain, making it challenging to discern real location information through semantic analysis.

In terms of the VANET, RSUs can generate fake beacons as dummy nodes to artificially increase the number of vehicles and maintain the anonymity. However, in Vehicular Ad Hoc Networks (VANETs), vehicles are constrained to move along roads, and due to the dense interconnections between frequently broadcasted beacons, attackers can more easily reveal complete vehicle trajectories through semantic analysis. Therefore, introducing fake beacons in VANETs presents greater challenges.

In the context of low vehicle density, C. Vaas et al. [26] enhanced the CMIX approach by introducing chaff beacons whenever a vehicle enters a junction to fill up exits. The quantity of chaff vehicles is regulated based on the current number of vehicles within the CMIX zone. Additionally, ref. [43] redesigned the authentication process of employing Cuckoo Filters to differentiate between genuine beacons and chaff beacons. M. Khodaei and P. Papadimitratos [27] suggested using relaying vehicles to cooperatively disseminate decoy traffic without the requirement from vehicles to provide their intended trajectory path to the RSUs. N. Guo et al. [22] introduced an indMZ scheme that involves collaborative vehicles generating randomized versions of a pseudonym until the desired level of anonymity is achieved.

Despite the privacy enhancement offered by fake beacons, a fundamental conflict between VANET functionality and privacy preservation persists. To enhance anonymity, the size of the mix zones is often extended beyond junctions to obscure critical information, such as steering angles, lane changes, and accelerations. Due to imperfect trajectory predictions, the need for increased fake beacons arises to reduce the success rate of tracking by attackers. However, this significantly amplifies the transmission overhead of the system and may potentially impact the Quality of Service (QoS) of the VANET [29,44].

Deep learning has been extensively applied in the field of trajectory prediction and pseudonym changing for vehicles, primarily due to its ability to model complex patterns and dependencies in sequential data. However, the effectiveness of these applications can vary significantly based on the specific architecture used and the context of the application.

Some studies [45,46] explored the use of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for generating vehicle trajectories. While RNNs are adept at processing sequential information, their performance in trajectory prediction at intersections was found to be less accurate due to the lack of practical intersection-related information during training. This shortfall makes the dummy vehicles generated by RNNs more susceptible to being distinguished by attackers. LSTMs, on the other hand, are better suited for long-distance trajectory predictions. They tend to underperform in scenarios involving short distances, such as those encountered at intersections, where the trajectory changes are abrupt and localized.

Zhang et al. [47] shifted focus to reinforcement learning, which is utilized to generate complete vehicle trajectories over large areas that closely mimic the real vehicular states. However, this approach does not specifically address trajectory generation at junctions. Since semantic attacks often exploit the continuity of vehicle states to track vehicles, gen-

erating realistic trajectories within junctions becomes crucial. The short duration vehicles spend at junctions necessitates a method that not only maintains accuracy but also speeds up trajectory generation to effectively confuse attackers.

One of the inherent limitations of deep learning highlighted by these studies is its relative inefficiency in real-time applications. Deep learning models, especially those involving complex architectures like LSTMs and reinforcement learning networks, require substantial computation, which can impede their ability to generate immediate responses needed for intersection scenarios.

These insights suggest a need for optimizing deep learning approaches that cater specifically to the unique dynamics at intersections, possibly by integrating more contextual data or employing models that balance complexity with computational efficiency. Further research could explore hybrid models or advanced training techniques that enhance the real-time capabilities of deep learning applications in vehicular systems.

The challenge of minimizing the generation of fake trajectories while preserving vehicle privacy in low-vehicle-density scenarios remains to be addressed.

3. System and Threat Models

3.1. Mix-Zone Deployment in VANETs Using Fake Beacons

Figure 2 illustrates the system model for implementing fake beacon approaches in mix zones within VANETs at junctions. The mix zone, depicted by the red circle, is established by an mRSU through the periodic transmission of beacons. In VANETs, vehicles periodically broadcast beacons containing their state information $x(t)$ at time t, which consists of coordinates $(x(t), y(t))$ and speeds $(v_x(t), v_y(t))$. These beacons form a chronological sequence, representing the vehicle trajectory r.

As depicted in Figure 2, the solid black line represents the vehicle trajectory with a pseudonym called Alice, which enters the mix zone. After receiving the beacon from the mRSU, the vehicle requests a change in the pseudonyms, to ensure anonymity within the mix zone. Upon receiving a request to change the pseudonym, the mRSU performs a check to determine if there are enough vehicles within the mix zone to maintain k-anonymity. If there are at least k vehicles present, the mRSU directly applies the k-anonymity algorithm to change the pseudonym of each vehicle. However, if the number of vehicles is insufficient, the mRSU generates fake beacons for each entering vehicle and modifies the pseudonyms accordingly. The dotted black lines, representing all communications and beacons within the mix zone, are encrypted to ensure privacy.

As illustrated in Figure 2, the vehicle's pseudonym is changed from Alice to Bob, while two additional fake trajectories are generated to represent vehicles with pseudonyms Juliet and Romeo, respectively. This design ensures that the attacker cannot distinguish the real trajectory of the vehicle among the three trajectories. By generating multiple fake trajectories, the system prioritizes the identity and location protection of individual vehicle identities.

However, various studies have shown that attackers can identify real trajectories from fake trajectories by analyzing the semantic information of beacons, especially when the generated fake trajectories lack authenticity. To address this concern, researchers often generate a large number of fake beacons to deceive attackers, but this approach leads to a significant increase in transmission overhead. In light of this, the study aims to address the threat model associated with beacon semantic analysis. Subsequently, the CPCROK scheme will be presented in Section 4, which safeguards vehicle identity and location privacy without compromising the core functionality of the VANETs.

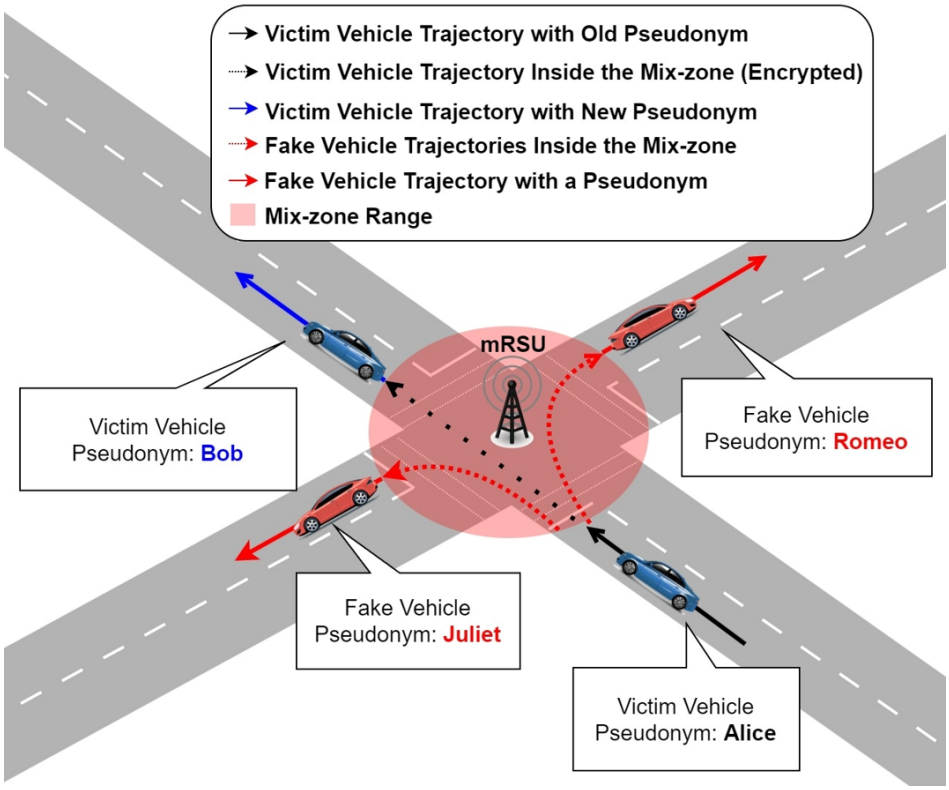


Figure 2. Mix-zone deployment in VANETs using fake beacons.

3.2. Threat Model Based on the Beacon Semantic Analysis

By intercepting the beacons transmitted within the VANET, the attacker conducts beacon semantic analysis using the beacons emitted by the targeted vehicle as it enters the mix zone (indicated by the black solid line in Figure 2), with the intention of predicting sensitive data, like GPS position, speed, and timestamp, once the targeted vehicle exits the mix zone (indicated by the blue solid line in Figure 2). The attacker then compares the beacons obtained outside the mix zone with the anticipated trajectory, identifying the targeted vehicle based on the closest match. It is important to note that there are two common assumptions regarding mix-zone vehicle beacon semantic analysis [30]: First, the attacker can access the beacons outside the mix zone (indicated by the solid line), while the beacons inside the mix zone (indicated by the dotted line) remain invisible due to encryption. Second, the attacker operates remotely instead of resorting to hacking cameras or tailing vehicles, as it requires a higher level of attacker capability and greater costs. In other words, the adversary relies solely on collecting beacons to gather information such as speed, position, and timestamps to uncover traces.

This study will take the multi-hypothesis-tracking (MHT) method originated from [48] and simplified in [24] as an example to illustrate how beacon semantic analysis can effectively unveil vehicle trajectories. In brief, the MHT method aims to match the new measurement driving outwards from a mix zone to a piece of trace that leads into the same junction, as detailed below.

3.2.1. Trajectory Prediction

For each mix zone, the adversary classifies all inbound beacons by their pseudonyms as a trace set R and then forms all first beacons of outbound traces as a candidate set S. To estimate the state of each trajectory r in R at the exit, the adversary employs a Kalman filter. The state of a vehicle x at a given time t is initially modeled as $x_t = x_{t-1} + w$, while the linear relation between the state measurement z and the state x at time t can be

represented as $z_t = Ax_t + v$, where H is a measurement matrix and v is the measurement noise. The noises w and v, are set to obey normal distribution with zero means and a very small standard deviation. The matrices A and H are:

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

where Δt represents the time interval between two estimations. For the iteration in r, Δt equals the beacon interval. For the final estimation of arrival at the exit, Δt equals the difference $\Delta t_{r_n,s}$ between the timestamp t_{r_n} of the last state in r and the timestamp t_s of each candidate s in S.

When a new measurement is iterated at time t, the Kalman filter updates the estimated state $\hat{x}$ without correction, along with the error covariance $\bar{P}$ affected by the disturbance noise.

$$\hat{x}_t = Ax_{t-1} \tag{1}$$

$$\bar{P}_t = AP_{t-1}A^T + Q \tag{2}$$

where P_{t-1} is the error covariance at the previous timestamp $t-1$, and Q is a carefully selected covariance matrix. Using $\hat{x}$ and $\bar{P}$, the control input matrix B can be calculated as follows:

$$B = HP_tH^T + R \tag{3}$$

where R is another covariance matrix that is carefully selected. The variables in the Kalman filter are then updated as follows:

$$K = \bar{P}H^TR^{-1} \tag{4}$$

$$\bar{x} = \hat{x} + K(z_t - H\hat{x}_t) \tag{5}$$

$$P_t = \bar{P} - \bar{P}H^T(H\bar{P}H^T + R)^{-1}H\bar{P} \tag{6}$$

where K represents the Kalman gain. The estimated state after correction x_t and the updated error covariance P_k are utilized in the subsequent calculation at the next timestamp $t+1$. Lastly, the adversary leverages $\Delta t_{r,s}$ to update the Kalman filter and acquire the estimation $x_{r,s}$ for each r at every timestamp at each timestamp t_s.

3.2.2. Victim Recognition

Using the final state estimation $x_{r,s}$, the adversary computes the probability of an end-initial state pair belonging to the same vehicle. This probability set, denoted as $\Phi_{r,s}$, is calculated as follows:

$$\Phi_{r,s} = N(s - x_{r,s}, B_r) \tag{7}$$

where B_r denotes the latest control input matrix for r, and $N(x, P)$ denotes the normal distribution:

$$N(x, P) \sim exp[-\frac{1}{2}x^TP^{-1}x]/\sqrt{2\pi|P|} \tag{8}$$

Next, the adversary eliminates probabilities in $Phi_{r,s}$ that fall below the threshold η in order to reduce computational load. Subsequently, the adversary matches each candidate from set S with another candidate from set R by selecting the pair with the highest probability in $Phi_{r,s}$. This pair is then removed, and the matching process is repeated until either set R or S becomes empty. Following these steps, each incoming trace is associated with a hypothetical outgoing trace at each junction. By connecting all the hypothetical

traces together, the success rate of the adversary can be measured by the number of fully revealed traces.

Originally, the MHT model [48] employed Bayes' theorem to compute the probability of the current hypothesis, typically based on the product of prior hypotheses since previous trajectory hypotheses were also uncertain. However, in our scenario, it has been established that beacons outside the mix zone are not encrypted, and hence the trajectories of the vehicles before entering the mix zone are known (as they use the same pseudonym). For prior trajectory segments, there is no need to apply Bayes' theorem; instead, we use Equations (7) and (8) as a simplification to select the highest probability hypothesis matches for trajectories entering and exiting the mix zone.

4. Proposed CPCROK Scheme

As previously discussed, in high vehicle density scenarios, there is no need to use CPCROK, as the abundance of neighboring vehicles is sufficient for a standard mix-zone strategy to meet the privacy protection needs of vehicles. However, in low-vehicle-density situations, applying CPCROK to create fake traces would be an appropriate method to enhance privacy protection.

Beacons inside the mix zone are encrypted, and therefore the vehicle behavior inside the mix zone is unknown and invisible to the adversary. In contrast, mRSUs can interpret trajectory preferences from internal beacons, including the pattern of acceleration and deceleration, the steering arc, and the duration of staying in a junction. Thus, the fake trace by trajectory prediction is more likely to attract the attention of attackers.

As less realistic fake trajectories have the potential to expose vehicle trajectories to attackers via the beacon semantic analysis mentioned above, most current fake beacon schemes address privacy vulnerabilities in low-vehicle-density scenarios by incorporating a large number of fake beacons. However, the generation of these fake beacons requires significant resource consumption to ensure desirable performance [24]. Hence, we propose the CPCROK approach to effectively protect vehicle privacy in VANET scenarios with minimal additional network overhead, even in situations of low vehicle density and limited real-time vehicle information.

4.1. Overview of the CPCROK Scheme

As depicted in Figure 3, the CPCROK approach comprises three modules: the IKF-based coarse-grained trajectory-generation module (IKF-CGTGM), the RNN-based fine-grained trajectory-generation module (RNN-FGTGM), and the fake beacon generation module (FBG). Each vehicle's trajectory V_i before entering a mix zone is represented by a series of beacons ($x(t_1)$, $x(t_2)$, etc.). Leveraging the steering angle θ in the mix zone, the IKF-CGTGM provides a coarse trajectory prediction P_i within the mix zone for each vehicle, along with corresponding timestamps (t_1, t_2, etc.) in beacons. The RNN-FGTGM then trains an RNN model using P_i to generate outputs that closely match the observed trajectories $V^{'}$ within the mix zone (z_{t+1}, z_{t+2}, etc.). Once the training phase is complete, the inference phase begins, where the IKF-CGTGM follows the same process for each new vehicle V, and the coarse prediction P is adjusted by the well-trained RNN model in the RNN-FGTGM. Finally, the FBG generates extended fake beacons based on the estimated states from the RNN-FGTGM to form the whole fake trajectory V_f. By working together, these modules distribute highly realistic fake beacons using a single fake trajectory.

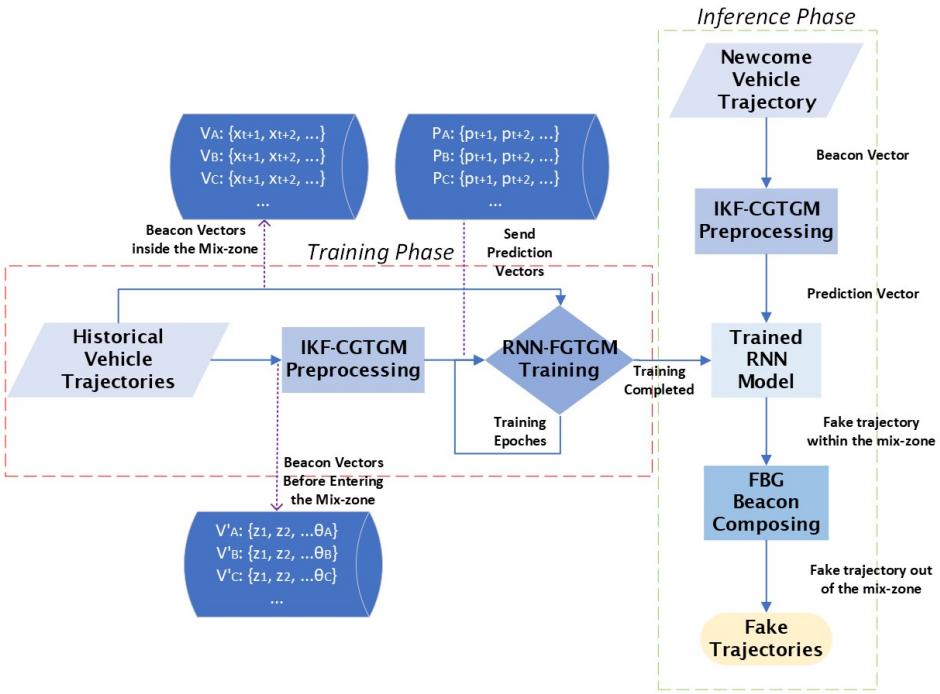


Figure 3. Flowchart of CPCROK modules.

4.2. IKF-CGTGM

In this method, mRSUs are required to store beacons and the number of pseudonym changes for vehicles. When a vehicle enters the mix zone, the mRSU checks if the number of pseudonym changes reaches the limit. If this vehicle has passed adequate mix zones, there is no need to generate fake beacons. However, if this vehicle is still vulnerable, the mRSU will retrieve beacons for this vehicle and report to the IKF-CGTGM.

This module is used to preprocess the real vehicle trajectories to gather coarse-grained predictions. The traditional Kalman filter makes state estimations in a linear dynamic system. However, the vehicle maneuver in the junction cannot be regarded as a linear movement. Steering angle and radius have a great influence on the results predicted by a linear model. Thus, the traditional Kalman filter is not suitable for estimating vehicle behaviors in junctions. As a result, an improved Kalman filter (IKF) is introduced using beacons before a vehicle enters the mix zone to establish a model. Then IKF-CGTGM involves the steering angle of this vehicle to continue updating the model by giving the predicted state a deflection direction.

When a vehicle has passed a junction, the information extracted from beacons is passed to the IKF, which includes vehicle ID, vehicle trajectory before entering the mix zone, length of the beacon sequence in the mix zone, and the system command.

Before the vehicle enters the mix zone, its trajectory information is open to everyone. The traditional Kalman filter can be used to gather estimations x by continuously updating the Kalman gain K and covariance matrix P in the iterative process. In accordance with the traditional Kalman filter, all trajectory points are involved in the iteration as measurements Z, in which the speed is focused only:

$$Z = \begin{bmatrix} v_x \\ v_y \end{bmatrix} \tag{9}$$

In a certain iteration i, the speed information of the trajectory point predicted by the Kalman filter is corrected by the measurement matrix Z_i from $\hat{x}_i$ to x_i.

However, the vehicle information inside the mix zone is not provided to the IKF since the objective is to predict the vehicle's complete journey within the mix zone. The iteration in the traditional Kalman filter stops at the edge of the mix zone. Therefore, a new observation matrix based on the predicted state and the system command is generated to continue the update step.

The system command is the angle θ that the vehicle is expected to turn in the mix zone, given the inbound angle θ_{in} and outbound angle θ_{out}:

$$\theta = \pm(\theta_{in} + \theta_{out}) \tag{10}$$

If the vehicle chooses to take a right turn within $(0°, 180°)$, θ takes a positive value; otherwise, θ takes a negative value. The direction of the predicted state is changed by the system command to create a definite measurement $Z^{'}$:

$$Z^{'} = \begin{bmatrix} v_y \times sin\theta_{correct} + v_x \times cos\theta_{correct} \\ v_y \times cos\theta_{correct} - v_x \times sin\theta_{correct} \end{bmatrix} \tag{11}$$

where:

$$\theta_{correct} = \frac{1}{2}\theta \tag{12}$$

The decision to use $\frac{1}{2}\theta$ for angle correction in our model was empirically determined. If $Z^{'}$ is amended by θ after each iteration, it will cause the vehicle to turn too quickly during predictions. Therefore, $\theta_{correct}$ is used instead to make the rotation smoother. We experimented with various correction factors and found that $\frac{1}{2}\theta$ consistently yielded a coarse-grained trajectory that closely approximates the real one. This specific adjustment is crucial not only for enhancing the performance of our model but also for optimizing computational efficiency. By reducing the complexity of the input to the RNN, we effectively decrease the computational overhead required, thereby accelerating the generation of fake beacons. Our use of the Kalman filter is primarily aimed at increasing the speed of this process, aligning with the overall goal of our system to produce quick responses in dynamic environments.

Hence, the estimation inside the mix zone is corrected by the predicted state in the IKF and the new rotated measurement:

$$x = \hat{x} + K(Z^{'} - H\hat{x}) \tag{13}$$

where K is the Kalman coefficient, which increases in iterations since the IKF chooses to believe the definite measurement created by the prediction from the previous step.

The length of the beacon sequence in the mix zone is used to determine the number of iterations. Ergo, the size of $Z^{'}$ is the same as the number of beacons inside the mix zone. This is designed to avoid padding processes in RNN training. After the iteration is completed, the final predicted trajectory for each vehicle is stored in the database according to the vehicle ID.

The improved Kalman filter differentiates itself from existing models like the extended Kalman filter (EKF) primarily through its handling of information gaps. In our scenario, the RSUs have access to the motion trajectories of the vehicles within the mix zone, including insights into potential exit points. This type of privileged information, crucial for strategic data handling, remains inaccessible to potential external attackers.

In contrast, the extended Kalman filter does not inherently require nor utilize specific directional (steering) information as part of its estimation process. The EKF, designed to handle non-linear systems, approximates state transitions by linearizing about the current

estimate, which, while effective under many scenarios, does not specifically tailor for the strategic manipulation of trajectory data as seen in our application.

The IKF is not devised to enhance the accuracy of trajectory predictions but rather to optimize the speed of computation. By rapidly calculating what we term the coarse-grained trajectory, our system is capable of quickly synthesizing effective and plausible fake beacons without compromising the operational efficiency of the system.

We will deliberately choose an exit different from the real vehicle's trajectory for later use.

4.3. RNN-FGTGM

The aim for this module is to adjust predictions provided by the IKF-CGTGM using a forward-pass RNN to generate precise fake traces containing similar states compared with real vehicles. All the results of the IKF-CGTGM are used as the input dataset to feed the RNN-FGTGM and all beacons inside mix zones are used as the validation set to correct and update the RNN model.

A Recurrent Neural Network (RNN) is an artificial recursive neural network that processes sequential data, evolving in the sequence's direction, and interconnects circular units in a chained network with feedback loops [49]. The RNN is commonly employed for time series prediction [50], especially for vehicle behavior prediction utilizing the sequential state information from beacons. For short distance or maneuver intention, RNNs can provide a higher-level understanding of vehicle behaviors, especially in a simple driving scenario [51]. For instance, Zyner et al. [52] presented a prediction method based on an RNN to predict the destination in a junction by labeling the distance and speed at each junction exit to give early warning of collisions.

As a combination of RNN and privacy protection, the following is a newly proposed algorithmic workflow for generating more accurate fake beacons through trajectory prediction using RNNs.

The RNN model designed for trajectory predictions operates through a series of time steps, each involving several components and processes. At any given time step t, the input vector x^t includes four features: the x-coordinate, y-coordinate, and the velocity components along the x-axis and y-axis. The hidden state h^t represents the output of the hidden layer, which incorporates information from all preceding time steps, ensuring the model captures temporal dependencies effectively.

The output o^t at each time step is generated from the current hidden state h^t, while y^t serves as the target or actual value used for validation purposes. The loss function L^t is calculated to measure the discrepancy between the predicted output o^t and the validation value y^t, aiding in model optimization.

Figure 4 shows the structure of the RNN model for trajectory predictions. In time step t, x^t denotes the input vector, including four features: the x-coordinate, y-coordinate, and the velocity components of the x-axis and y-axis. h^t denotes the value of the hidden unit influenced by all previous hidden units, o^t denotes the output value based on the current hidden unit, y^t denotes the validation value, and L^t denotes the loss function.

There are three key parameters in this RNN model.

- U: The weight matrix applied to the input vector x^t;
- W: The weight matrix applied to the previous hidden state h^{t-1};
- V: The weight matrix that transforms the hidden state h^t into the output o^t.

During the forward pass, the next hidden state h^t is computed using the input x^t and the previous hidden state h^{t-1}, incorporating a linear bias b and processed through an

activation function σ. This function introduces non-linearity into the model, enabling it to capture complex patterns in trajectory data.

$$h^{(t)} = \sigma(z^{(t)}) = \sigma(Ux^{(t)} + Wh^{(t-1)} + b) \tag{14}$$

The generation of the output o^t at time step t is relatively simple, where c is another linear bias:

$$o^{(t)} = Vh^{(t-1)} + c \tag{15}$$

The loss function calculates the difference between $o^{(t)}$ and y^t for each time step. In this model, since the objective is composing precise fake beacons that have similar states as real ones, the Euclidean distance d is selected as the loss function. In each time step t, d is calculated as:

$$d(z_{t,r}, z_{t,o}) = \sqrt{\sum_{i=1}^{n}(x_{t,r_i} - z_{t,o_i})^2} \tag{16}$$

where z denotes the state vector in time step t for the real vehicle r and the RNN output o, including coordinate $(x(t), y(t))$ and speed $(v_x(t), v_y(t))$.

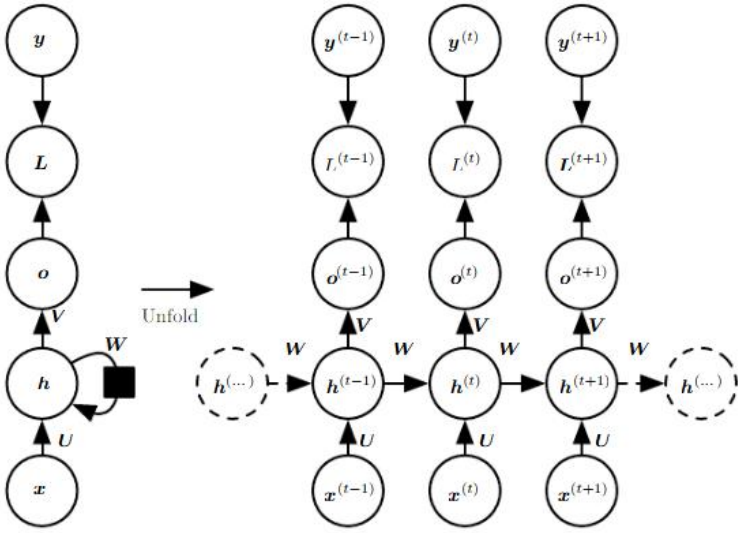


Figure 4. The structure of the RNN model [53].

Gradients are calculated using backward propagation so that the hyperparameters of the RNN are updated using the optimizer. The gradient calculation for V and c is relatively simple, as illustrated in the partial derivatives:

$$\frac{\partial L}{\partial c} = \sum_{t=1}^{\tau} \frac{\partial L^{(t)}}{\partial c} = \sum_{t=1}^{\tau} o^{(t)} - y^{(t)} \tag{17}$$

$$\frac{\partial L}{\partial V} = \sum_{t=1}^{\tau} \frac{\partial L^{(t)}}{\partial V} = \sum_{t=1}^{\tau} (o^{(t)} - y^{(t)})(h^{(t)})^T \tag{18}$$

The gradient of W, U, b is then calculated. During backward propagation, the gradient loss at a certain time step t is determined by the gradient loss corresponding to the output of t and $t+1$. Therefore, the gradient loss of W at t needs to be calculated step by step. First, the gradient of the hidden state at t is defined as:

$$\delta^{(t)} = \frac{\partial L}{\partial h^{(t)}} \tag{19}$$

Thus, the state at $t-1$ can be derived from the state at t:

$$\begin{aligned}\delta^{(t-1)} &= (\frac{\partial o^{(t-1)}}{\partial h^{(t-1)}})^T \frac{\partial L}{\partial o^{(t-1)}} + (\frac{\partial o^{(t)}}{\partial h^{(t-1)}})^T \frac{\partial L}{\partial o^{(t)}} \\ &= V^T(o^{(t-1)} - y^{(t-1)}) + W^T diag(1-(h^{(t)})^2)\delta^{(t)}\end{aligned} \tag{20}$$

For the last state τ, since there are no more time steps after it, the calculation is:

$$\delta^{(\tau)} = (\frac{\partial o^{(\tau)}}{\partial h^{(\tau)}})^T \frac{\partial L}{\partial o^{(\tau)}} = V^T(o^{\tau} - y^{\tau}) \tag{21}$$

Thus, the gradient calculation for W, U, and b can be expressed as:

$$\frac{\partial L}{\partial W} = \sum_{t=1}^{\tau} diag(1-(h^{(t)})^2)\delta^{(t)}(h^{(t-1)})^T \tag{22}$$

$$\frac{\partial L}{\partial U} = \sum_{t=1}^{\tau} diag(1-(h^{(t)})^2)\delta^{(t)}(x^{(t)})^T \tag{23}$$

$$\frac{\partial L}{\partial b} = \sum_{t=1}^{\tau} diag(1-(h^{(t)})^2)\delta^{(t)} \tag{24}$$

The above steps are repeated for several epochs until a specified number of epochs is reached to obtain a low overall loss function. At last, each vehicle has a prediction route corrected by the RNN model. The route consists of many beacons in both the IKF and real vehicle trajectories.

The LSTM model, while powerful in handling long-term dependencies, is not employed in the current scenario primarily due to the nature and scope of the data involved. LSTMs are indeed suitable for trajectory prediction and have been effectively applied in contexts such as intersection path prediction, as demonstrated in [54]. However, in the specific context of the mix zone, the data comprising encrypted beacons represent only a short sequence. The limited spatial extent of the mix zone restricts the length of these sequences, thus diminishing the advantage offered by LSTM's long-term memory capabilities. Furthermore, employing an LSTM in such scenarios introduces a significant computational overhead. Studies, including those reported in [55], have shown that when handling identical time-series prediction tasks, LSTMs require approximately 151% of the computational time needed by simpler RNN models under the best circumstances and up to 227% under the worst scenarios. This increase in computational demand can be attributed to the LSTM's complex internal architecture, which includes multiple gates that manage the flow of information.

Given these considerations, for the specific application of predicting vehicle trajectories within the relatively confined space of a mix zone, the additional computational cost of employing an LSTM does not justify the potential gains in prediction accuracy. Instead, we have opted for a more lightweight RNN model. This choice is driven by the need for rapid computation and efficiency in generating plausible fake beacons, where the simplicity of RNNs offers a more balanced solution. The RNN's less complex structure allows it to process the necessary data sequences quickly and efficiently, making it an ideal choice for applications where speed is crucial and the data sequences are inherently short.

4.4. FBG

This module generates fake beacons according to the predictions provided by the RNN-FGTGM. When a new vehicle approaches a mix zone that has deployed the RNN-

based fake beacon scheme, the mRSU first collects all beacons before this vehicle enters the mix zone. If this vehicle requires a fake trace, the mRSU will continue with the algorithm.

At this time, the direction of this vehicle has been detected by either indicator light information from beacons or steering angle analysis. Thus, the mRSU chooses another random exit and uses the angle of deflection to this exit as the system command to build an IKF and make preliminary predictions. Each time the mRSU receives a new interval beacon before the vehicle leaves the mix zone, an iteration is updated in the IKF.

Next, the data from the IKF are put into the RNN model to make final predictions. To enhance privacy protection in VANET simulations, asynchronous timestamps are used, meaning each vehicle is randomly assigned a start time to broadcast beacons. Vehicles decide whether to encrypt their beacon based on their location relative to the mix zone. Upon entering the mix zone, vehicles continue to broadcast using their previous timestamp and pseudonym, but only RSUs can receive these encrypted beacons. As a vehicle reaches the edge of the mix zone, the RSU assigns a new pseudonym to the vehicle and begins broadcasting at a new random time. This strategy is also applied to fake beacons. Since the length of the RNN prediction sequence is the same as that of the real trajectory, the last state of the prediction sequence becomes the first fake beacon, i.e., the start of the fake trace driving out of the mix zone. Starting from the first point of the fake trace, if the position indicated in the fake beacon is still within the range of the mix zone, this fake beacon will be encrypted by the mRSU.

Prior to the real vehicle exiting the mix zone, each new beacon received by the mRSU is used to update the IKF and the RNN model. Regardless of where the fake beacon generated by the RNN model terminates within the mix zone, the generation of fake beacons will continue following the protocol outlined in [24] to form a complete fake trajectory. As the RNN eventually generates beacons at the edge of the mix zone, the RSU also assigns new pseudonyms and new random times to these fake beacons. Attackers cannot trace vehicles by merely analyzing time intervals. If a fake beacon terminates inside the mix zone, all subsequent fake beacons generated within the mix zone will need to be encrypted to maintain security. This ensures that the continuation of fake beacons adheres to the established security protocols whether or not the real vehicle has left the mix zone.

With the help of the RNN model, the fake trace behaves very closely to a real vehicle, with similar patterns of steering angle and speed transformation in the junction. Therefore, fake traces are accurate enough to deceive attackers, and a small number of fake traces can effectively protect the privacy of vehicles under a low-vehicle-density situation.

5. Evaluation

To evaluate the performance of the CPCROK scheme, we constructed a mix-zone scenario in a VANET environment using Python 3.8.12 and PyCharm (version 2022.3.1; JetBrains, Prague, Czech Republic). This scenario included vehicles entering and exiting the mix zone, the generation of mRSU beacons, beacon encryption, semantic-level pseudonym-changing attacks, and the implementation of various pseudonym-changing schemes. The simulations were conducted on a high-performance desktop computer with an Intel Core i7-11700F @ 2.50 GHz 16-core CPU, 16 GB RAM, and NVIDIA GeForce RTX 3060 graphics card. The computer ran on the Windows 10 64-bit operating system. The experiment setup and discussion are elaborated upon in the subsequent sections.

5.1. Scenario Setting

In a prior study [24], we meticulously designed a traffic simulation that incorporated numerous traffic rules, including overtaking allowance and connected junctions with roads to form a map. We also set up a Poisson vehicle arrival pattern to simulate vehicle entries

and exits. The experimental results indicated that these factors did not significantly impact the attacker, while the success rate of an attacker recognizing a vehicle dropped exponentially as the vehicle passed through several intersections (i.e., changed pseudonyms multiple times). This is because, during other times on the road when the vehicle did not change its pseudonym, attackers could easily track its trajectory segment. Since this paper focuses on the success rate of changing pseudonyms once, the experiment was simplified to only involve passing through junctions of different shapes, removing additional influencing factors.

As illustrated in Figure 5, the simulation encompasses five commonly encountered junction shapes. These include two T-junctions with three potential exits, two crossroads with four possible exits, and a multi-way junction offering five potential exits. The mix zones, depicted as red circles, extend beyond the boundaries of the junctions. For performance evaluation, the radius of all the mix zones is set to 10 m, similar to the attacking strategy proposed in [24].

In designing traffic rules, we considered the worst-case scenario, where the vehicle's dwell time within the mix zone is minimal (with no red light delays or traffic congestion). This approach ensures that attackers receive more closely linked segments of vehicle trajectories. As we have demonstrated in [24] and also seen in [56], traffic simulation experiments spanning multiple mix zones indicate that prolonged stays within the mix zone are highly disadvantageous for attackers, significantly reducing their success rate in tracking trajectories.

For each junction in the simulation, we randomly select two ends. One end is designated as the vehicle entrance, while the other end is designated as the vehicle exit. The arrival interval of vehicles entering the mix zone λ is used to describe the vehicle density in the simulation. We set λ to 2, 4, 6, 8, and 10, representing a new vehicle entering the mix zone every 2, 4, 6, 8, and 10 seconds, respectively. For example, when λ is set to 2, multiple vehicles will appear in the junction simultaneously, with some having just arrived and others about to leave. In contrast, when λ is set to 10, vehicles will rarely encounter other vehicles during their journey inside the junction. The vehicles' speed ranges from 2 to 8. The minimum speed of 2 m/s ensures vehicles keep moving, reducing congestion risks, while the maximum of 8 m/s enhances safety, allowing drivers enough time to react to sudden changes. These limits are consistent with typical urban traffic rules, making the experimental conditions realistic. Their speed gradually decreases as they approach the mix zone, remains constant within the mix zone, and slightly increases at the exit of the mix zone to simulate real-world traffic conditions. All other parameters for the traffic and VANET simulations are shown in Table 1.

Table 1. Simulation parameters.

Module	Parameter	Value
RNN	Learning rate	0.001
	Number of epochs	50
	Number of training vehicles	5000
	Number of inference vehicles	250
Traffic	Maximum speed	8 m/s
	Minimum speed	2 m/s
	Maximum acceleration	0.8 m/s^2
	Maximum deceleration	0.2 m/s^2
VANET	Mix-zone radius	10 m
	Beacon interval	0.3 s

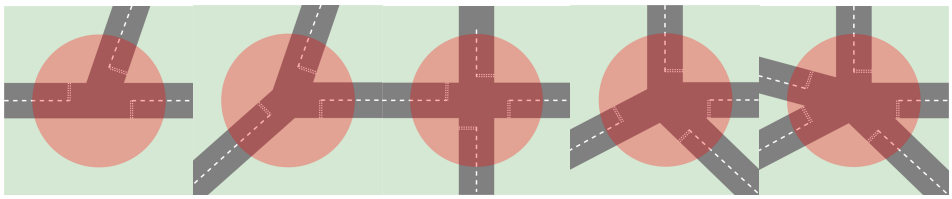


Figure 5. Five commonly encountered junction shapes utilized in the simulation. (**a**) T-junction 1, (**b**) T-junction 2, (**c**) crossroad 1, (**d**) crossroad 2, (**e**) multi-way. The red circular areas indicate the mix-zone ranges, and the green areas represent non-road regions.

5.2. CPCROK Implementation

5.2.1. Data Preparation

By generating a large number of vehicle trajectories passing through the mix zone, as mentioned earlier, we have successfully collected a substantial amount of data. This data have been carefully organized into a dedicated dataset, specifically designed for offline model training in the RNN-FGTGM module. The dataset comprises four features: the x-coordinate, y-coordinate, and velocity components of the x-axis and y-axis, which are fed chronologically. Similarly, the output dataset also contains the same four features, which are used to form fake beacons. To ensure comprehensive RNN training and cover all possible steering angles, random noises are introduced to both entrance and exit. This ensures that even vehicles following the same route will have slightly different steering actions, covering the range of angles from -90° to 90° and enhancing the diversity and effectiveness of the training process.

5.2.2. Model Setting

Our RNN-FGTGM module utilizes the Adam Optimizer, and we have tested combinations of different numbers of neurons (16, 32, 64, 128) and different numbers of hidden layers (1, 2) to evaluate their impact on the success rate and the size of the loss function. Ultimately, we selected 64 neurons and one hidden layer as the hyperparameters, because this combination achieves the lowest loss function value while minimizing the attacker's success rate, all within the constraints of limited computational time. Other parameters are displayed in Table 1.

5.3. Comparison Schemes

The simulations employed the MHT method mentioned in Section 3 as the attack model. Three pseudonym-changing schemes were implemented to compare the performance of the proposed CPCROK scheme, as described below:

- The plain mix-zone (MZ) scheme proposed in [30] changes the pseudonym of the vehicle when it enters the mix zone, but without any additional protection.
- The fake beacon (FB) scheme proposed in [24] generates a fake trajectory that directs the vehicle towards an alternative exit point within the mix zone, bewildering potential adversaries. This fake trajectory is carefully crafted based on the vehicle's pre-entry state and the distance between the entrance and the selected exit.
- The advanced fake beacon (AFB) scheme [24] utilizes an approach similar to the FB scheme but enhances privacy protection by generating two distinct fake trajectories with different estimated states.

Since the traditional Kalman filter cannot provide turning information, it is unable to offer possibilities other than going straight when generating fake trajectories; therefore, there is no need for comparative analysis against the traditional Kalman filter.

5.4. Evaluation Metrics

To evaluate the effectiveness of the pseudonym-changing schemes, we assess the capability to ensure privacy protection and communication efficiency by analyzing the attacking success rate and the number of fake trajectories required for preserving identification and location privacy.

5.4.1. Success Rate

The attacking success rate ρ is defined as

$$\rho = \frac{1}{N} \sum_{i \in V} \lambda_{V,V'} \times 100, \quad \lambda_{V_i,V_i'} = \begin{Bmatrix} 1 & V_i \equiv V_i' \\ 0 & otherwise \end{Bmatrix}$$

where V' represents the set of hypothetical traces, V represents the actual trace set of vehicles, N denotes the number of vehicles (i.e., the size of V), and V_i and V_i' represent individual traces in V and V', respectively. A higher value of ρ indicates that the adversary has successfully matched more vehicle trajectories, which suggests a weaker performance of the pseudonym-changing scheme. Hence, an effective pseudonym-changing scheme should strive to minimize the value of ρ.

5.4.2. Minimal Number of Fake Trajectories

The required number of generated fake trajectories ψ is defined as

$$\psi = \log_{\rho_\lambda} E(\rho) \tag{25}$$

where ρ_λ denotes the success rate when the arrival interval is λ, and $E(\rho)$ denotes the expected success rate. A pseudonym-changing scheme utilizing fake beacons is considered more efficient if it can achieve a certain level of protection while generating fewer fake trajectories.

5.5. Visualization of Fake Beacon Trajectories

Figure 6 visualizes real and fake vehicle trajectories during the inference phase. In the actual experiment, vehicles can enter from any end of the junction, whilst Figure 6 only lists an instance of entering from one direction in each junction and all mix zones are of the same size and settings. Referring to Figure 6, we can obtain the following conclusions: (1) Different types of junctions lead to variations in turning angles and distances covered between fake and genuine trajectories; (2) taking Figure 6c as an example, the distance between the fake vehicle trajectory predicted by the IKF-CGTGM (red dotted line) and the preset exit is very different. The significant deviations when leaving the mix zone are attributed to the lack of post-entry status information, which makes them easily distinguishable by attackers; (3) the fake beacon trajectory (blue dotted line) generated by CPCROK illustrates one beacon leaving from the 12 o'clock direction and another from the 6 o'clock direction at the same time, both deviating by 90 degrees from the entrance direction. The state of this fake vehicle closely approximates the state when the victim vehicle chooses the 12 o'clock exit, effectively confusing attackers.

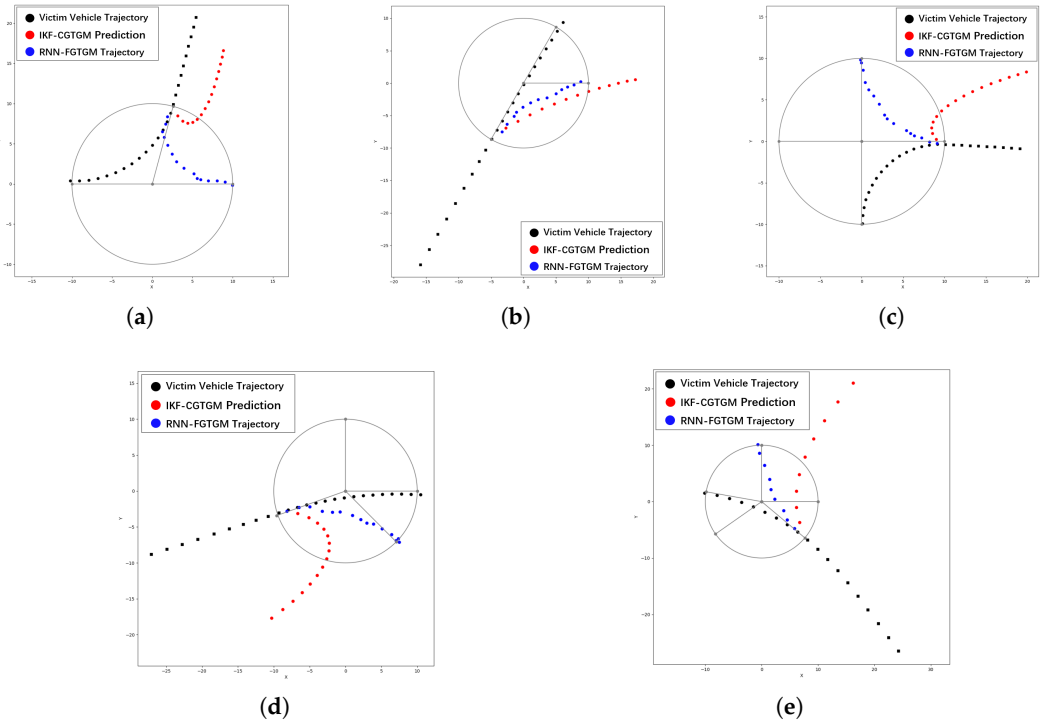


Figure 6. Visualization of fake beacon trajectories generated in CPCROK across various types of junctions. (**a**) T-junction 1, (**b**) T-junction 2, (**c**) crossroad 1, (**d**) crossroad 2, (**e**) multi-way.

5.6. Vehicle Privacy-Preserving Evaluation

Each experimental scenario underwent five trials with different 5000 vehicles and the average value is taken as the result to eliminate randomness.

Figure 7 illustrates the average success rate of attackers under different pseudonym-changing schemes as the vehicle's arrival interval varies.

From Figure 7, the following observations can be made: First, the MZ scheme exhibits the highest attacking success rate, highlighting the vulnerability of vehicles to pseudonym-linking attacks. Then, vehicle privacy can be adequately maintained with an arrival interval $\lambda = 2$, even without additional fake traces. However, as vehicle density decreases, the success rate of attackers gradually increases, reaching nearly 100%. Thus, incorporating fake traces becomes crucial in low-vehicle-density scenarios to enhance privacy protection. Thirdly, CPCROK demonstrates a significant decrease in the adversary's success rate, reducing it by over 90% across varying vehicle density scenarios compared to the mix-zone scheme. Additionally, CPCROK performs better compared to the other two fake beacon schemes, by 60% in low vehicle density and over 50% in normal vehicle density. Lastly, and most importantly, when comparing the FB and AFB schemes, despite the latter generating one additional fake trajectory, there is hardly any improvement in the success rate. This is because the fake trajectories generated by traditional methods do not behave like real vehicles, causing the extra fake trajectory to be identified by attack algorithms as inconsistent with real vehicle patterns. As a result, this trajectory is not matched with the real vehicle.

Simulations were conducted to assess attacks on vehicles as they traverse multiple junctions. Figure 8 shows the success rates for passing through one to four mix zones, considering different vehicle density scenarios with λ values of 2, 6, and 10. The results indicate that the attacker's success rate decreases exponentially with each additional mix zone. This is due to the increased frequency of pseudonym changes as vehicles pass through multiple mix zones, making it challenging for attackers to analyze the beacons.

Additionally, while the FB and AFB schemes offer some privacy protection compared to the MZ scheme, they require vehicles to pass through more mix zones. In contrast, the attacking success rate when CPCROK is implemented drops to be lower than 10% with only a few pseudonym changes, even in scenarios with limited mixing opportunities.

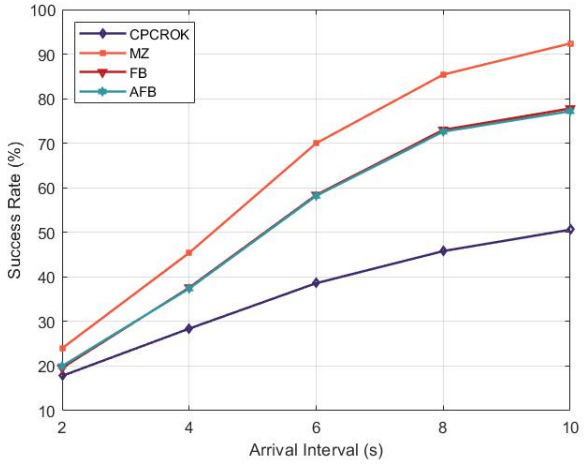
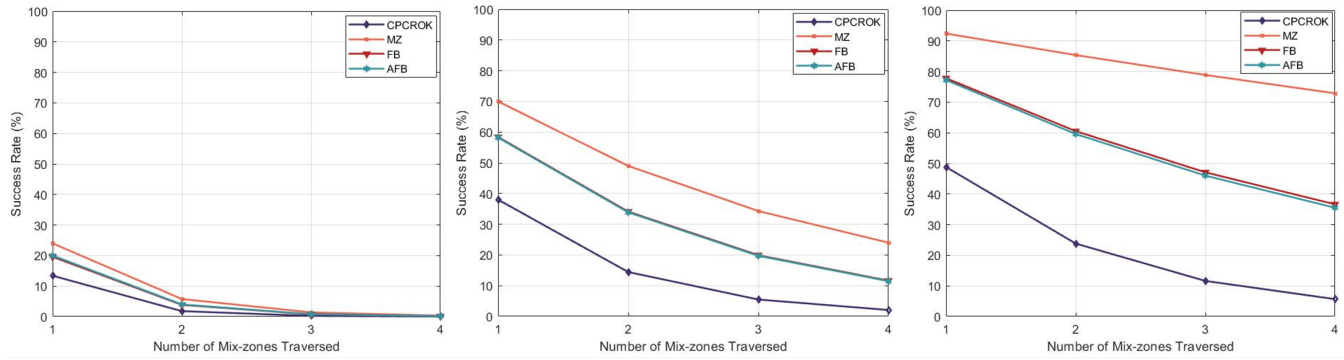


Figure 7. Success rate under different pseudonym-changing schemes as the vehicle's arrival interval varies.



(**a**) (**b**) (**c**)

Figure 8. The success rates of vehicles traversing different numbers of mix zones under varying vehicle densities. (**a**) $\lambda = 2$, (**b**) $\lambda = 6$, (**c**) $\lambda = 10$.

5.7. Communication Efficiency Evaluation

We aim to trade computational overhead for reduced transmission overhead to enhance the stability of VANETs. This approach is justified as the mRSUs leverage backbone networks, which are more robust and reliable, while vehicles are limited to mobile wireless networks.

Figure 9 illustrates the minimum number of fake trajectories ψ required to achieve various success rates for fake beacon schemes. The MZ scheme is excluded since it does not generate fake beacons. For instance, when the success rate is 30%, both the AFB and FB schemes become resource-intensive, particularly at low vehicle density. In contrast, the CPCROK scheme achieves a 30% success rate reduction for adversaries, with only one-third of the overhead of the FB scheme and one-sixth of the AFB scheme when the arrival interval is 10 s. Therefore, CPCROK strikes a balance between privacy protection and resource conservation by generating a minimal number of fake trajectories.

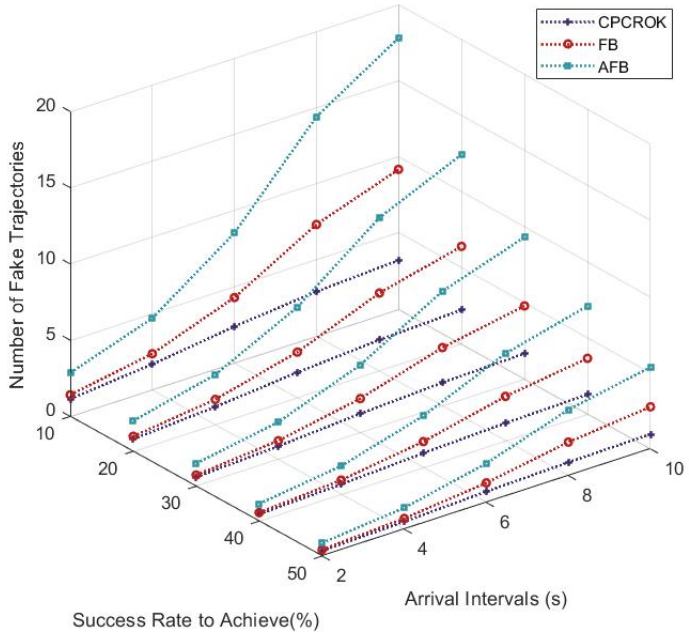


Figure 9. The minimum number of fake trajectories required to achieve various success rates for fake beacon schemes.

6. Conclusions

In this paper, we propose CPCROK to preserve vehicle identity and location privacy in low-density VANET scenarios, without imposing an obvious increase in transmission overhead. In CPCROK, three modules cooperate to generate highly authentic fake beacons, forming a single fake trajectory. This trajectory is authentic enough to deceive attackers, even in low-density VANET scenarios with limited online knowledge. The evaluation results demonstrate that CPCROK can reduce the success rate of attacks by approximately 90% compared to the plain mix-zone approach, and by 60% compared to the state-of-the-art fake beacon-based mix-zone approach. Additionally, CPCROK significantly reduces the additional transmission overhead by generating fewer fake beacons. Specifically, while the FB scheme requires three fake traces and the AFB scheme requires six, CPCROK only necessitates one to achieve the same protection level, resulting in a transmission overhead reduction of 67% compared to the FB scheme and 83% compared to the AFB scheme.

In future work, we will evaluate CPCROK against other pseudonym-changing schemes. We also plan to train and evaluate CPCROK with other datasets, such as NGSIM and HighD. Finally, we plan to extend our threat model to consider other attack strategies (e.g., deep learning-based trackers) and develop a pseudonym-changing mechanism that can adapt to different attack scenarios.

Author Contributions: Conceptualization, J.W.; methodology, J.W. and Y.S.; software, J.W. and H.L.; writing—original draft preparation, J.W.; writing—review and editing, A.M. and D.G.; visualization, J.W. and H.L.; supervision, Y.S. and C.P.; validation, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Kumar, A.; Bansal, M. A review on VANET security attacks and their countermeasure. In Proceedings of the 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 21–23 September 2017; pp. 580–585.
2. Molina-Masegosa, R.; Gozalvez, J.; Sepulcre, M. Comparison of IEEE 802.11 p and LTE-V2X: An evaluation with periodic and aperiodic messages of constant and variable size. *IEEE Access* **2020**, *8*, 121526–121548. [CrossRef]
3. Shahid, M.A.; Jaekel, A.; Ezeife, C.; Al-Ajmi, Q.; Saini, I. Review of potential security attacks in VANET. In Proceedings of the 2018 Majan International Conference (MIC), Muscat, Oman, 19–20 March 2018; pp. 1–4.
4. Papathanassiou, A.; Khoryaev, A. Cellular V2X as the essential enabler of superior global connected transportation services. *IEEE 5G Tech Focus* **2017**, *1*, 1–2.
5. Tan, H.; Zheng, W.; Vijayakumar, P. Secure and efficient authenticated key management scheme for UAV-assisted infrastructure-less IoVs. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 6389–6400. [CrossRef]
6. Luo, G.; Yuan, Q.; Zhou, H.; Cheng, N.; Liu, Z.; Yang, F.; Shen, X.S. Cooperative vehicular content distribution in edge computing assisted 5G-VANET. *China Commun.* **2018**, *15*, 1–17. [CrossRef]
7. Tanuja, K.; Sushma, T.M.; Bharathi, M.; Arun, K.H. A survey on VANET technologies. *Int. J. Comput. Appl.* **2015**, *121*, 1–9.
8. Molina-Masegosa, R.; Sepulcre, M.; Gozalvez, J.; Berens, F.; Martinez, V. Empirical models for the realistic generation of cooperative awareness messages in vehicular networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 5713–5717. [CrossRef]
9. Chen, J.; Lu, H.; Xu, W.; Yang, Y.; Amoon, M.; Kumari, S. A blockchain-based trusted vehicle service recommendation scheme in intelligent transport systems. *Secur. Priv.* **2025**, *8*, e452. [CrossRef]
10. Chen, C.M.; Hao, Y.; Kumari, S.; Amoon, M. An Intelligent Blockchain-Enabled Authentication Protocol for Transportation Cyber-Physical Systems. *IEEE Trans. Intell. Transp. Syst.* 2025, *early access*.
11. Sheikh, M.S.; Liang, J. A comprehensive survey on VANET security services in traffic management system. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 2423915. [CrossRef]
12. Kaur, R.; Singh, T.P.; Khajuria, V. Security Issues in Vehicular Ad-Hoc Network (VANET). In Proceedings of the 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–12 May 2018; pp. 884–889.
13. Bellatti, J.; Brunner, A.; Lewis, J.; Annadata, P.; Eltarjaman, W.; Dewri, R.; Thurimella, R. Driving habits data: Location privacy implications and solutions. *IEEE Secur. Priv.* **2017**, *15*, 12–20. [CrossRef]
14. Hahn, D.; Munir, A.; Behzadan, V. Security and privacy issues in intelligent transportation systems: Classification and challenges. *IEEE Intell. Transp. Syst. Mag.* **2019**, *13*, 181–196. [CrossRef]
15. Manivannan, D.; Moni, S.S.; Zeadally, S. Secure authentication and privacy-preserving techniques in Vehicular Ad-hoc NETworks (VANETs). *Veh. Commun.* **2020**, *25*, 100247. [CrossRef]
16. European Telecommunication Standard Institute (ETSI). *ETSI TS 102 941 Version 2.1.1—Intelligent Transport Systems (ITS); Security; Trust and Privacy Management*; ETSI: Valbonne, France, 2021.
17. European Telecommunications Standards Institute (ETSI). *ETSI TS 102 867 v1.1—Intelligent Transport Systems (ITS), Security, Stage 3 Mapping for IEEE 1609.2*; ETSI: Valbonne, France, 2012.
18. Wiedersheim, B.; Ma, Z.; Kargl, F.; Papadimitratos, P. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. In Proceedings of the 2010 Seventh International Conference on Wireless On-Demand Network Systems and Services (WONS), Kranjska Gora, Slovenia, 3–5 February 2010; pp. 176–183.
19. Boualouache, A.; Senouci, S.M.; Moussaoui, S. A survey on pseudonym changing strategies for vehicular ad-hoc networks. *IEEE Commun. Surv. Tutor.* **2017**, *20*, 770–790. [CrossRef]
20. Buttyán, L.; Holczer, T.; Vajda, I. On the effectiveness of changing pseudonyms to provide location privacy in VANETs. In Proceedings of the European Workshop on Security in Ad-Hoc and Sensor Networks, Cambridge, UK, 2–3 July 2007; pp. 129–141.
21. Beresford, A.R.; Stajano, F. Location privacy in pervasive computing. *IEEE Pervasive Comput.* **2003**, *2*, 46–55. [CrossRef]
22. Guo, N.; Ma, L.; Gao, T. Independent mix zone for location privacy in vehicular networks. *IEEE Access* **2018**, *6*, 16842–16850. [CrossRef]
23. Svaigen, A.R.; Boukerche, A.; Ruiz, L.B.; Loureiro, A.A. BioMixD: A bio-inspired and traffic-aware mix zone placement strategy for location privacy on the internet of drones. *Comput. Commun.* **2022**, *195*, 111–123. [CrossRef]
24. Wang, J.; Sun, Y.; Phillips, C. Fake Beacon: A Pseudonym Changing Scheme for Low Vehicle Density in VANETs. In Proceedings of the 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 20–23 June 2023; pp. 1–7.
25. Wang, J.; Sun, Y.; Phillips, C. Enhanced Pseudonym Changing in VANETs: How Privacy is Impacted Using factitious Beacons. In Proceedings of the 2023 Wireless Telecommunications Symposium (WTS), Boston, MA, USA, 19–21 April 2023; pp. 1–6.
26. Vaas, C.; Khodaei, M.; Papadimitratos, P.; Martinovic, I. Nowhere to hide? Mix-Zones for Private Pseudonym Change using Chaff Vehicles. In Proceedings of the 2018 IEEE Vehicular Networking Conference (VNC), Taipei, Taiwan, 5–7 December 2018; pp. 1–8.
27. Khodaei, M.; Papadimitratos, P. Cooperative location privacy in vehicular networks: Why simple mix zones are not enough. *IEEE Internet Things J.* **2020**, *8*, 7985–8004.

[CrossRef]
28. Weinfeld, A. Methods to reduce DSRC channel congestion and improve V2V communication reliability. In Proceedings of the 17th ITS World CongressITS JapanITS AmericaERTICO, Tokyo, Japan, 7–11 September 2010.
29. Sarkar, N. The impact of transmission overheads on IEEE 802.11 throughput: Analysis and simulation. *J. Sel. Areas Telecommun. (JSAT)* **2011**, *2*, 49–55.
30. Freudiger, J.; Raya, M.; Félegyházi, M.; Papadimitratos, P.; Hubaux, J.P. Mix-zones for location privacy in vehicular networks. In Proceedings of the ACM Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS), Number CONF, Vancouver, BC, Canada, 14 August 2007.
31. Sampigethaya, K.; Li, M.; Huang, L.; Poovendran, R. AMOEBA: Robust location privacy scheme for VANET. *IEEE J. Sel. Areas Commun.* **2007**, *25*, 1569–1589. [CrossRef]
32. Emara, K.; Woerndl, W.; Schlichter, J. CAPS: Context-aware privacy scheme for VANET safety applications. In Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks, New York, NY, USA, 22–26 June 2015; p. 21.
33. Lefevre, S.; Petit, J.; Bajcsy, R.; Laugier, C.; Kargl, F. Impact of v2x privacy strategies on intersection collision avoidance systems. In Proceedings of the 2013 IEEE Vehicular Networking Conference, Boston, MA, USA, 16–18 December 2013; pp. 71–78.
34. Humbert, M.; Manshaei, M.H.; Freudiger, J.; Hubaux, J.P. On the optimal placement of mix zones: A game-theoretic approach. In Proceedings of the 16th ACM Conference on Computer and Communications Security P. Citeseer, Chicago IL, USA, 9–13 November 2009; pp. 324–337.
35. Sun, Y.; Zhang, B.; Zhao, B.; Su, X.; Su, J. Mix-zones optimal deployment for protecting location privacy in VANET. *Peer- Netw. Appl.* **2015**, *8*, 1108–1121. [CrossRef]
36. Palanisamy, B.; Liu, L. Mobimix: Protecting location privacy with mix-zones over road networks. In Proceedings of the 2011 IEEE 27th International Conference on Data Engineering, Hannover, Germany, 11–16 April 2011; pp. 494–505.
37. Lu, R.; Lin, X.; Luan, T.H.; Liang, X.; Shen, X. Pseudonym changing at social spots: An effective strategy for location privacy in vanets. *IEEE Trans. Veh. Technol.* **2011**, *61*, 86–96. [CrossRef]
38. Boualouache, A.; Senouci, S.M.; Moussaoui, S. PRIVANET: An Efficient Pseudonym Changing and Management Framework for Vehicular Ad-Hoc Networks. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3209–3218. [CrossRef]
39. Khodaei, M.; Papadimitratos, P. Poster: Mix-Zones everywhere: A dynamic cooperative location privacy protection scheme. In Proceedings of the 2018 IEEE Vehicular Networking Conference (VNC), Taipei, Taiwan, 5–7 December 2018; pp. 1–2.
40. Song, J.H.; Wong, V.W.; Leung, V.C. Wireless location privacy protection in vehicular ad-hoc networks. *Mob. Netw. Appl.* **2010**, *15*, 160–171. [CrossRef]
41. Wang, S.; Yao, N.; Gong, N.; Gao, Z. A trigger-based pseudonym exchange scheme for location privacy preserving in VANETs. *Peer- Netw. Appl.* **2018**, *11*, 548–560. [CrossRef]
42. Mdee, A.P.; Saad, M.M.; Khan, M.; Khan, M.T.R.; Kim, D. Impacts of location-privacy preserving schemes on vehicular applications. *Veh. Commun.* **2022**, *36*, 100499. [CrossRef]
43. Al-Marshoud, M.S.; Al-Bayatti, A.H.; Kiraz, M.S. Improved Chaff-Based CMIX for Solving Location Privacy Issues in VANETs. *Electronics* **2021**, *10*, 1302. [CrossRef]
44. Zhang, S.; Li, M.; Liang, W.; Sandor, V.K.A.; Li, X. A Survey of Dummy-Based Location Privacy Protection Techniques for Location-Based Services. *Sensors* **2022**, *22*, 6141. [CrossRef]
45. Huang, R.; Wei, C.; Wang, B.; Yang, J.; Xu, X.; Wu, S.; Huang, S. Well performance prediction based on Long Short-Term Memory (LSTM) neural network. *J. Pet. Sci. Eng.* **2022**, *208*, 109686. [CrossRef]
46. Li, Y.; Yin, Y.; Chen, X.; Wan, J.; Jia, G.; Sha, K. A secure dynamic mix zone pseudonym changing scheme based on traffic context prediction. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 9492–9505. [CrossRef]
47. Zhang, Z.; Wong, W.C.; Sikdar, B. A Hybrid Reinforcement Learning-Based Method for Generating Privacy-Preserving Trajectories in Low-Density Traffic Environments. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 14740–14757. [CrossRef]
48. Reid, D. An algorithm for tracking multiple targets. *IEEE Trans. Autom. Control* **1979**, *24*, 843–854. [CrossRef]
49. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall PTR: Hoboken, NJ, USA, 1998.
50. Connor, J.T.; Martin, R.D.; Atlas, L.E. Recurrent neural networks and robust time series prediction. *IEEE Trans. Neural Netw.* **1994**, *5*, 240–254. [CrossRef] [PubMed]
51. Mozaffari, S.; Al-Jarrah, O.Y.; Dianati, M.; Jennings, P.; Mouzakitis, A. Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 33–47. [CrossRef]
52. Zyner, A.; Worrall, S.; Nebot, E. A recurrent neural network solution for predicting driver intention at unsignalized intersections. *IEEE Robot. Autom. Lett.* **2018**, *3*, 1759–1764. [CrossRef]
53. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

54. Zyner, A.; Worrall, S.; Nebot, E. Naturalistic driver intention and path prediction using recurrent neural networks. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1584–1594. [CrossRef]
55. Wang, J.; Li, X.; Li, J.; Sun, Q.; Wang, H. NGCU: A new RNN model for time-series data prediction. *Big Data Res.* **2022**, *27*, 100296. [CrossRef]
56. Emara, K. Poster: Prext: Privacy extension for veins vanet simulator. In Proceedings of the 2016 IEEE Vehicular Networking Conference (VNC), Columbus, OH, USA, 8–10 December 2016; pp. 1–2.