Novel Centralized Pseudonym Changing Scheme for Location Privacy in V2X Communication

Ahmed Didouh, Yassin El Hillali, Atika Rivenq, Houda Labiod

energies

MDPI

Article

Novel Centralized Pseudonym Changing Scheme for Location Privacy in V2X Communication

Ahmed Didouh [1,*], Yassin El Hillali [1], Atika Rivenq [1] and Houda Labiod [2]

[1] IEMN DOAE, UMR CNRS 8520, Polytechnic University Hauts-de-France, 59300 Valenciennes, France; Yassin.ElHillali@uphf.fr (Y.E.H.); Atika.Menhaj@uphf.fr (A.R.)
[2] INFRES Computing and Networks ENST, Telecom, 91120 Paris, France; houda.labiod@telecom-paris.fr
* Correspondence: ahmed.didouh@uphf.fr

Abstract: Vehicular ad hoc networks allow vehicles to share their information for the safety and efficiency of traffic purposes. However, information sharing can threaten the driver's privacy as it includes spatiotemporal information, and the messages are unencrypted and broadcasted periodically. Therefore, they cannot estimate their privacy level because it also depends on their surroundings. This article proposes a centralized adaptive pseudonym change scheme that permits the certificate's authority to adjust the pseudonyms assignment for each requesting vehicle. This scheme adapts dynamically depending on the density of the traffic environment and the user's privacy level, and it aims to solve the trade-off problem between wasting pseudonyms and Sybil attack. We employ a Knapsack problem-based algorithm for target tracking and an entropy-based method to measure each vehicle's privacy. In order to demonstrate the applicability of our framework, we use real-life data captured during the interoperability tests of the European project InterCor. According to the experimental results, the proposed scheme could easily estimate the level of confidentiality and, therefore, may best respond to the adaptation of the pseudonyms.

Keywords: privacy measurement in V2X wireless communications; adapted pseudonym change scheme; authorization authority backup

Citation: Didouh, A.; El Hillali, Y.; Rivenq, A.; Labiod, H. Novel Centralized Pseudonym Changing Scheme for Location Privacy in V2X Communication. *Energies* **2022**, *15*, 692. https://doi.org/10.3390/en15030692

Academic Editors: Fouzia Boukour Elbahhar, Luca De Nardis and Daniele D. Giusto

Received: 30 November 2021
Accepted: 1 January 2022
Published: 18 January 2022

1. Introduction

The vehicular transport sector is frequently affected by issues such as traffic congestion and accidents. It was thus essential to evolve a cooperative system between vehicles to minimize accidents and permit vehicles and road managers to share information freely. This new ecosystem uses different communication methods such as vehicle to vehicle (V2V), vehicle to infrastructure (V2I), and vehicle to anything (V2X).

Recently, technologies have provided communication models that can be used by vehicles in different application contexts. For example, the European Telecommunications Standards Institute (ETSI) has standardized the ITS-G5 standard, using the IEEE 802.11p standard. It is based on 10 MHz bandwidth channels in the 5.9 GHz band (5.850–5.925 GHz) [1]. ITS-G5 is a suitable standard for cooperative–intelligent transport systems (C-ITS) applications for the following reasons: low-latency communications; no infrastructure requirement; reliable communications; communications range 200–1000 m [2].

The main components in the V2X ecosystem are onboard units (OBUs), which operate in vehicles, and the roadside units (RSUs), which act as the infrastructure by broadcasting information in I2V mode. ITS-G5 technology enables vehicles to operate as an ad hoc network on a V2V mode without the need for RSU intervention [3].

Therefore, it is mandatory to secure these wireless communications to ensure that all technologies meet security requirements [4]. Furthermore, safety should be particularly considered in connected autonomous vehicles, where a vulnerable system component can be exploited to cause dangerous consequences, such as injury or even loss of life.

For these reasons, several types of security architectures linked to V2X have been proposed. The current V2X security architecture is based on a centralized architecture where

all vehicles are identified, authenticated, authorized, and connected through central cloud servers that use a public key infrastructure (PKI) [3]. It should ensure the following security requirements: *Trust* of the provision to ITS stations of certificates allows them to affirm their permission to use the ITS system and use specific ITS services and applications; Access control should be ensured by giving ITS stations cryptographically signed certificates of authorization, which allow them to use specific services or send specific information; Confidentiality of information transmitted in a unicast communication is protected by encryption of messages within an established security association; Privacy is based on the use of pseudonyms that can replace meaningful and traceable identifiers.

There is a compromise between the waste of certificates and the Sybil attack as explored by [5] since, on the side of the authority, we can not differentiate between the "honest" vehicles that only use the certificates excessively and the others that use the pool of pseudonyms to run Sybil Attacks. This contribution focuses on improving the privacy of V2X communications by proposing a dynamically adaptive system that allows certificate authorities to monitor the pseudonym-changing process. Our contribution allows the authorization authorities to anticipate users' needs in terms of confidentiality and to adapt the pool of pseudonyms to avoid both ends of the problem.

Our contributions in this paper can be summarized as follows:

- Propose a context-adaptive and authority-centric privacy scheme for VANET.
- Knapsack problem-based algorithm for the trajectories combinations and users traceability.
- Evaluate the privacy of real-life users based on data shared from OBUs developed by different countries (France, Germany, Holland, Norway, and Austria).

This article is organized as follows: Section 2 gives an overview of traditional V2X security and pseudonym change strategies. We also present the metrics of measuring the vehicles' privacy. Then, in Section 4, we describe our dynamic recommending schemes framework. Next, we present an experimental analysis of the proposed solution in Section 5, with some discussions in Section 6. Finally, Section 7 concludes the paper, presents valuable insights from our work, and introduces future work.

2. Related Work

2.1. Conventional Security Architecture

The security architecture for V2X is a public key infrastructure (PKI) adapted to the context of C-ITS. It is a hierarchical architecture composed of different authorities. The root certificate authority (RCA) acts at the top of the hierarchy of certificate authorities. It controls all the subordinate certification authorities and the final entities in its scale. A trusted certificate is provided to each last legitimate entity and may be revoked or blocked.

The C-ITS system is based on the provision of certificates and access control management [6]. The RCA manages the certificate of revocation list (CRL), and certificate trusted list (CTL). The RCA also manages two authorities: enrollment authority and the authorization authority.

Enrollment authority: This authority provides enrollment certificates to ITS-S such as RSUs and OBUs. Each node has a unique long-term identifier, an agreement between the car manufacturers and the authorities where each identity is associated with a pair of cryptographic keys and a set of node attributes. The attributes reflect the node's equipment's technical characteristics and its role in the system.

Authorization authority: This authority provides short-term certificates, also known as authorization tickets, to all ITS stations (OBUs, RSUs, . . .). The tickets are obtained based on key pairs generated by the OBU's HSM using its EC to authenticate with the AA. The AA signs each of the public keys and generates a set of pseudonym certificates (PC) for the station. Each PC contains information about the issuer CA as well as information specific to the OBU station.

According to the IEEE standard and European standard, ETSI [3,4,7], here is an overview of some functions that the C-ITS system offers:

- Secures the private keys corresponding to public keys via the hardware and software security modules implemented in OBUs.
- Logging actions (in centralized archives).
- Archiving certificates over time.
- Misbehavior detection and certificate revocation.

The global architecture is operated under the Security Credential Management System (SCMS) proposition explained by [8]. In addition to the certificates authorities, two more entities ensure the unlinkability of vehicles' identities and ensure their privacy:

Linkage authority (LA): Generates pre-linkage values, forming linkage values in the certificates and supporting efficient revocation. There are two LAs in the SCMS, referred to as LA1 and LA2. The splitting prevents the operator of an LA from linking certificates belonging to a particular device. In Figure 1, the linkage process to obtain the misbehavior identity is shown.

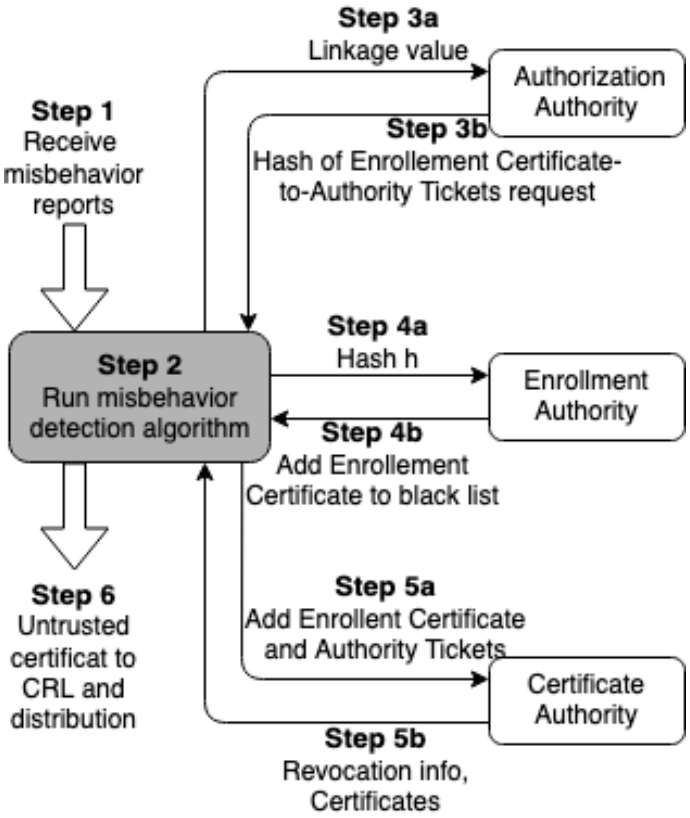

Figure 1. Reports of OBUs to misbehavior authority of malicious vehicles and the authorities' conducted process to the linkage of authorization tickets and their corresponding ECs and attributes to report them into the CRL.

Location obscurer proxy (LOP): Hides the location of the requesting device by changing source addresses, thus preventing linking network addresses to locations.

2.2. Identifiers

There are many different addresses, IDs, or other identifying information scattered around the network layers.

- *GeoNetworking:* Each GN node is identified by [9], containing information about the ITS-S type (passenger car, cyclist, pedestrian, RSU, . . .) and 48 bits derived from the link-layer address. In the case of a pseudonym change, only the latter part is supposed to change. GN packets have a basic, a common, and an optional extended header. The basic header contains information such as the packet's maximum lifetime and the remaining hop limit. This information is noncritical for identification.
- *Facilities layer:* The facilities layer introduces a StationID, an integer identifying the ITS system. The standard document already mentions that this ID may be a pseudonym.
- *IPv6:* While each IPv6-capable network interface can have multiple addresses, it has at least one link-local address with the interface ID (the lower 64 bits) uniquely derived from its data-link layer address. The mapping of the IPv6 link-local address and GNADDR is straightforward, as both addresses are deterministically derived from the same link-layer address. Additionally to the IPv6 address, the IPv6 header can also contain a flow label which could lead to partial linkability of packets even after an address change: Although a flow shall be identified by the triplet of the flow label, source, and destination address, an equal flow label could indicate the resumption of a connection even after an address change.

2.3. Pseudonyms Changes Strategies

Pseudonym certificates are stored and managed in pseudonym pools, with their corresponding private keys kept in the HSMs. To keep the privacy of vehicles and avoid tracking or linking their real identities to the used pseudonym certificates (PCs), the authorization tickets are changed frequently according to various rules [10]. This ensures that each VC has precisely one key pair (own pseudonym and private key) active during each period. VCs cannot reuse the pseudonym once it has been changed, even if the PKI certificate has not yet expired.

The ETSI standard on trust and privacy management [6] mentions the goal of pseudonymity and unlinkability of ITS nodes and their messages as the way to achieve ITS privacy. This privacy goal is subdivided into two dimensions: The privacy of ITS registration and authorization shall be achieved by limiting the knowledge of a node's canonical (fixed) identifier to a limited number of authorities. Furthermore, the responsibility for verifying the validity of a canonical identifier is given to an enrollment authority (EA) and split from the authorization to services by the authorization authority (AA). These authorities are parts of the needed public key infrastructure (PKI) and need to be operated in different control areas to achieve a surplus of privacy. During manufacture, the following data is to be stored in an ITS node using a physically secure process:

- A globally unique canonical identifier.
- Contact addresses + public keys of an EA and AA.
- A set of trusted EA and AA certificates.

There needs to be some ambiguity regarding which node changed to which pseudonym, there shall be other nodes present within the reception range, coordination, and frequency of change matter, and all identifiers need to be changed simultaneously with buffers being flushed or discarded. Finally, control metadata, such as sequence numbers in GN packets, have to be reset as well.

The ETSI, ITS working group, gathers several concepts for pseudonym change strategies (PCS) in a technical report [10]: The parameters deciding a PCS (e.g., period or length) shall be randomized to prevent linkability by analyzing the periodicity of changes. After changing pseudonyms, random-length silent periods shall be abided, in which nodes stop sending any packages. When using a vehicle-centric strategy, pseudonym change time, frequency, and duration of silent periods are influenced by the vehicle's mobility and trajectory to make linking pseudonyms based on broadcasted movement parameters harder. In the

density-based approach, pseudonyms are changed only if enough other vehicles are around to avoid unnecessary unambiguous pseudonym changes. Mix-zones are geographical areas where no messages of location-aware services are exchanged. This concept is supposed to make the linkage of ingoing and outgoing vehicles from the zone difficult. These zones are especially effective in high-density and high-fluctuation areas such as intersections or parking spots. Vehicles could collaboratively change pseudonyms within these zones by announcing them via broadcast messages and then changing synchronously.

However, as stated in the report, the efficiency of that approach depends heavily on the density of the situation. A particular variant is cryptographic mix-zones: Within these zones with a size limited to the radio coverage of an RSU, no identifying data is sent in plain text, but everything is encrypted with the same symmetric key provided by the RSU. Thus, it allows the usage of location-aware collision detection messages while preventing an outsider from eavesdropping without switching off essential safety features. An alternative to just changing from one pseudonym to the next from a node's internal storage is swapping pseudonyms randomly between nearby vehicles. We find this approach to be limited, though, by the inclusion of vehicle-specific data into messages and legal requirements demanding the possibility of an identity resolution for law enforcement.

The ETSI survey [10] also gives an overview of used strategies in existing standards or projects. These include some interesting further approaches: The SCOOP project proposes a timeslot-based, round-robin pseudonym selection. The exciting thing about this is that using pseudonyms from the local pool is explicitly allowed, as the selection mechanism ensures they are not always reused in the same order. This is a practical approach against the problem of pseudonym refill (acquiring new pseudonyms) not always being possible.

The strategy proposed by the car-2-car communication consortium is dividing each trip into at least three segments: The first one from the start of the trip to a middle segment, the middle segment being familiar to several people and unassociated with specific origins and destinations, and the last segment to the intended destination of the trip. This shall achieve that locations significant to a user can neither be linked together, nor the user, thus preventing individual movement profiles. Some safety requirements of the ETSI standard affect pseudonym change: In critical situations when a receiving station would need to take immediate action in response to received safety information, pseudonyms have to be locked. The reason behind that is that cooperation collision avoidance depends on all vehicles broadcasting their location and trajectory.

2.4. Tracking Attacks

We set the "unlinkability" as the concept that the greater the distance in time and space between two transmissions from the same device, the harder it is to determine that those two transmissions did come from the same device. Accordingly, vehicles in a silent period due to a pseudonym change would not be considered, and vehicles changing pseudonyms without a silent period could appear as duplicate or ghosting vehicles hindering collision evasion. Furthermore, recognizing such critical situations and initiating the pseudonym locking is performed by the receiving ITS vehicle, which decreases the risk of an attacker trying to lock pseudonyms without a critical situation being present deliberately, as shown in Figure 2.

Therefore, there is a real challenge in the trade-off between road safety and cybersecurity. Full message encryption does not meet temporary road safety requirements. The message linked to crash information must be accepted before the event to avoid a crash.

In the linkability process, the crucial metric is the period of silence after each pseudonym change, as shown in Figure 2, depending on the disseminated messages, as vehicles could not be silent in a period of TTC.

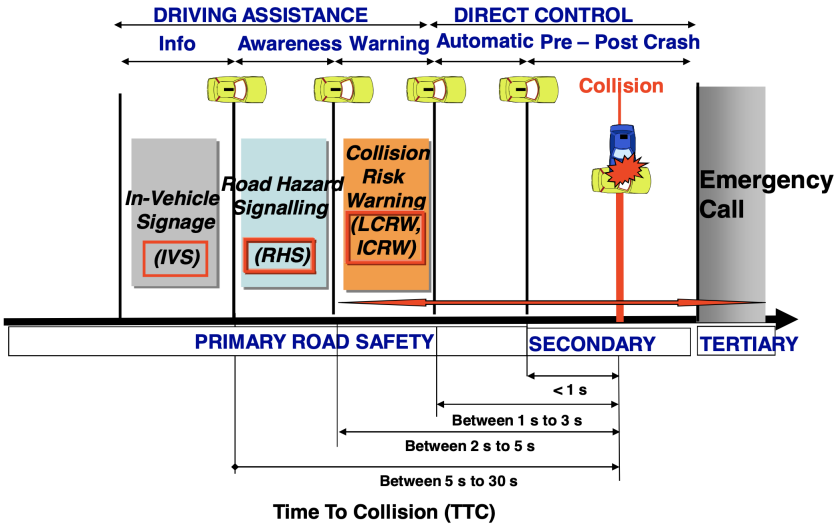

Figure 2. Applications being served by transmission showing the time to collision [11].

2.5. Pseudonym Privacy

Security architecture is a security design. It addresses the necessities. Moreover, potential risks are involved in a specific environment and when and where to apply security controls. Standards provide detailed requirements on how a policy must be implemented. In VANET, many groups [5,8] have presented the credential security architecture. The privacy and linkability of pseudonyms are essential issues in V2X communications. Researchers have contributed to resolving several issues for linkability. For example, Rebollo-Monedero et al. [12] suggested a trusted third-party system where privacy depends on collaboration among multiple untrusted users. This solution is related to a situation where the service provider is not trusted. In this way, the untrusted service provider will be unable to access the privacy information of any user. Yao et al. [13] proposed a novel lightweight, secure, and privacy-preserving pseudonym changing scheme and proposed an asynchronous key agreement scheme.

3. Problem Formulation

There is a significant trade-off in the changing pseudonyms scheme. The certificates must be changed periodically for privacy reasons. Although one option is to have many certificates, each valid one after the other for a short period, this would result in many unused certificates, leading to the waste of certificates, and could even be used to operate the Sybil attack. In addition, the authorities should revoke misbehaving or malfunctioning devices, but placing all valid device certificates on the CRL would make it very large. This paper aims to dynamically adapt the number of PCs given by the authorization authorities to each vehicle. This should help to regulate all these problems related to the pool of PCs given by the AA.

4. Machine Learning-Based Framework

4.1. Attacker Model

The confidentiality level of an individual's location is always relative to the control of an attacker trying to follow a person in the network.

In this article, we assume a passive attacker can listen to all messages sent over the network. Thus, what the attacker can gain from observing transmissions in the network is to trace the identity of the drivers.

The assumption of the attack model used is based on the attacker's strong ability to link an identity to a vehicle MAC address at the beginning of the node's lifetime. The individual remains anonymous when the departure has not been linked to an origin/destination pair.

The modeling of an attacker is linked to the tracking algorithm. Therefore, the learning of the attacker is highly dependent on the mobility used and the pseudonym-changing

strategies used by the driver. If, for example, nodes do not change pseudonyms, or drive in a very predictable way, the tracking algorithms will work much better.

Therefore for our calculations, we choose to use a probabilistic attacker model: Attacker strength is defined as the probability with which an attacker can follow a nickname exchange between two nodes. The entropy H for an attacker who cannot follow a pseudonym exchange for each individual in the network would then be zero.

The force attacker also affects the increased privacy level when a new location in the nickname pool becomes active, i.e., when all nodes start using new nicknames. If we assume that two nodes very close to each other could confuse an attacker by exchanging their nicknames (the extent being dependent on its strength), that attacker will also be confused when these two nodes simultaneously change nicknames. From this, it follows that the level of confusion is based on the number of candidates directly neighboring the node.

4.2. System Model

Our system model is based on the network architecture proposed by the European committee [14], as illustrated in Figure 3. This architecture is peculiar in that the national node is linked to all the users who operate in cellular technology and the road manager, allowing him to receive the messages broadcast in ITS-G5. Furthermore, this configuration will allow certificate authorities to receive messages circulating in the network.

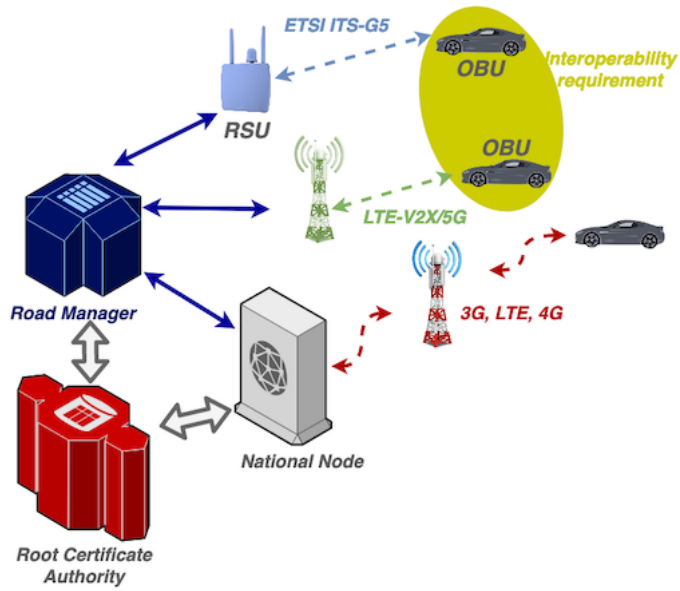

Figure 3. Network setup.

Our solution aims to set up a scheme of pseudonym changing dynamically. Our system's first actors are the authorization authority, as it is the responsible entity for providing the pseudonyms to the vehicles. With this framework, it should optimize the size of PC pools provided to VCs. In our proposition, we assume full connectivity between vehicles and the authorities. This proposed framework could be used as a backup to the conventional solutions to optimize resources, and it also helps to avoid some attacks, such as the Sybil attack.

In order to give an adapted proposition of changing scheme, our solution is to be placed in the shoes of the attacker by trying to track vehicles, and this calculates their entropy (privacy metric explained in Section 4.4) and gives a global PCs changes scheme.

4.3. Tracking Algorithm

The attacker is assumed to have access to all transiting messages in the network. Thus, our algorithm computes several informative characteristics of each communicating node to relate each MAC address to an origin/destination pair.

Our algorithm permits us to solve our problem in the form of the knapsack problem. It has all vehicles' messages of a specific region as input and also a couple of O/D pairs.

The optimization target is to attribute each MAC address m to an O/D pair. As shown in Figure 4, the output of our algorithm is the probabilities of m MAC address to carry out the corresponding O/D trajectory.

We determine the best candidate for each O/D pair in real time, as vehicles keep changing their pseudonyms and MAC addresses. Moreover, this algorithm permits to solve just a first step of the tracking problem, as it is based on the MAC address as an identity.

We formulate our knapsack problem using the well-studied multiple multidimensional knapsack problem (MMKP) [15,16].

The weights w_{ij}^k correspond to the distance of each vehicle's trajectory to go to each destination pair, and the profits p_{ij}^k correspond to the probability of the set of trajectories corresponding to different MAC addresses to do the O/D pair k. In this problem, we want to maximize the combination of the probabilities of several paths corresponding to different MAC addresses. Respecting the capacity of each O/D pair,

$$\begin{cases} Maximize \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}^k x_{ij}^k, for\ k = 1, \ldots, s \\ Subject\ to \sum_{j=1}^{n} w_{ij}^k x_{ij}^k \leq c_k, for\ i = 1, \ldots, m, and\ j = 1, \ldots, n \\ \prod_{p}^{q} x_{ij} = 1, for\ i = 1,2, \ldots, m \end{cases} \quad (1)$$

x_{ij} : Set of trajectories.
w_{ij} : The weight of the j^{th} trajectory corresponds to k^{th} O/D pair.
p_{ij} : The profit of the i^{th} trajectory in the j^{th} MAC address in terms of probability.
c_j : The capacity constraint of every k^{th} combination to correspond to the right O/D pair.

We first calculate the matched combination to the O/D pair and then calculate each combination probability using Algorithm 1. As shown in Figure 4, the algorithm aims to minimize the gap between every identity origin/destination ($P_s(i)/P_e(i)$) and the O/D pair.

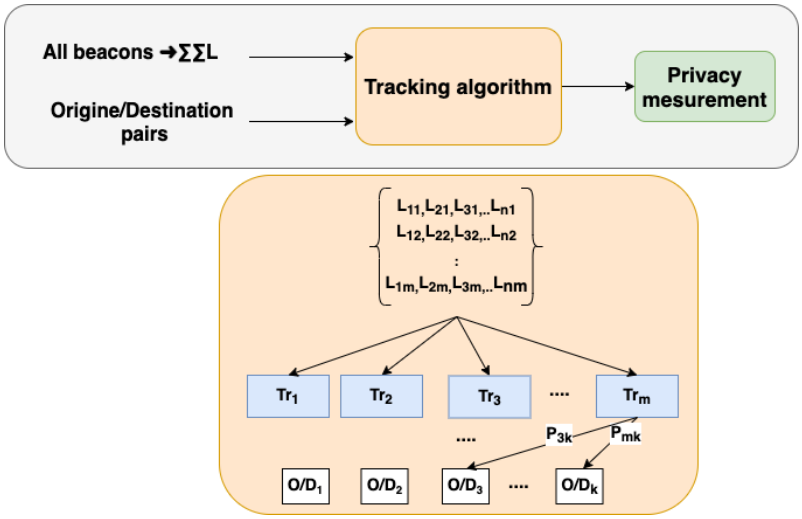

Figure 4. Tracking algorithm steps.

The output of this algorithm is the Matrix E, given as the following:

$$E = \begin{pmatrix} Tr_1 \\ Tr_2 \\ \vdots \\ Tr_n \end{pmatrix} \begin{pmatrix} ID_{MAC}(3) & ID_{MAC}(1) & 0 & 0 & \cdots \\ ID_{MAC}(4) & ID_{MAC}(5) & ID_{MAC}(7) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ ID_{MAC}(12) & ID_{MAC}(2) & ID_{MAC}(9) & 0 & \cdots \end{pmatrix} \tag{2}$$

We then calculate the gap between IDs (ID_{MAC}) in each Tr. These gaps could be considered the period of silence used by vehicles to transit from one pseudonym to another. The silence period could be estimated by estimating the number of disseminated security messages: as seen in Section 2.4 and Figure 2, the silence period is linked to the TTC period as the OBU could not change the pseudonym or make a silence period in TTC. The dissemination of C-ITS security messages in each geographical zone depends on the C-ITS transmission range (R) and, therefore, nodes' interdistances. We use the truncated exponential distribution to estimate the number of vehicles with interdistances $0 < X_R < R$ in a given segment:

$$E[X_R] = E[x|x < R] = \frac{\int_0^R \mu x e^{-\mu x} dx}{1 - e^{-\mu R}} \times \frac{1}{\phi} = \frac{1 - e^{-\mu R}(\mu R + 1)}{\mu(1 - e^{-\mu x})} \times \frac{1}{\phi} \tag{3}$$

where μ is the interdistance distribution parameter, and ϕ is the ratio of security messages upon all disseminated messages.

The probability of silent period is given by

$$\delta_s = argminPr(E[X_R]) \tag{4}$$

Algorithm 1: Algorithm of community construction.

Input: $ID_{MAC}[]; O/Dpair$
Output: H
Function `KnapSack linking(`$ID_{MAC}[]; O/D\ pair$`):`
  s← size(ID_{MAC});
  while $i > S$ **do**
    VAR← ID_{MAC}(i); $ID_{MAC} \leftarrow \forall ID_{MAC} \setminus \{ID_{MAC}(i)\}; i \leftarrow i - 1;$
    $P_s \leftarrow Pos_{start}(VAR); P_e \leftarrow Pos_{end}(VAR); D_{start/O} \leftarrow distance(P_s, O);$
    $D_{end/D} \leftarrow distance(P_e, D);$
    if $D_{start/O} > 0.1km$ **then**
      for j<S **do**
        $dis \leftarrow distance(P_s, Pos_{end}(ID_{MAC}(j)))$ **if** `dis <` $D_{start/O}$ **then**
          $ID_{MAC} \leftarrow \forall ID_{MAC} \setminus \{ID_{MAC}(j)\};$
          $ID_{MAC} \leftarrow add(ID_{MAC} + VAR);$
        end
      end
    end
    if $D_{end/D}$ `> 0.1 km` **then**
      for j<S **do**
        $dis \leftarrow distance(P_e, Pos_{start}(ID_{MAC}(j)))$ **if** $dis < D_{end/D}$ **then**
          $ID_{MAC} \leftarrow \forall ID_{MAC} \setminus \{ID_{MAC}(j)\};$
          $ID_{MAC} \leftarrow add(VAR + ID_{MAC});$
        end
      end
    end
    if $D_{start/O} > 0.1km\ and\ D_{end/D} > 0.1km$ **then**
      $E \leftarrow add(VAR)$
    end
  end
  return NS_i, n_i
End Function

4.4. The Measurement Model

For the metric that is used to quantify location privacy in V2X systems, the level of privacy is quantified based on the uncertainty about that user. In [17,18], they introduced the method calculation of the privacy metric based on the entropy of exchanged information. In this second part of our framework, we use the results of our knapsack algorithm as input to calculate the privacy of each vehicle.

We calculate the confidentiality of the geographical position of each person. In order to prove the traceability of a vehicle, it is necessary to ensure that the person corresponds to the vehicle which served the O/D (origin/destination) pair.

We give the mathematical model inspired from [18], and we can model the vehicular communications as a weighted directed graph $G = (V, E, p)$.

G has several unique properties. G contains all information relative to its trajectory, and vertices in G are connected with directed edges. The probability distributions on the edges model depend on the adversary's knowledge of the users and their movements in the system from the previous algorithm. Moreover, the sum of the probabilities on outgoing edges from a vertex is defined $o \in O$ or $d \in D$ to be 1, $\sum_{k=1}^{m} p(i_j, o_k) = 1$, $\sum_{k=1}^{m} p(o_j, d_k) = 1$,$\sum_{k=1}^{n} p(d_j, i_k) = 1$.

In order to determine the probability distributions and quantify the privacy in the measurement model, we use the information entropy developed by Shannon [19]. We extract the entropy based on the probability distribution, which represents the quantitative measure of information content and uncertainty. Entropy has been accepted as an applicable measure in the privacy research community [18,20,21]. However, the main challenge here is to rely on the entropy calculation to give an optimal pattern of change of pseudonyms. By definition, for a probability distribution with values $p_1, \cdots, p_n$, the entropy is

$$H = -\sum p_i log(p_i)$$

where p_i is the ith element of the probability distribution, and H is the balance of information measure and uncertainty related to the probability distribution. High entropy means an increase in uncertainty and, therefore, a higher level of privacy. The entropy is maximal if the probability values are equal. In order to calculate entropy, we are interested in the source of the information that the adversary captures. For example, we are interested in information linking individuals to their geographical movements to determine who moves from where to where.

For nonzero probabilities, the computation of entropy for $p_i = 0$ means that there is no uncertainty and that the sum of the probability distribution must be equal to 1. Therefore, we compute the entropy for a specific individual as

$$H(i_s) = -\sum_{j=1}^{m}\sum_{k=1}^{m} \hat{p}_{jk} log(\hat{p}_{jk}) \tag{5}$$

where $\hat{p}_{jk}$ is the probability of traveling from o_j to d_k.

The value of $p\hat{}jk$ is given as

$$\hat{p}_{jk} = \frac{p(i_s, o_j)p(o_j, d_k)p(d_k, i_s)}{\sum_{j=1}^{m}\sum_{k=1}^{m} p(i_s, o_j)p(o_j, d_k)p(d_k, i_s)} \tag{6}$$

The maximum entropy for an identity depends on the number of possible trajectories.

4.5. Dynamic Pseudonym Change

After identifying the level of privacy of each vehicle, the authorization authority proceeds to the clustering model (K-Means or others) based on vehicle information and the results obtained by the previous algorithm. The AA classifies vehicles into three categories, as shown in Figure 5: these categories represent vehicles in a definite range of privacy levels. Therefore, the AA will adapt the pseudonym-changing scheme proposal and the

number of PCs in the pools. The latter could be personalized for each vehicle, depending on the route it usually takes.

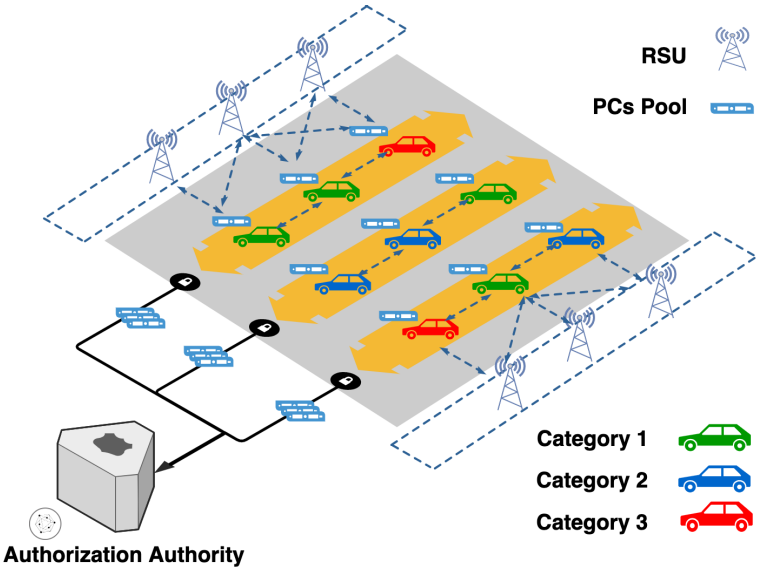

Figure 5. The authorization authority adapts the pseudonym pools sent to each privacy category.

5. Performance Evaluation

We tested the performance of our solution via data collected from real-life tests in the European project InterCor [22]. We analyzed the raw data using Wireshark. We implemented and tested our algorithm using the Matlab tool.

5.1. Mobility Model

This scenario is based on the actual data obtained during the TestFest in Holland. Using a sniffer, we captured the messages sent by all the surrounding vehicles in addition to PCAP files received from the other participants. Using this, we performed reverse engineering on the identity of each vehicle. Finally, we applied our solution to identify each vehicle and calculate its privacy level. These tests aim to test interoperability between the European partners. For all the test cases, vehicles have the same trajectory using one origin/destination pair. The test site corresponds to the start and arrival points.

5.2. Data Analysis

In Figure 6, we illustrate all the sniffed MAC addresses in their locations. All figures show the positions of each captured MAC address, and each of the five figures represents half a day of tests. We notice that *Test 2* represented the peak of the participation of tester vehicles, as we received a more significant number for MAC address.

In Table 1, we detail one of the first captions tests. The table gives information about the first day of tests. All the information given in this table is based on the received messages. We calculated the distance traveled and the distance between origin (travel start point) and destination (travel endpoint). We also give the different StationIDs used during the travel and the type of messages sent. The IVI message is sent only by RSUs.

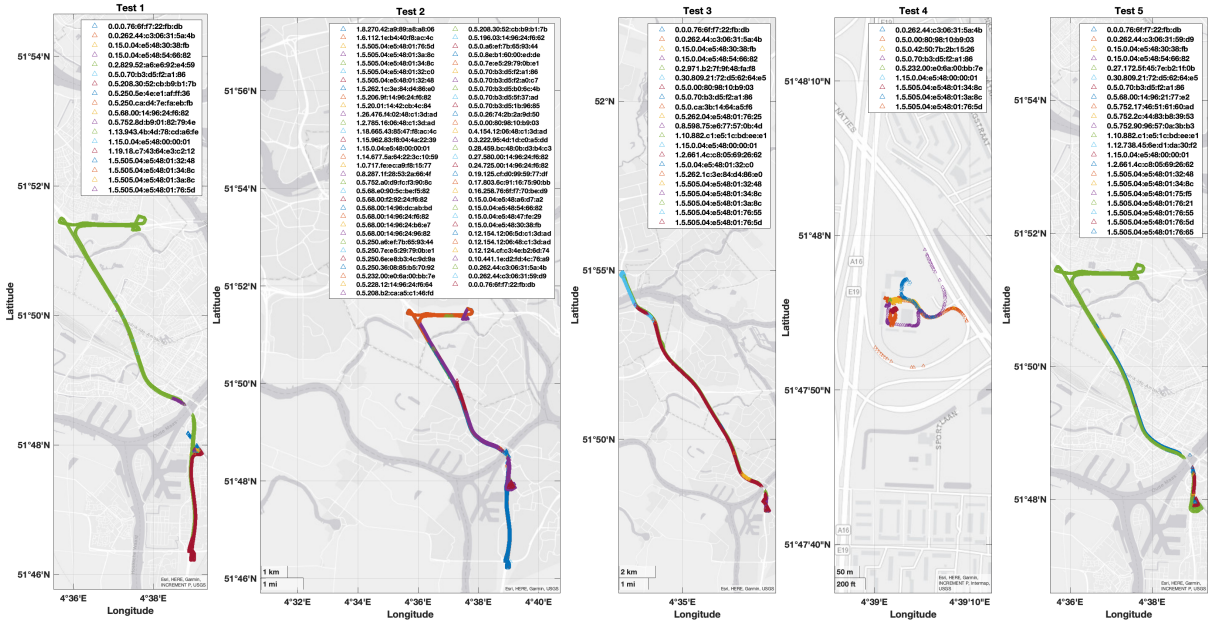

Figure 6. Applications being served by transmission showing the time to collision.

Table 1. Test 1 details of the analyzed data from Wireshark tool.

Test 1						
Adresse Mac	**Distance Orig/Dist (m)**	**Global Distance (m)**	**Nbr**	**StationID**	**Nbr Messages**	**Messages**
0.0.0.76:6f:f7:22:fb:db	94.5262534395616	5184.53670513403	2	9819	56	CAM
				1018	2	DENM
0.0.262.44:c3:06:31:5a:4b	41.10	4,716.54	1	103897675	102	CAM
0.15.0.04:e5:48:30:38:fb	0	19,600.66		1003	952	IVI, DENM,
0.15.0.04:e5:48:54:66:82	0	0	2	10031004	402	IVI, DENM,
0.2.829.52:a6:e6:92:e4:59	51.21	57,523.99	1	3693631938	2109	CAM
0.5.0.70:b3:d5:f2:a1:86	124.13	32,463.03	1	168084	1025	CAM
0.5.208.30:52:cb:b9:b1:7b	61.90	2,068.51	1	1	180	CAM
0.5.250.5e:4e:e1:af:ff:36	152.12	14,238.65	1	10127	6793	CAM
0.5.250.ca:d4:7e:fa:eb:fb	159.74	2,519.89	1	10127	2355	CAM
0.5.68.00:14:96:24:f6:82	75.56	23,408.76	1	2519004802	424	CAM
0.5.752.8d:b9:01:82:79:4e	77.13	2,825.36	2	8666661. 8666662	92	CAM
1.13.943.4b:4d:78:cd:a6:fe	50.95	402,364.18	1	81449815	6406	CAM
1.15.0.04:e5:48:00:00:01	0	22,262.54	3	1018. 1015. 1014	3150	IVI, DENM,
1.19.18.c7:43:64:e3:c2:12	129.12	149,748.17	1	3843896860	2240	CAM
1.5.505.04:e5:48:01:32:48	34.70	14,558.60	1	302050072	276	CAM
1.5.505.04:e5:48:01:34:8c	96.26	2,627.46	1	302052140	107	CAM
1.5.505.04:e5:48:01:3a:8c	79.29	2,508.76	1	302058140	26	CAM
1.5.505.04:e5:48:01:76:5d	45.98	12,985.55	1	302118093	283	CAM

In Figure 7, each box represents the variation of steps distance between all received messages from each MAC address in the first session of tests. This metric is very useful for our tracking algorithm.

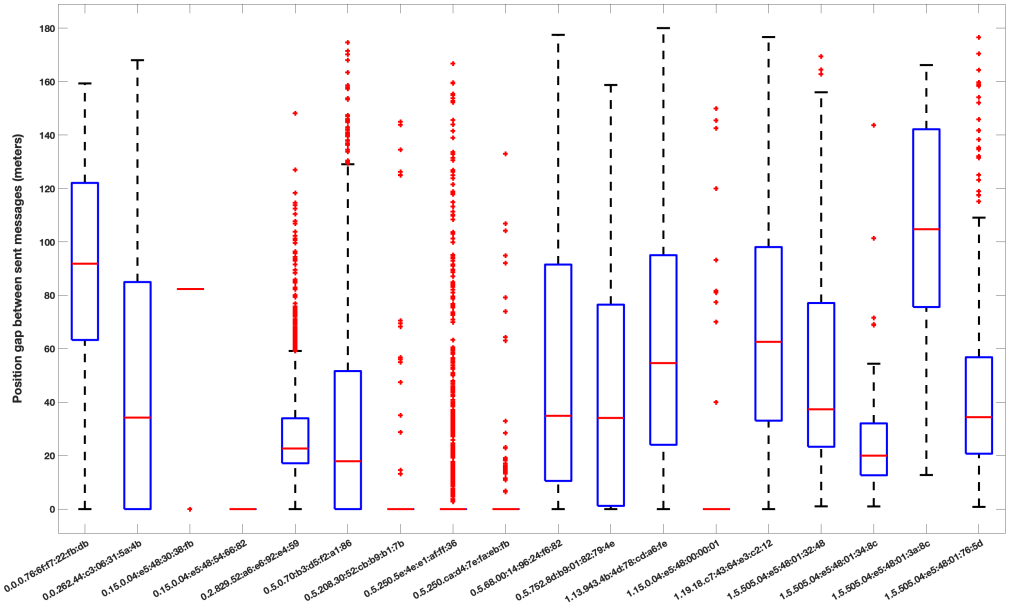

Figure 7. Representation of all steps distance between the received messages in the first tests.

In order to apply our algorithm, we took the second set of data (Test 2) as a case study. Our algorithm analyzed all cases based on the different metrics and information in Table 1. We calculated their probabilities and their privacy entropy in order to estimate the identitie,s as seen in Section 4. This analysis gave place to the three clusters. All the explanations are based on the assumptions of the attacker model in Section 4.1.

Cluster 1: It is a trivial case for an attacker because even with changing the pseudonym certificate and the StationID, the attacker could quickly identify users using the same MAC address for all their journeys. Figure 8 shows two cases from this cluster.

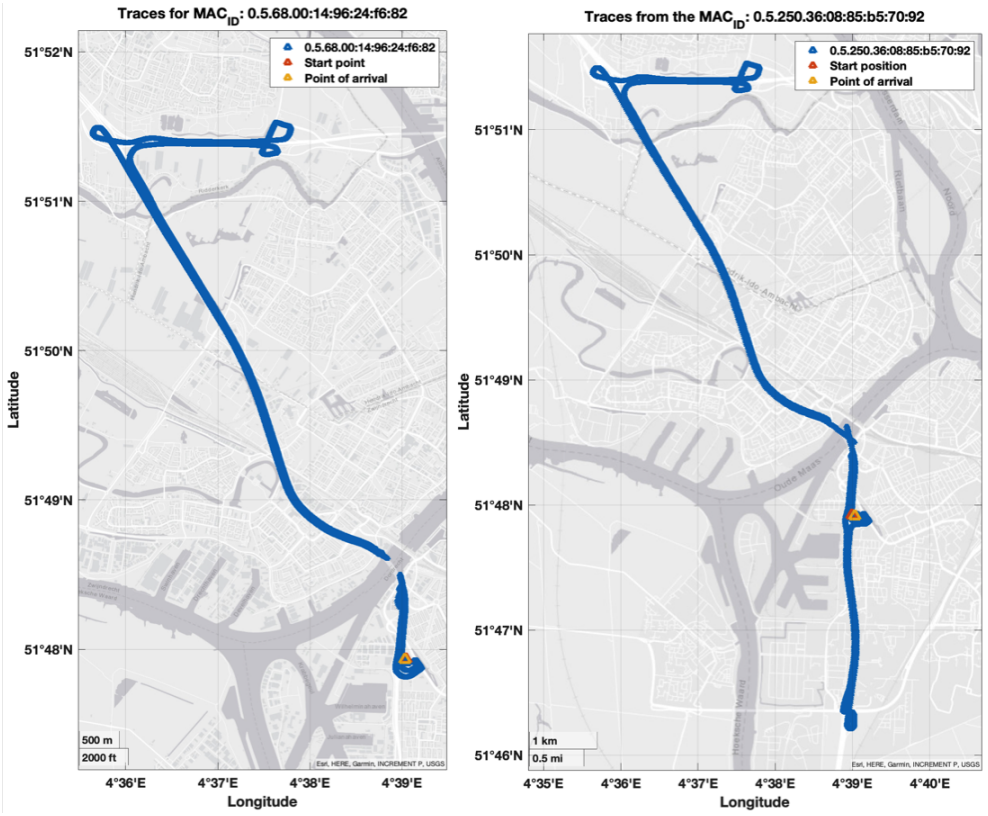

Figure 8. Cluster 1: high vulnerability.

Cluster 2: In this case, our algorithm could successfully link two different MAC addresses to a single identity that could have carried out the O/D trajectory. As there is a period of silence in the changing pseudonym strategies, it decreases the truth's probability. In Figure 9, we give indications of different assumed steps that the OBU could have carried out: (1) is the starting point of the driver's (i_x) journey; (2) is the point that i_x decided to change its pseudonym; (3) represents the silent period; (4) is the starting point with the new ID_{MAC} which ended in the point of arrival (destination D).

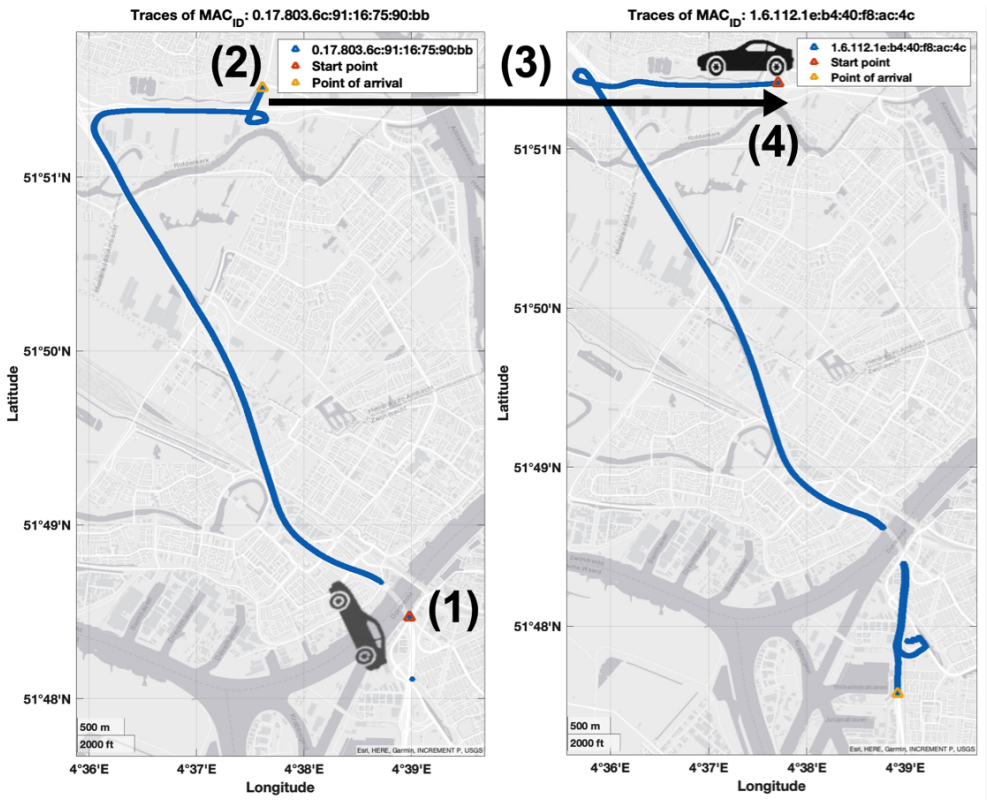

Figure 9. Cluster 2: medium vulnerability.

Cluster 3:This case is considered as the more secured case that could not be identified or linked. In Figure 10, our algorithm could not link the MAC addresses, which means that the users have different pseudonym-changing strategies.

Figure 11 shows the results of clustering of all the MAC addresses captured for the five tests according to several criteria taken into consideration by our algorithm to classify the privacy.

In Figure 12, we illustrate the ROC diagram of our algorithm performance in terms of precision.

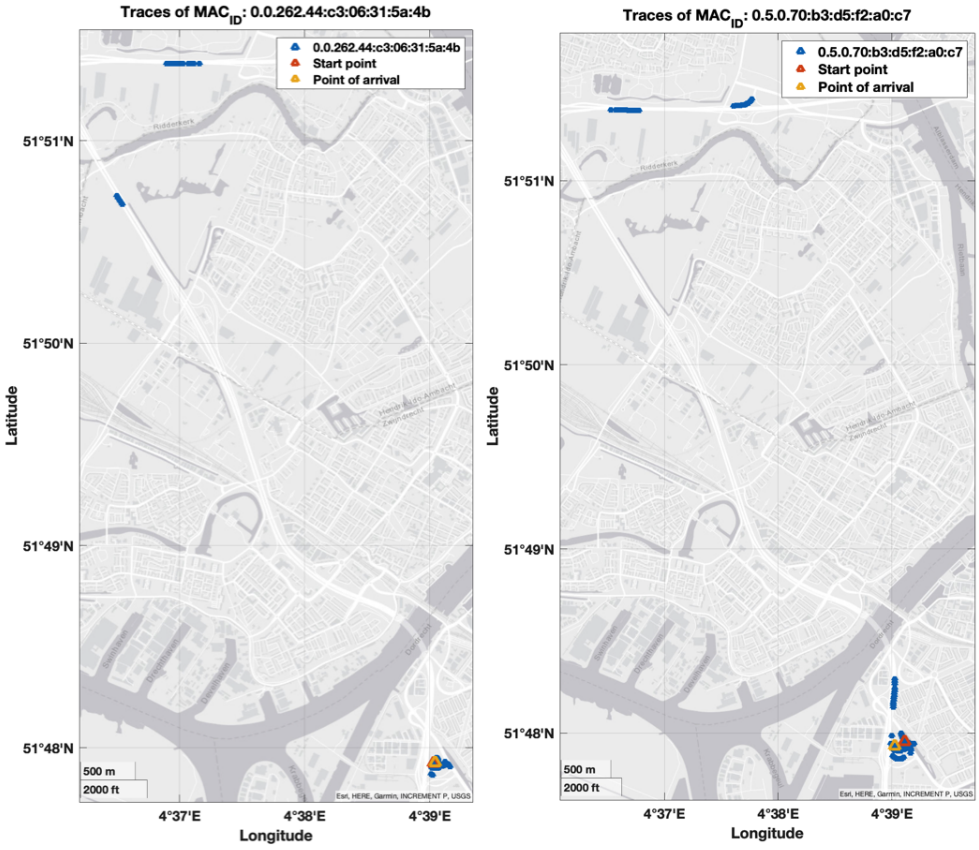

Figure 10. Cluster 3: secured.

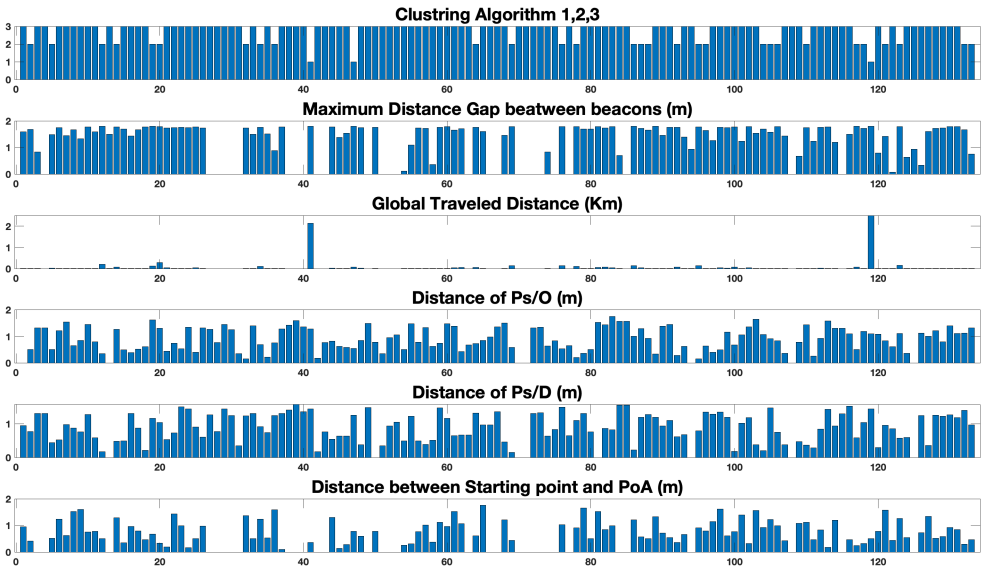

Figure 11. Clustering indication based on vehicle's journey parameters.

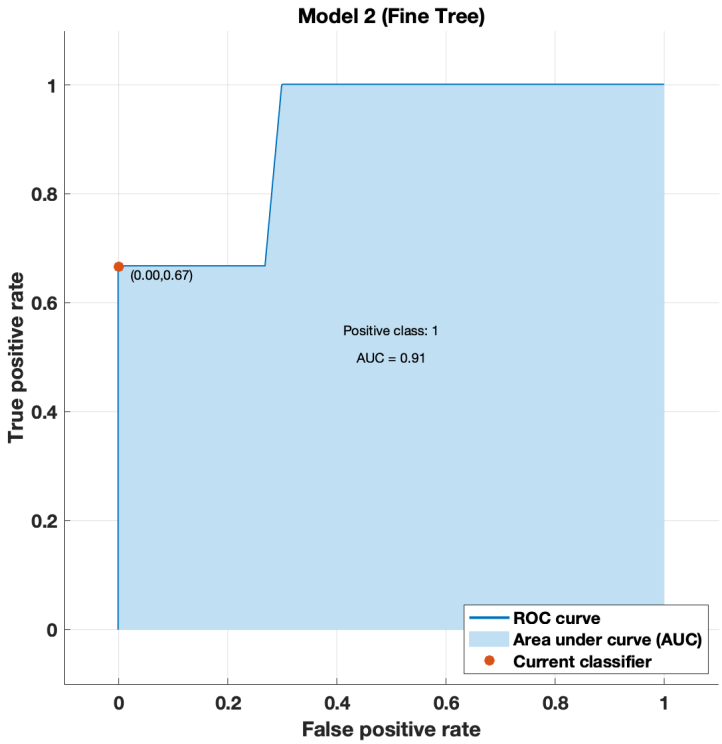

Figure 12. ROC curve of processed data.

6. Discussion

The authorization authority needs different information on the identity of the vehicles and the common routes for a fleet of vehicles to be able to compare the O/D pairs with the identities. We notice that in our case of tests, c is trivial given that the only O/D pair that was possible is the departure from the test site and the arrival on this same site. The AA will have direct access to messages circulating in the network via its link with the national node.

The clustering process shows precisely the privacy level of all users. The three categories represent the existing configurations well. Nevertheless, this framework is flexible and could be used with more categories to classify better. After the classification, the AA should propose an alternative PCS; nevertheless, this process should be nondeterministic. Therefore, an unsupervised machine learning model should fit perfectly into the framework. This framework could perfectly guard against tracking attacks as the attacker carries out the same process we underwent during these experiments. They stand on the listening mode to receive all the messages through the network and try to detect the identity of each MAC address which passes or at least tracks a particular identity.

7. Conclusions and Future Work

This paper applies an algorithm to users' privacy verification. We summarize three different categories of users' privacy. Thus, a formal verification framework for privacy is established. Based on this framework, AA could propose an adapted PCS. This contribution could help to resolve three major issues of the PKI system: 1. It allows to hollow out the wasting certificate problem; 2. The waste of certificates can lead to their use in Sybil attacks; 3. Reducing the CRLs size.

This work shows solid results and is the first algorithm applied to real-life data to estimate their privacy level. Furthermore, these results represent the most common cases in real life, as the tests were carried out with all the European participants. Thus, we can interpolate the results in all cases. This demonstrates that our framework is real-world applicable.

In the future, we will complete the framework with an unsupervised ML model to propose a PCS. We will improve the verification model with more real-life data. Our goal is to adapt the framework to all types of PCS. In addition, we plan to develop a decentralized manner to collaborate with the certificates authority. It will be a meaningful exploration and attempt in the field of V2X communication privacy.

Author Contributions: Conceptualization, A.D. and H.L.; methodology, A.D., Y.E.H. and H.L.; software, A.D.; validation, H.L. and A.R.; formal analysis, A.D., Y.E.H. and H.L.; investigation, A.D. and H.L.; resources, Y.E.H.; data curation, A.D.; writing—original draft preparation, A.D.; writing—review and editing, A.D., H.L., Y.E.H. and A.R.; visualization, A.D.; supervision, H.L. and A.R.; project administration, H.L. and A.R.; funding acquisition, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This work is partially supported by the DIR Nord (Road operator of the north of France) and supported by the EU project InDiD (Infrastructure Digitale de Demain) co-financed by the connecting Europe facility the European Union.

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

Abbreviations

The following abbreviations are used in this manuscript:

PCS	pseudonym change strategies
AA	Authorization Authorities
OBUs	On-Board Units
RSUs	Road Side Units
PKI	Public Key Infrastructure
CRL	Certificate of Revocation List
RCA	Root Certificate Authority
V2V	Vehicle to Vehicle
V2I	Vehicle to Infrastructure
V2X	Vehicle to Anything
ETSI	European Telecommunications Standards Institute
C-ITS	Cooperative- Intelligent Transport Systems

References

1. *302 663 (V1.2.1) ETSI Standard*; Access Layer Specification for Intelligent Transport Systems Operating in the 5 GHz Frequency Band. ETSI: Sophia Antipolis, France, 2013.
2. Kuhlmorgen, S.; Llatser, I.; Festag, A.; Fettweis, G. Performance Evaluation of ETSI GeoNetworking for Vehicular Ad Hoc Networks. In Proceedings of the 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, Scotland, 11–14 May 2015; pp. 1–6.
3. *102 940 (V1.3.1) ETSI Standard*; ITS Communications Security Architecture and Security Management. ETSI: Sophia Antipolis, France, 2018.
4. *IEEE Std 1609.2-2016 (Revision of IEEE Std 1609.2-2013)*; IEEE Standard for Wireless Access in Vehicular Environments Security Services for Applications and Management Messages. IEEE Standards: Piscataway, NJ, USA, 2016; pp. 1–240.
5. Whyte, W.; Weimerskirch, A.; Kumar, V.; Hehn, T. A security credential management system for V2V communications. In Proceedings of the 2013 IEEE Vehicular Networking Conference, Boston, MA, USA, 16–18 December 2013; pp. 1–8.
6. *102 941 (V1.3.1) ETSI Standard*; Trust and Privacy Management. ETSI: Sophia Antipolis, France, 2019.
7. *IEEE Std 1455-1999*; IEEE Standard for Message Sets for Vehicle/Roadside Communications. IEEE Standards: Piscataway, NJ, USA, 2006; pp. 1–134. [CrossRef]
8. Brecht, B.; Hehn, T. A security credential management system for V2X communications. In *Connected Vehicles*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 83–115.
9. *302 636-4-1 (V1.3.1) ETSI Standard*; Vehicular Communications, GeoNetworking. ETSI: Sophia Antipolis, France, 2017.
10. *103 415, (V1.3.1): Intelligent Transport Systems (ITS)*; Pre-Standardization Study on Pseudonym Change Management. ETSI: Sophia Antipolis, France, 2018.
11. *101 539 (V1.1.1) ETSI Standard*; Longitudinal Collision Risk Warning (LCRW) Application Requirements Specification. ETSI: Sophia Antipolis, France, 2013.

12. Rebollo-Monedero, D.; Forné, J.; Solanas, A.; Martínez-Ballesté, A. Private location-based information retrieval through user collaboration. *Comput. Commun.* **2010**, *33*, 762–774. [CrossRef]
13. Yao, Y.; Chang, X.; Wang, J.; Mišić, J.; Mišić, V.B.; Wang, H. LPC: A lightweight pseudonym changing scheme with robust forward and backward secrecy for V2X. *Ad Hoc Netw.* **2021**, *123*, 102695. [CrossRef]
14. Car2Car CAR 2 CAR Communication Consortium. Available online: https://www.car-2-car.org/about-c-its/#c176 (accessed on 29 November 2021).
15. Huang, B.; Li, J.; Lih, K.W.; Wang, H. Approximation Algorithms for the Generalized Multiple Knapsack Problems with K Restricted Elements. In Proceedings of the 2015 7th IEEE International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 26–27 August 2015; pp. 470–474. [CrossRef]
16. Shmoys, D.B.; Tardos, É. An approximation algorithm for the generalized assignment problem. **1993**, *62*, 461–474. [CrossRef]
17. Ma, Z.; Kargl, F.; Weber, M. Measuring location privacy in V2X communication systems with accumulated information. In Proceedings of the 2009 IEEE 6th International Conference on Mobile Adhoc and Sensor Systems, Macau, China, 12–15 October 2009; pp. 322–331.
18. Eckhoff, D.; German, R.; Sommer, C.; Dressler, F.; Gansen, T. Slotswap: Strong and affordable location privacy in intelligent transportation systems. *IEEE Commun. Mag.* **2011**, *49*, 126–133. [CrossRef]
19. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
20. Ma, Z.; Kargl, F.; Weber, M. A location privacy metric for v2x communication systems. In Proceedings of the 2009 IEEE Sarnoff Symposium, Princeton, NJ, USA, 30 March–1 April 2009; pp. 1–6.
21. Pfitzmann, A.; Hansen, M. *Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management-A Consolidated Proposal for Terminology*; Citeseer: Princeton, NJ, USA, 2005.
22. InterCor Project: Interoperable Corridors Deploying Cooperative Intelligent Transport Systems. Available online: https://intercor-project.eu/ (accessed on 29 November 2021).