*Manuscript

Journal Logo

Ad Hoc Networks 00 (2015) 1–16

Providing *k*-anonymity and revocation in ubiquitous VANETs

C. Caballero-Gil[a], J. Molina-Gil[a], J. Hernández-Serrano[b], O. León[b], M. Soriano[b,c]

[a] *Department of Computer Science. University of La Laguna. Spain.*
[b] *Department of Telematics, Universitat Politècnica de Catalunya, Spain*
[c] *Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Spain*

Abstract

Vehicular ad hoc networks (VANETs) is considered a milestone in improving the safety and efficiency in transportation. Nevertheless, when information from the vehicular communications is combined with data from the cloud, it also introduces some privacy risks by making it easier to track the physical location of vehicles. For this reason, to guarantee the proper performance of a VANET it is essential to protect the service against malicious users aiming at disrupting the proper operation of the network. Current researches usually define a traditional identity-based authentication for nodes, which are loaded with individual credentials. However, the use of these credentials in VANETs without any security mechanism enables vehicle tracking and therefore, violate users' privacy, a risk that may be overcome by means of appropriate anonymity schemes. This comes at the cost, however, of on the one hand preventing VANET centralized authorities from identifying malicious users and revoking them from the network, or on the other hand to avoid complete anonymity of nodes in front of the CA thus to allow their revocation. In this paper, a novel revocation scheme that is able to track and revoke specific malicious users only after a number of complaints have been received while otherwise guaranteeing node's *k*-anonymity is described. The proper performance of these mechanisms has been widely evaluated with NS-2 simulator and an analytical model validated with scripts. The results show that presented work is a promising approach in order to increase privacy protection while allowing revocation with little extra costs.

Keywords: Ubiquitous, VANET, revocation, *k*-anonymity, privacy, security.

1. Introduction

Vehicular Ad-hoc NETworks (VANETs) [1] are a subset of Mobile Ad-hoc NETworks (MANETs) aimed for providing the infrastructure for developing new communication systems that enhance drivers and passengers safety and comfort. Consequently, VANETs are considered to be part of the Intelligent Transport System (ITS) [2] and will be an important part in the development of strategic plans for ITS in the future [3]. Its fundamental structure consist of a set of On Board Units (OBUs) placed in vehicles and several stationary nodes called Road Side Units (RSUs) placed at fixed locations along the roads. While a high effort has been made to design new standards for VANETs services and interfaces [4], a real deployment on a large scale has not been carried out yet.

The need for ubiquitous data transmission capacity to offer complimentary services such as mobile Internet, video conferencing, television and downloading files, culminates with the emergence of first the General Packet Radio Service (GPRS) or 2G, then the Universal Mobile Telecommunications System (UMTS) or 3G, and finally the new LTE (Long Term Evolution) or 4G. These 3 different coexisting technologies are hereafter referred to as Mobile Telecommunications (MTs). The presence of MTs along with the emergence of powerful devices, such as smartphones, have

Email addresses: ccabgil@ull.es (C. Caballero-Gil), jmmolina@ull.es (J. Molina-Gil), jserrano@entel.upc.edu (J. Hernández-Serrano), olga@entel.upc.edu (O. León), soriano@entel.upc.edu (M. Soriano)

allowed the rise of new applications that provide real-time information about what happens on the road; examples of these applications are Waze, GMaps, TomTom, etc. Therefore, reusing current deployed MTs may save the additional associated cost for deploying new infrastructure as proposed in the standard. Furthermore, the technology is not limited and there are other researches exploiting other V2V communications methods [5], [6].

This paper proposes a new scheme for the deployment of VANETs that combines both the VANETs communications standard, relying on the OBUs and current Wi-Fi technologies [7], and the technology provided by the current MTs. Thus, on the one hand, Wi-Fi ad-hoc communications reduce energy consumption per node, enhance users' privacy and increase the number of potential users that may provide data to the network and on the other hand, current MTs provide widespread, often faster, communications to nodes.

Conventional VANETs and current traffic applications do not protect users' privacy. They can breach the privacy of the user of the vehicle because they manage information that allows knowing where users are in every moment. Nodes and/or users' privacy may be violated by the Telecommunications Service Providers (TSPs) that must know the telephony cell where the users are connected in order to establish connection. This level of privacy is lost since current mobile phone standards do not allow any modification. However, the Certification Authority (CA) that has to provide user certificates, and companies managing traffic data (Traffic Authorities - TAs), which must have information about traffic conditions, can locate and track vehicles based on their transmissions. Protecting users against tracking from the CA and TAs can be solved by providing complete user anonymity, but this lack of tracking avoids the revocation of malicious/misbehaving nodes disrupting the service operation.

This paper proposes a novel scheme that protects users' privacy in front of other users and authorities like the TAs and even the CAs while also offering the possibility to track malicious user and therefore, to throw them out of the system when a predefined amount of complaints have been received. The use of a new way of k-anonymity protection allows that the information for nodes cannot be distinguished from at least $k-1$ individuals in different locations whose information is signed by the same certificate.

The remainder of this paper is organized as follows. Section 2 gives an overview of the state of the art on privacy and revocation in VANETs. In Section 3 the scenario for the Ubiquitous VANET is defined. Section 4 presents the revocation control, details about the use of certificates and nodes authentication. Section 5 shows details about the group formation scheme and its security analysis. Section 6 includes a NS-2 simulation to demonstrate the performance in a real-based scenario, a theoretical analysis of the proposed privacy scheme and its verification with simulations. Finally, Section 7 gives the conclusions.

2. Related Research

It is supposed that VANETs will be deployed following the Wireless Access for Vehicular Environment (WAVE)[4] architecture. This architecture defines a new standard for vehicular communication that merely relies on the IEEE *802.11p* standard of the IEEE 802.11 family (wireless LANs - aka Wi-Fi) for transmissions.

However, from our point of view, WAVE architecture is incomplete as long as it does not consider the complementary use of GPS-based applications for smartphones, which have become quite popular and are thus widely spread among users, as a potential and very useful source of data for enhancing VANET services.

At the present time, there is an explosion in GPS-based applications offering traffic services based on information provided by local road authorities, police departments, systems that track traffic flow or information from other users, such as Google Traffic [8], TomTom [9], Sygic [10] or Waze [11]. Some of these apps provide near real-time data about traffic status and congestions. However, for these services to properly work, users should provide information with at least their location to the companies offering these services without any guarantee that these companies will use this data for other considerations like in [12]. Therefore, users' privacy are in risk.

There are some works dealing with this privacy and security risk in the literature [13], [14], some of them relying on providing user anonymity by means of the use of pseudonyms [15], revocation schemas [16] and/or k-anonymity schemes. The paper [17] is a short previous version of this paper.

As explained below, in this proposal we deal with this privacy risk by guarantying user k-anonymity in front of any other entity, even TAs and the system CA. In our scenario, a system provides k-anonymity if the authentication data for each user/node cannot be distinguished from at least other $k-1$ users, which can be easily achieved by providing group credentials for Logical Groups (LGs) of k users.

Sweeney proposed k-anonymity at first in 2002 [18]. Its original intention was to thwart the ability to link field-structured databases, but viewed more broadly has been applied to many other fields, such as VANET. Authors in [19] present an Efficient and Privacy-Aware revocation Mechanism (EPA) based on the use of Merkle Hash Trees (MHT) and a Crowds-based anonymous protocol. This proposal protects user privacy and traceability over other users or compromised RSUs by using pseudonyms. These pseudonyms are, however, generated by the RSU and therefore cannot protect users' privacy against the server or non-compromised RSUs in Ubiquitous VANETs as we propose in this paper. In [20], the authors propose a key management scheme based on Temporary Anonymous Certified Keys called TACKing that efficiently protects the authentication, revocation, and privacy in VANETs. The privacy of this method is limited since the group manager can trace the identity of all users in his group at every moment. Proposal in [21] uses k-anonymity in VANET applications where k-anonymity is provided by a centralized *anonymizer* that is based on the users' real location. In the paper, it is proposed a homomorphism for the location of a group of users that are near to others users. However, tracking with less precision is possible and users need to wait until at least $k - 1$ other nodes are close to their location to achieve enough anonymity. This delay is a problem because it reduces the quality of the users' localization in time and space which compromises real-time service availability and accuracy. Therefore, the approach does not work for real-time services or in low density areas. Authors of [22] propose k-anonymity in its aggregation model for VANETs that has intrinsic privacy benefits. Authors in [23] propose a hybrid and social-aware location-privacy in Opportunistic mobile social networks (HSLPO). This is a collaborative and distributed obfuscation protocol that offers location-privacy k-anonymity. The work in [24] presents a location privacy approach that uses group navigation combined with the use of mobility prediction and prospective path confusion, in order to mitigate tracking of a vehicle. Combined usage of Path Prediction and Paths Intersection (performed by a group leader) create a series of intersected paths, which provide a prospective form of path confusion. However, this proposal is not valid for roads with low density and the tracking is only mitigated but not avoided. In [25], the authors propose a new system to avoid tracking in a VANET by using pseudonyms. These methods only avoid that a node could be tracked by other users. The CA and companies that maintain navigation software can already track every user. Authors in [26] proposed an anonymous authentication scheme based on Biometric Encryption in VANETs that ensures authentication, resists multiple attacks, and provides k-anonymity. Since the privacy of this scheme relies on pseudonyms certified by the CA, it has the same problems that the proposal in [25]. The authors in [27] present a privacy preserving secure communication protocol for VANETs. This protocol anonymously authenticates safety messages in order to preserve vehicle privacy and prevent unauthorized tracking. To this purpose, the CA periodically generates a set of keys for every vehicle, so, users' privacy is lost. The proposal in [28] uses traditional k-anonymity to provide users' privacy in VANETs. Nevertheless, the paper does not clearly address the revocation problem. The authors in [29] use k-anonymity for location privacy by using a cloaking users' location technique that demonstrates the failure of the traditional uses of k-anonymity for protecting location privacy. Unlike other studies, current work proposes a different use of k-anonymity to protect privacy. In [30] authors propose a self-managed VANET without CA based on Certificates Graphs. In this paper, each node has a pseudonym and many sub-pseudonyms that periodically change. Therefore, passive users cannot track other users. In this scenario there is neither a RSU nor any cloud connection. As a result, tracking from the cloud is impossible. The unique way to track other user is physically because the user must be authenticated with other user in order to track it.

To the best of our knowledge, none of the solutions in the literature provides anonymity (even against the central authorities who provide the credentials) while allowing to later identify an anonymous and malicious node in order to revoke him/her from the system. Our proposal achieves this aim only and only if enough complaints against a malicious user are received.

3. Scenario for Ubiquitous VANETs

A VANET, as is often presented in the literature, consists of vehicles equipped with OBUs that allow them to communicate with other vehicles (Vehicle-to-Vehicle - V2V) and RSUs that are fixed infrastructures located on the road that allow Vehicle-to-Infrastructure (V2I) communications. However, this model of VANET would be extremely expensive both for users, who would have to buy and install some devices in their vehicles, and for the state, which would have to deploy RSU on the roads to support VANET services. For this reason, the authors of this paper think that VANETs should have different technologies involved such as Wifi, Wimax or MTs (2G, 3G, LTE, etc.). to take advantage of all the efforts made in the deployment of MTs technology, instead of adding new ones, since

it is a completely tested technology and its performance has been thoroughly proved. This paper presents the idea of a ubiquitous VANET where devices communicate in two ways: using the proposed VANETs standard ad-hoc communications and the currently deployed MTs communications.

In a VANET scenario, in order to allow users to authenticate data from other users, every vehicle must own valid credentials issued by a trusted third party or CA. However, the use of credentials linked to users' vehicles identities may violate the users' privacy. In order to overcome this risk, we assume in our proposal the use of logical groups and pseudonyms as beacons [31]. However, since a persistent use of a pseudonym becomes an identity, in order to avoid tracking by servers [32], nodes will change their group of pseudonyms each time they renew the logical group.

A node does not share its pseudonym with the server and therefore the server cannot track the user with this information. Every pseudonym is certified by a trusted CA and thus it is related to: a public certificate that holds the pseudoidentity data and a public key PKu, and a user-owned private key PKs associated to that public key. As later detailed in Section 5, our proposal can provide k-anonymity while still being able to revoke specific users by making the pseudonyms refer to groups of k members (k members sharing the same identifier).

According to the proposal in [33], two entities with different roles are mainly proposed for the management of a ubiquitous VANET scenario: on the one hand, the CA is responsible for maintaining the relationship between pseudonyms, keys, and certificates; on the other hand, the Traffic Authority (TA) is responsible for collecting traffic information sent by nodes. Besides, the TA provides and disseminates traffic information to the network. Communication between the two authorities will be carried out through a secure channel. All the scope is summarized in Figure 1(1).

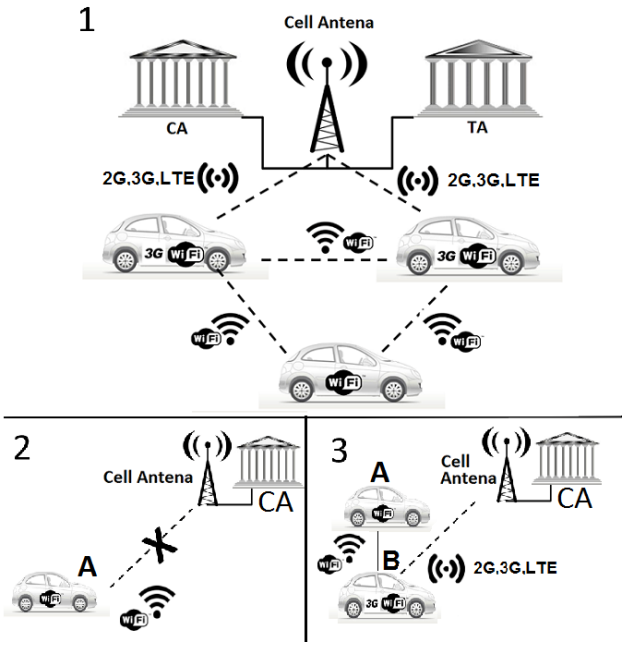

Figure 1. Scope of the ubiquitous VANET. Nodes connect to the network directly or through other nodes. Different ways to update certificates with authorities.

Besides the general description of the scenario, we have assumed the following premises during the definition of our proposal:

- Nodes can inject false information intentionally or due to a system malfunction.
- Both the CA and the TA have a high computational power.
- Not all the OBUs have MTs connection so that, communication with the CA and the TA are intermittent for some nodes.

- Nodes revocation is performed based on the complaints emitted by other nodes.
- The cryptography used will be based on elliptic curves to reduce the computational cost.
- The CA public and private keys are not compromised.
- The CA is reliable and will properly perform the Logical Groups operation.

4. Revocation Control

The traditional method for obtaining data about revoked nodes is that every node periodically or on-demand downloads a certificate revocation list signed and maintained by the CA. Every day or even more often, the OBUs have to update or download the list by connecting with the CA. As shown in [34] this approach is impractical for VANET. The high mobility of nodes, the limited capacity of the channel, the time required to download the CRL and the amount of memory required to store it, are some of the problems that make this proposal infeasible.

In order to overcome these problems, in this work we propose a system where each node has to demonstrate that it is a trusted node and it is not revoked by the CA. The system is merely based on assigning short-life pseudonyms to users. Users are responsible for updating their certificate before its expiration; otherwise, they will not be part of the network and thus will not take profit of the information relayed on it.

In the following we detail how the credentials are issued to the nodes, how users/OBUs authenticate other ones and how a single user/OBU is flagged as revoked when a given amount of complaints against it are received.

4.1. Obtaining credentials from the CA

When a node A first enters the system it is loaded with their own non-changeable pair: private key PKs_A and public certificate $Cert(A, t)$. The public certificate is a document signed by the CA containing the pseudoidentity data of A, A's public key PKu_A and an expiration date t.

As later explained in Section 5.1, in order to achieve k-anonymity, new credentials of users are generated periodically by the server and shared with other $k-1$ users belonging to the same logical group. The implications of this credentials sharing will be detailed in that section and, for the sake of clarity, in this section we will just focus on the fact that every user owns valid credentials.

When A's certificate is about to expire, A should query the CA (authenticating with its current credentials) for updating its certificate. The CA checks if the user is revoked based on the number of complaints registered against that user and, if it is not the case, issues a new valid certificate $Cert(A, t')$ that will expire in time t'. Notice that a revoked user will be expelled from the system once it has to update the certificate as long as it will not be able to acquire a new valid certificate.

The certificates updating process is a key point for the VANET safety and operation. On the one hand, if the time for updating the certificate is short, it greatly increases processing data on the server and number of server-OBUs communications. On the other hand, if this measure is too long, the attackers have more time to send false information to other nodes.

As a first trivial approach, given that the average trip time by car is about 22 minutes each way [35], if we want to discover and remove misbehaving/malicious nodes with bad behavior from the system in a single day, the lifetime of a certificate should be no more than $\frac{22*2}{N_{complaints}}$, with $N_{complaints}$ the number of complaints that the CA needs to detect with high probability a malicious node. As later explained, this is a privacy policy related to the privacy level we want to provide and the probability of having false positives and/or negatives.

Obviously, the proper behavior of this proposal relies on the connectivity between nodes and the CA. This connectivity can be easily assumed when the nodes have Internet access through currently available MTs or, when they are in the Wi-Fi range of an Access Point (AC). However, it may happen that a node A, without MTs connection may find a path to the AC through an ad-hoc Wi-Fi connection to a node B which has MTs connection. In this case, node B can connect to the CA thus allowing the node to update its certificate (see Figure 1(2)(3)). As long as the use of certificates provides end-to-end mutual authentication, the process can be securely carried out regardless of the number of intermediate nodes in the same manner as it was when the node is in the AC range.

Obviously, a long-term lack of connectivity between a node and the CA can lead to the unwanted situation of a legitimate node holding an expired certificate, which is the same as being revoked. Mainly two options can be took in such a case:

- 1) to keep the node out of the services (it can neither send nor receive messages) until it connects to the AC and thus obtains a valid certificate as shown in Figure 1(3); and less strictly
- 2) to allow to use the expired certificate during a time of reconnection t_r after its expiration. At this way, users can use the network without MTs connection. This t_r will be longer if the density of users with MT connection on the road is low.

4.2. Authentication between OBUs

When two nodes meet in the network, they exchange data that they have collected by any of the two available interfaces, either Ad-hoc 802.11p or currently available MTs. For this purpose, each participant node in communication, let us say node A and node B, must demonstrate that they have not been revoked.

In the first communication step, both nodes exchange their LG certificates as in Figure 2(1). Assuming that the sent certificates are not expired, both nodes proof that they have not been revoked. However, a quite easy attack in this context would consist on reusing any certificate that has been collected previously from another node. In order to avoid it, nodes must pass a revocation test. Therefore, as shown in Figure 2(2), both nodes exchange random challenge messages encrypted with their private keys, so that anyone can check that they are the only potential senders, and reencrypted with the public key of the destination node, so that the destination is the only one that can decrypt the message. Once the receiver decrypts the message, it replies with the message encrypted with the sender's public key as in Figure 2(3). Upon the sender receives it, the sender can decrypt the message and checks that it corresponds with the one it previously sent. In such a case, authentication is over and both nodes are sure that the other has a valid certificate at this moment.

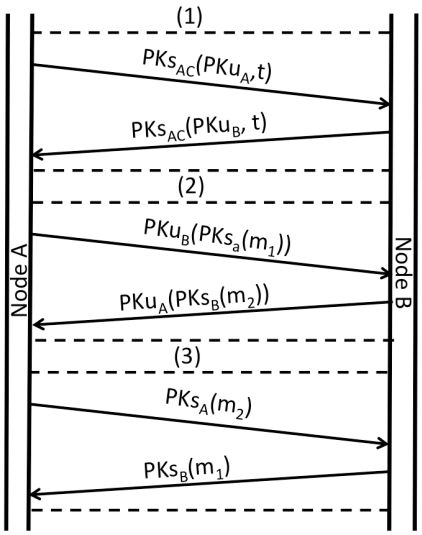

Figure 2. Packets exchange for authentication between OBUs.

5. The Traceable Distributed K-Anonymity Method

This paper proposes a method that provides k-anonymity [18] in VANETs guaranteeing that malicious nodes will be traceable. The method operation is described as follows. Every node is randomly associated to a group n with k members that share cryptographic material, i.e., a pair of private-public keys (PKu_{G_n}, PKs_{G_n}), and a group

certificate $Cert(G_n, t)$, which will be used to sign messages, complaints or any authenticated data. The system will use cryptography based on elliptic curves [36] that is better in computational cost. The CA will be the responsible for creating the logical groups, maintaining a database with the group membership and complaints, distributing the cryptographic material among users or revoking users that misbehave, etc.

When a node detects an undesired behavior from another node, such as false data, it presents a complaint to the CA. Such complaint must be signed with its group key and contain information such as the group identifier of the malicious user, position, date, infraction, etc. In its turn, the CA flags all the nodes belonging to that group and keeps all the relevant information in its database. Since nodes do not reveal their particular identities but only their group identity, both the malicious node and the one who sends the complaint cannot be distinguished from other nodes belonging to the same group. Therefore, they cannot be traced by the CA or by other nodes. This feature protects users' privacy, but makes harder the task of revoking and isolating malicious nodes from the network.

To achieve the traceability of malicious users is a must to be able to revoke them. In our proposal we use group certificates with short-term expiration dates that henceforth we will call round r. When a node detects that its current group certificate is about to expire, it automatically sends a query to the CA to update it. The CA will check if this particular node is revoked based on the number of flags that it has received. The rationale behind such mechanism is that nodes will change of group frequently over time and will be flagged whenever they belong to one group which has been accused in a complaint.

With this method, the bigger the group size k is, the better the level of anonymity. However, as previously mentioned, the number of required complaints for revoking users with k-anonymity is bigger as well. Later in section 6.4 we derive an analytical expression for the number of false positives and false negatives affecting the system, i.e., the number of honest nodes being revoked or the number of malicious nodes that remain in the network without being detected. Furthermore, provided by this method as a function of the anonymity value k, the number of complaints required to revoke nodes and the number and period of group changes.

The CA should guarantee that the assignment of a node to a given group is random, kept in secret and exclusively known to the user and itself. In addition to this, only the members of the group can have access to the cryptographic material of the group. For this reason every node should be provided with a public/private key pair and the corresponding certificate when entering the network. Such cryptographic material would be exclusively used for communication with the CA or other users of the VANET in order to authenticate the node against them or renew group certificates.

5.1. Logical Groups (LG)

A LG consists in a group of OBUs formed without following any physical criteria such as location, traffic density, data aggregation, etc. All nodes belonging to the same LG share a pair of keys (PKu_{G_n}, PKs_{G_n}) instead of having a different pair of keys (PKu_n, PKs_n) for each one. The same idea is used for the certificates, i.e., each node has a LG certificate $Cert(LG_n)$, with n being the group to which it belongs. The CA is responsible for randomly create these groups, where each one has approximately the same number of nodes k, and also the cryptographic information for each group. When a node presents a complaint, it uses its LG key to sign it and reports the group identifier of the LG to which the malicious nodes belongs. Because nodes only provide LG information and not their particular identity, they avoid being tracked by other nodes and also by the CA.

However, as pointed out above, k-anonymity makes impossible to distinguish a well-behaved node from a malicious one belonging to the same group, and thus to revoke and isolate malicious nodes from the network. In order to overcome this limitation, in our proposal each node is required to change to a new group after a given time interval. In this way, every time a new complaint for a given LG is received, all nodes belonging to that group receive a penalty mark. Note that, because the CA knows the membership list for each LG at any point in time, it can easily assign penalty marks when a complaint is received and a malicious node will not be able to acquire a new valid certificate when he has several complaints. With this mechanism, well-behaved nodes will receive a penalty mark whenever they belong to the LG reported in the complaint. However, if we assume that malicious nodes repeatedly misbehave, they will always receive at least the same amount or more penalty marks than any other honest node in the network. Section 6.2 discuss the requirements of the system, in terms of number of penalty marks and group size k in order to properly trace malicious nodes and revoke them while keeping an acceptable level of anonymity.

The performance of the proposed method is illustrated in Figure 3, which shows an example with 4 groups or LGs and 9 members or nodes per LG. As it can be observed, the GPS coordinates of the vehicles belonging to the same

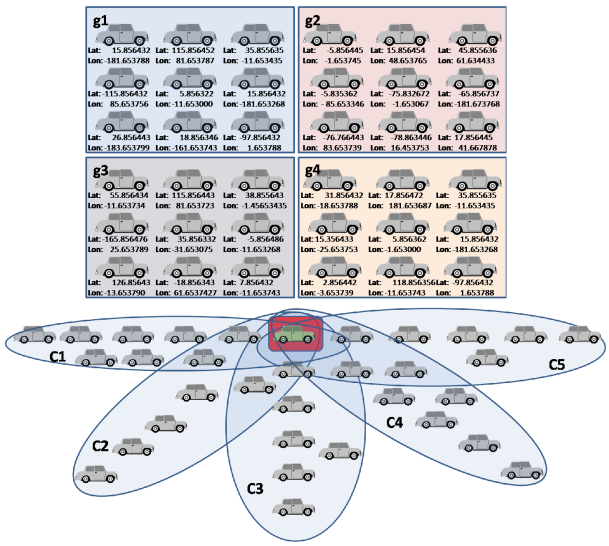

Figure 3. Logical groups, vehicles share the same group but they can be in very different physical positions, an attacker is discovered when enough complaints are received.

group are very different because groups are randomly created. Thus, using LGs, any node can sign data or appear in data signed without being recognized by other nodes nor the CA.

The proposed mechanism to detect malicious nodes is presented in figure 3. Let $PKu_{G_A}(Complaint(PKu_{G_B}, t_0))$ be the complaint sent by a member of group A reporting a bad behavior from a member belonging to group B. Consider that the CA receives the following complaints in different time intervals:

$PKu_{S_1}(Complaint(PKu_{S_2}, t_0))$
$PKu_{S_2}(Complaint(PKu_{S_4}, t_1))$
$PKu_{S_3}(Complaint(PKu_{S_1}, t_2))$
$PKu_{S_4}(Complaint(PKu_{S_2}, t_3))$
$PKu_{S_5}(Complaint(PKu_{S_3}, t_4))$

Note that not all the complaints are considered, since the CA discards those which refer to the same group as another complaint which has been received in the same time interval. Because the green vehicle is the only one who was associated to LGs 2, 4, 1, 2 and 3 in each time interval, it has received five penalty marks (more than the remaining vehicles) and thus the CA identifies it as the malicious node.

Also, to prevent a group of nodes from colluding against a legitimate node, our proposal only takes into account one complaint per group during a time interval, i.e., the amount of time during which a node remains as a member of a given group. Because after such amount of time every node moves to a different group, the colluding nodes would find rather difficult to determine the new group of the victim, and thus the probability of success of the attack would be considerably reduced. Finally, it must be taken into account that is possible that nodes generate wrong information due to a device failure or a user misuse. In that case, nodes could report an incorrect attack attempt. For these reasons, the system should have some fault tolerance.

5.2. Possible attacks to the LG schema

In this subsection some of the most frequent attacks to VANETs are analyzed regarding to the point of view of Logical Groups.

Impersonate: An attacker with the identity and privileges of an authorised node could inject false information to the network. This attacker would be discovered with LG if he attacks in a period of time. Otherwise, it wouldn't be discovered, but the result of the attack would be very limited.

Sybil: An attacker that controls multiple nodes could inject false information to the network. In short period of time the attacker wouldn't be detected with LG scheme. But at this way, the result of the attack is limited. Some

mechanisms such as the use of social networks and phone numbers to the user authentication process would make difficult the task of the attacker to control multiple nodes.

Collusion attack: Collusion attacks, where colluders are members of different LGs is not an easy task since every node must have a signed certificate by the authority. As it will be explained in Section 6.5, the system can detect collusion attacks with a low error rate. At this way, the authority would detect several attacks in a short period of time. Hence, colluders will be detected and revoked avoiding that they can enjoy the network again.

Session hijacking / Identity revealing: A strong authentication protocol would difficult this task to attackers.

Location Tracking: Logical Groups itself protects users location privacy.

Repudiation: This paper presents a mechanism to discover attackers providing k-anonymity. When an attacker is unequivocally traced, he cannot repudiate the data related to it.

Eavesdropping: The use of changing pseudonyms protects the confidentiality of users. Confidential data, that is the main goal of this attack, is protected with a strong authention protocol.

Denial of Service: DoS attacks can be carried out in many ways. With Logical Groups, most of DoS attacks would decrease the time that an attacker could attack because it would be discovered. If there are multiple requests in a short period of time, only the first request of every Logical Group would be taken into account.

Routing attack: Attacks which exploits the vulnerability of network layer routing protocols in Ubiquitous VANETs has no sense since data can reach the objective from different sources.

6. Analysis and Simulation of K-Anonymity Method

6.1. Simulation of LG in real-based scenario

Both the feasibility and effectiveness of our proposal have been tested through simulation. This section presents some details and results of these simulations. In particular, we have used NS-2 [37] and SUMO [38] (see Fig. 4) to simulate a VANET with a variable number of nodes ranging from 1300 to 2000. Due to computational constraints, we cannot provide simulation results for a larger number of nodes, which would be more realistic. The total number of vehicles with OBUs varies from 1% to 50% of the total number of nodes, the vehicle proximity is 75 meters according the study presented in [30], and a simulation time of 1000 seconds. We have considered that the number of malicious nodes in the network varies from 1 to 10 and, regarding the proposed method, we have considered different values k of anonymity ranging from 10 to 100 and a variable number of rounds r or group changes between 6 and 12. The number of complaints s needed to consider a node as malicious is computed as explained in section 5.1, as a function of the anonymity value k and the number of rounds r.

In the simulator, the vehicle mobility layer manages the node movement according to the movement pattern, which defines roads, lines, different speed limits for each line, traffic congestions, etc. Figure 4 shows a screenshot of a simulation performed with NS-2 and SUMO, where we can see the mobility model and the communications and group formation model.

The node energy layer is used to distinguish between vehicles with and without OBUs because vehicles without OBUs are on the road but they cannot communicate with other nodes. By means of such layer, we have computed the number of approximations between nodes in the network related to the percentage of vehicles with OBUs and size of every group (see Figure 5). This allows us to get an idea of the minimum value of anonymity k needed for the proposed method as a function of the number of nodes.

6.2. Traceability with LGs

In this section, the number of reports required to revoke a node is discussed. With this purpose, we compute the probability of unequivocally trace a node as a function of the number of complaints.

Note that it is possible to find different nodes having been associated to the same groups over time, but the probability of two or more nodes being at the same group at each time interval, i.e., to follow the same path in terms of group changes, tends to zero as the number of groups is increased. Let n be the number of users in the system, $g < n$ the number of groups with $k = \frac{n}{g}$ nodes, s the number of penalty marks or complaints required to consider a user as malicious. Also, let us assume that all groups are equiprobable, so that the probability of a given user moving to a given group is $\frac{1}{g}$.

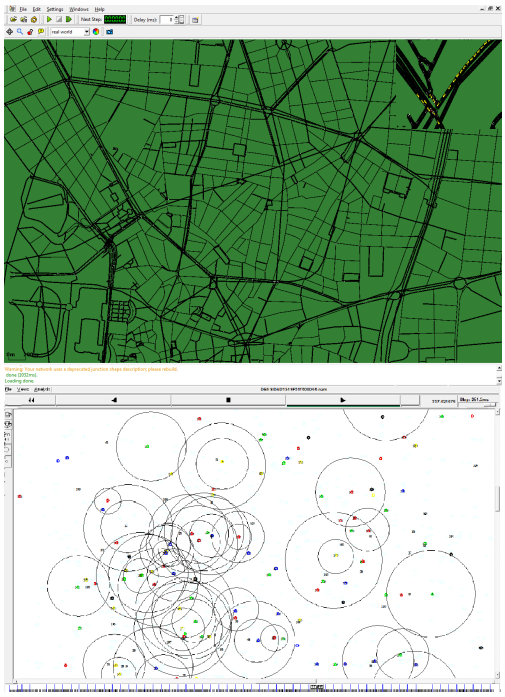

Figure 4. Simulation of LG with SUMO, MOVE and NS-2 in center of Madrid.

The probability of a given user following a unique path in terms of changes of groups can be easily computed as the probability that no other user follow such path. Let's consider a single group change for a given user. Then, the probability of the rest of group users ($k-1$ users) going to different groups can be expressed as in (1).

$$\left(1-\frac{1}{g}\right)^{k-1} \tag{1}$$

From the previous reasoning the same probability after $s-1$ group changes can be expressed as in (2).

$$p=\left(1-\left(\frac{1}{g}\right)^{s-1}\right)^{k-1}=\left(1-\left(\frac{k}{n}\right)^{s-1}\right)^{k-1} \tag{2}$$

The probability of being unequivocally traced depending on the size of groups and the number of complaints depends on the size of the group, the smaller the group is, the bigger the probability of being traced. Note that, although the level anonymity increases with the size of the group, it also does the number of needed complaints. This implies that a malicious node can attack the network for longer without being detected. For example, in a scenario with 2^{22} users, with group size of 2^{17} nodes, the number of needed complaints to discover an attacker with a high probability is 5, this shows that it is possible to get a high level of privacy and detect a malicious node quite fast. Furthermore, in order to prevent accumulative incorrect data, each complaint has a validity period. Once this period expires, if the complaint has not been used to revoke any node, it is dismissed.

The CA must update three lists. The first one is the list of complaints, which stores data about complainant, denounced, date, coordinates and type of complaint. The second list stores information about the groups assigned by the CA: the group number, the nodes that belong to this group at this moment and the period of validity of the certificate. The third list associates the complaints generated and received in every group that the node has been associated to. These reports are needed to unequivocally trace a node. Furthermore, this list stores nodes that are revoked.

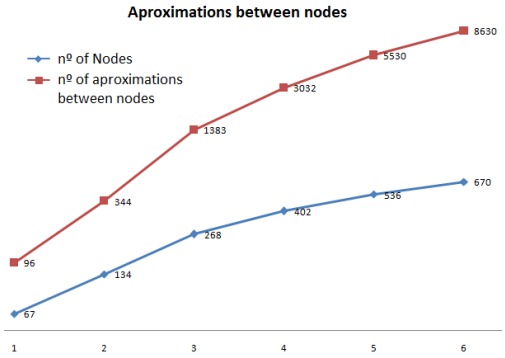

Figure 5. Number of physical approximations between nodes that belong the same LG in the SUMO and NS2 simulation in center of Madrid.

6.3. Optimal Group Size for k-anonymity

The revocation of a node from the network is performed when it has received a certain number of penalty marks or complaints. This number depends on the number of groups and the group size k. If the group size is big, the number of required complaints increases so the malicious node detection process lasts for a long time and is very expensive. However, if the group size is small, the system is faster identifying malicious nodes but the user privacy is put in an awkward position easily. Therefore, there is a tradeoff between privacy and speed to detect a malicious node.

Assuming that malicious nodes misbehave at least once every time they move to a new group, to determine the required number of complaints in order to achieve a given k-anonymity is analogous to compute the number of group changes s and can be derived by isolating s in equation (2) as in (3).

$$s = \left\lceil \frac{\log\left(\frac{k}{n}\left(1 - p^{\frac{1}{k-1}}\right)\right)}{\log \frac{k}{n}} \right\rceil \tag{3}$$

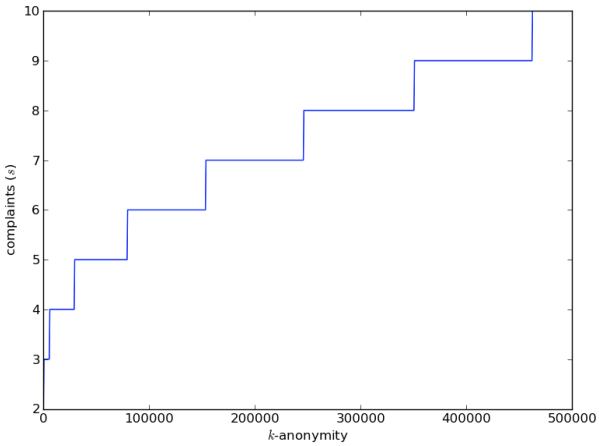

Figure 6. Complaints required to unequivocally discover an attacker using k-anonymity with equation 3.

The representation of this function is shown in figure 6. It represents the optimal k-anonymity depending on the number of complaints. In order to avoid revoked nodes can get benefits from the network, the number of complaints should be low but, to ensure a good level of k-anonymity, the number of nodes per group must be high. According to the graph, 6 or 7 complaints provided a good level of k-anonymity with a good performance to detect an attacker. 7 complaints can offer varying k-anonymity in groups sizes between 150.000 and 250.000 nodes. With this in mind, if