On the revocation of malicious users in anonymous and non-traceable VANETs

Cándido Caballero-Gil,
Jezabel Molina-Gil
Departamento de Estadística,
Investigación Operativa y Computación
Universidad de La Laguna
Email: {ccabgil|jmmolina}@ull.es

Juan Hernández-Serrano,
Olga León,
Miguel Soriano-Ibañez
Departamento de Ingeniería Telemática,
Universitat Politècnica de Catalunya (UPC)
Email: {jserrano|olga|soriano}@entel.upc.edu

***Abstract*—For the proper performance of Vehicular Ad-hoc NETworks (VANETs) it is essential to protect the service against malicious nodes aiming at disrupting the proper operation by injecting fake, invalid data into the network. It is common to define a traditional identity-based authentication for vehicles, which are loaded with individual credentials. However, the use of these credentials in VANETs may allow vehicle tracking and thus violate users' privacy, a risk that can be overcome by means of anonymity schemes. This comes at the expenses, however, of on the one hand preventing VANET authorities from identifying malicious users and revoking them from the network, or on the other hand to avoid anonymity of users in front of the CA thus to allow their revocation. In this work, we describe a novel revocation scheme that is able to track and revoke specific users only after a number of complaints have been received while otherwise guaranteeing vehicle's anonymity.**

***Index Terms*—VANET, revocation, k-anonymity**

I. Introduction

Nowadays, there is a bunch of GPS-based applications offering traffic services based on information provided by local road authorities, police departments and systems that track traffic flow. Some of these applications, such as Google Traffic [7], TomTom [16], Sygic [15] or Waze [18], can even 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 [1]. Therefore, users' privacy may be at risk.

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 reveals the location of vehicles at every moment. Nodes and/or users' privacy may be violated by the Certification Authority (CA) as long as it provides their certificates, but also by companies managing traffic data or Traffic Authorities (TAs), which can locate and track vehicles based on their transmissions. Protecting users against tracking can be solved by providing user anonymity but, at the same time, this lack of tracking avoids the revocation of malicious/misbehaving nodes disrupting the service operation.

In this work we propose a novel scheme that protects users' privacy in front of other users, TAs and even the CA while also offering the possibility to track malicious user and thus to throw them out of the system when a predefined amount of complaints have been received. As later explained malicious user can only be tracked after a predefined amount of complaints are received by the CA. To do this we will use k-anonymity protection that allow that the information for nodes contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release.

This paper is organized as follows. Section II covers related research about privacy in VANETs. In Section III we present the proposed method to guarantee anonymity of users while allowing tracking malicious users. In section IV we derive an analytical model to analyze the efficiency of the method. This model has been validated via simulation in section V, where we also provide the results obtained by means of the model for a real scenario. Finally, the conclusions of this work are presented in Section VI.

II. Related Research

Sweeney proposed k-anonymity at first in 2002 [14] and its original intention was to thwart the ability to link field-structured databases, but has been viewed more broadly, and have been applied to many other fields, such as VANET.

In [13], authors use k-anonymity in VANET applications where k-anonymity is provided by a centralized *anonymizer* based on the users' real location. The author proposes a homomorphism for the location of a group of users that are near to others users. However, tracking with less precision is already possible and users need to wait until at least $k-1$ other users are close to their location to achieve enough anonymity. This delay reduces the quality of the users localization in time and space which compromises real-time service availability and accuracy. Thus, this approach does not work in case of real-time services and also in low density areas of users.

[19] proposes a hybrid and social-aware location-privacy in Opportunistic mobile social networks (HSLPO), a collaborative and distributed obfuscation protocol that offers location-privacy k-anonymity.

ISBN: 978-84-9717-323-0

In [2] authors propose a self-managed VANET without CA based on Certificates Graphs where every node has a pseudonym and many sub-pseudonyms that change frequently in a range of time, at this way, passive users cannot track other users. In this scenario there is neither a RSU nor any cloud connection. Therefore, tracking from the cloud is impossible. The unique way to track another user is physically because the user must be authenticated with another user to track it.

In [20] vehicles entering a group can anonymously broadcast vehicle-to-vehicle (V2V) messages, this is another way to preserve privacy but the TM has the ability to retrieve the real identity of dishonest vehicles that are sending fake messages to other vehicles to disrupt traffic, so the privacy is violated.

[4] present different privacy-preserving variants to ensure that vehicles volunteering to generate and/or endorse trustworthy announcements do not have to sacrifice their privacy.

[11] proposes a protocol to exclude malicious network nodes based on complaints received from other vehicles. [17] presents another protocol that uses decentralized revocation voting. [12], [8], [9] are other approaches for conditional anonymity in VANET.

To the best of our knowledge, none of the proposals in the literature provide both complete anonymity (even against the CA providing the credentials) while allowing to later identify an anonymous user in order to revoke him/her from the system. In our proposal we achieve this goal only and only if several complaints against a user is received.

III. A TRACEABLE K-ANONYMITY METHOD FOR VANETS

In this section, we propose a method that provides k-anonymity [14] in VANETs while still guaranteeing that malicious users will be traceable. The method operation is as follows. Every user 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 and authenticate data. In order to reduce the computational cost, we assume the use of cryptography based on elliptic curves [3]. We also consider the existence of a CA, which is responsible for creating the groups, maintaining a database with the group membership, distributing the cryptographic material among users and revoking users that misbehave.

When a user detects an undesired behavior from another user, such as the injection of false data, it presents a complaint to the CA. Such complaint is signed with its group key and must report the group identifier of the malicious user. In its turn, the CA flags all the users belonging to that group and stores this information at its database. As users do not reveal their particular identities but only their group identifier, both the malicious user and the one sending the complaint cannot be distinguished from other users belonging to the same group. Thus, they cannot be tracked by the CA or by other users. This feature protects users' privacy, but it makes a hard task to revoke and isolate malicious users from the network.

In order to achieve the traceability of malicious nodes to revoke them, we propose the use of group certificates with short-term expiration dates that henceforth we will call $roundr$. When the group certificate of a user is about to expire it must send a query to the CA to update it, which will check if the user is revoked based on the number of flags it has received. The rationale behind such mechanism is that users will change of group over time and will be flagged whenever they belong to the group reported in a complaint. Due to the fact that you can not identify what node is having a bad behaviour, only one complaint is valid in each round. Assuming that malicious users repeatedly misbehave, they will be flagged at least the same amount of times or more than any other honest node in the network and thus they can be easily revoked by the CA. Note that, a revoked user will be expelled from the system once it has to update its certificate, as it will not be able to acquire a new valid certificate.

The certificate updating process is a key point for the VANET safety and operation. On the one hand, if the time for updating the certificate t is short, it greatly increases processing data on the server and number of server-users communications. On the other hand, if this measure is too long, malicious users would remain in the network for a longer time without being revoked. As a first, trivial approach, given that the average trip time by car is about 22 minutes each way [5], the lifetime of a certificate t should be no more than $\frac{22*2}{f}$, with f the number of complaints that the CA needs to revoke a user if we want to remove malicious users from the system in a single day. Furthermore, in order to reduce overhead, multiple certificates can be issued without interaction between CA and the user [10].

The level of anonymity provided by this method increases as the group size k does. However, as previously mentioned, k-anonymity complicate the process of revoking users. In section IV we derive an analytical expression for the number of false positives and false negatives, i.e., the number of honest users being revoked and the number of malicious users that remain in the network, provided by this method as a function of the anonymity value k, the number of complaints needed to revoke users and the number of group changes.

The CA should guarantee that the assignment of a user to a given group is kept secret and exclusively known to the user and itself, and that only the members of the group have access to the group cryptographic material. Because of this, every user 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 in order to authenticate the user against it and renew group certificates.

IV. ANALYTICAL MODEL

In this section, we derive the probability of false positives and false negatives of the proposed k-anonymity mechanism, i.e., the probability of an honest user being regarded as malicious one and the probability of not detecting and actual attacker. For the sake of clarity, table I presents the specific notation considered from now on.

TABLE I
NOTATION

r	number of rounds
n	number of users in the system
a	number of malicious users
p	prob. of a malicious user performing an attack in a round r
k	number of users in a group
f	number of flags needed to revoke a user
t	time of the certificate expiration
p_h	prob. of an honest user being flagged in a round
p_a	prob. of an attacker being flagged in a round
$p_h^{f,r}$	prob. of a honest user being flagged f times after r rounds
$p_a^{f,r}$	prob. of an attacker being flagged f times after r rounds
FP	r.v. number of false positives in r rounds and f flags
FN	r.v. number of false negatives in r rounds and f flags

Given these definitions, we denote as p_h and p_a the probabilities of an honest user and an attacker, respectively, receiving a flag in a given round r, and can be computed as in (1) and (2) with $\alpha_h = \min(k-1, a)$ and $\alpha_a = \min(k-1, a-1)$.

$$\begin{aligned} p_h &= \\ &= \sum_{i=1}^{\alpha_h} \binom{k-1}{i} \prod_{j=1}^{i} \tfrac{a-j+1}{n-j} \prod_{j=1}^{k-i-1} \tfrac{n-a-j}{n-i-j} \left(1-(1-p)^i\right) \\ &= \sum_{i=1}^{\alpha_h} \binom{k-1}{i} \tfrac{a!\,(n-a-1)!\,(n-k)!}{(a-i)!\,(n-1)!\,(n-a-k+i)!} \left(1-(1-p)^i\right) \end{aligned} \tag{1}$$

$$\begin{aligned} p_a = p \cdot 1 + (1-p) \cdot \\ \sum_{i=1}^{\alpha_a} \binom{k-1}{i} \prod_{j=1}^{i} \tfrac{a-j}{n-j} \prod_{j=1}^{k-i-1} \tfrac{n-a-j+1}{n-i-j} \left(1-(1-p)^i\right) \end{aligned} \tag{2}$$

Then, we can compute the probability of an honest user and a real attacker being regarded as attackers after r rounds as in (3) and (4), respectively.

$$p_h^{f,r} = \sum_{i=f}^{r} \binom{r}{i} p_h^i (1-p_h)^{r-i} \tag{3}$$

$$p_a^{f,r} = \sum_{i=f}^{r} \binom{r}{i} p_a^i (1-p_a)^{r-i} \tag{4}$$

From the above equations, it can be easily derived the probabilities of false positives and false negatives as a function of the number of rounds r, i.e., the probability of a honest user being flagged as an attacker and the probability of an attacker being regarded as a honest user after r rounds.

The probability of false positive after r rounds p_{fp}^r is the probability of at least one honest user having more than f flags, which is equal to 1 minus the probability of all honest users having less than f flags, and can be expressed as in (5).

$$p_{fp}^r = 1 - \left(1 - p_h^{f,r}\right)^{(n-a)} \tag{5}$$

Analogously, the probability of false negative p_{fn}^r is the probability of at least one attacker having less than f flags after r rounds, which is equal to 1 minus the probability of all attackers having f or more flags, and can be expressed as in (6).

$$p_{fn}^r = 1 - \left(p_a^{f,r}\right)^a \tag{6}$$

In order to analyze with more precision the goodness of the mechanism, it can be useful to estimate the expected number of false positives and false negatives and their variance, as a function of the number of flags f and the number of rounds r. Let us define $FP_{f,r}$ and $FN_{f,r}$ as two discrete random variables following a binomial distribution that account for the values of false positives and false negatives respectively. Then, we can define their respective probability mass functions as in (7)), with expected values μ as in (8) and variance σ^2 as in (9).

$$\begin{aligned} f_{FP_{f,r}}(i) &= \mathrm{P}(FP_{f,r} = i) \\ &= \binom{n-a}{i} \cdot (p_h^{f,r})^i \cdot (1-p_h^{f,r})^{n-a-i} \\ f_{FN_{f,r}}(i) &= \mathrm{P}(FN_{f,r} = i) \\ &= \binom{a}{i} \cdot (1-p_a^{f,r})^i \cdot (p_a^{f,r})^{a-i} \end{aligned} \tag{7}$$

$$\begin{aligned} \mu_{FP} &= \mathrm{E}[FP_{f,r}] = \sum_{i=1}^{n-a} i \cdot \mathrm{P}(FP_{f,r} = i) \\ \mu_{FN} &= \mathrm{E}[FN_{f,r}] = \sum_{i=1}^{a} i \cdot \mathrm{P}(FN_{f,r} = i) \end{aligned} \tag{8}$$

$$\begin{aligned} \sigma^2_{FP_{f,r}} &= \mathrm{V}[FP_{f,r}] = \sum_{i=1}^{n-a} (i - \mu_{FP_{f,r}})^2 \cdot f_{FP_{f,r}}(i) \\ \sigma^2_{FN_{f,r}} &= \mathrm{V}[FN_{f,r}] = \sum_{i=1}^{a} (i - \mu_{FN_{f,r}})^2 \cdot f_{FN_{f,r}}(i) \end{aligned} \tag{9}$$

V. Performance evaluation

In this section we first validate in section V-A the analytical model presented in section IV via simulation, and then in section V-B we evaluate the goodness of our proposal when applied to a real scenario such as the current Spain's vehicle fleet.

In order to make the simulation easier, we have assumed in the following that an attacker always attacks in a round, that is to say that the protocol operation properly detects all the attackers and thus cannot lead to false negatives. However, there are still chances of leading to false positives (honest users flagged as attackers) and therefore the following analysis mainly focuses on the mean and variance of false positives.

A. Validation of the analytical model

Figure 1 shows the mean number of false positives obtained by simulation (100 iterations per each possible combination) in dashed line vs the mean and standard deviation in continuous line obtained by the analytical model. Due to space constraints, we present just a specific case; the purpose of it is just to show that the analytical model properly fits the protocol behavior.

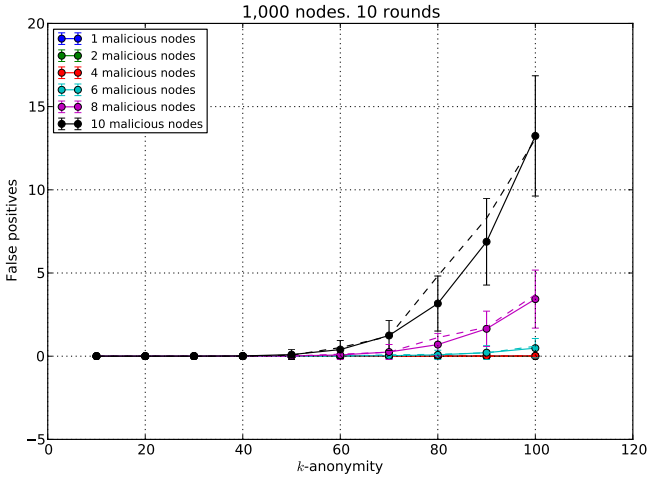

Fig. 1. False positives after 10 rounds with 1,000 nodes

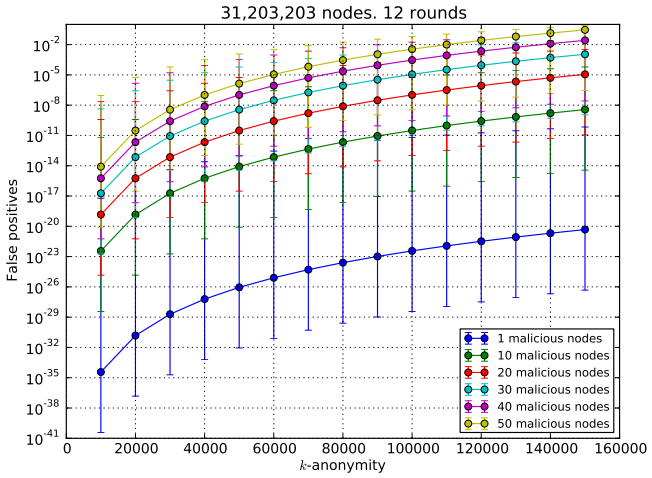

Fig. 2. Mean and STD of false positives vs k-anonymity after 12 rounds for a varying number of attackers

Obviously, it also fits in any other case but in the figures we present the average number of false positives after 10 rounds of operation in a network with 1000 nodes, a varying number of malicious nodes from 1 to 10 which are selected randomly and values k of anonymity ranging from 10 to 100 nodes per group also selected randomly. As per the figure, one can clearly notice that simulation values are within the range of expected values as per the analytical model.

Once showed the validity of the analytical model, in the following we evaluate the goodness of our proposal applied to the Spain's vehicle fleet.

B. Application to the Spain's car fleet

In this section we present the results obtained from the analytical model for the Spain's vehicle fleet in 2012[6], which rises up to 31,203,203 vehicles.

Figure 2 shows the mean and standard deviation of false positives obtained for a varying number of concurrent attackers

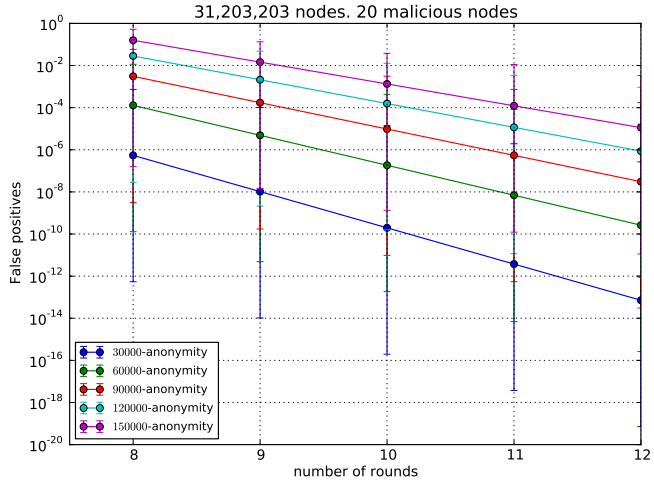

Fig. 3. Mean and STD of false positives vs rounds with 20 concurrent malicious nodes for a varying value of k-anonymity

ranging from 1 to 50 and k-anonymity from 10,000 to 150,000.

First impressions from the figure are that, as expected, the performance of the protocol increases with the number of nodes (now is 30,000 times greater than in section V-A). This is because, for the same level of k-anonymity, the number of groups increases, and therefore the probability of an honest user coinciding in every group with a malicious node diminishes.

Following the above reasoning and the results in the figure, the protocol performance decreases with the value of k-anonymity. That is to say that the more anonymity the worse performance. However, the average number of false positives is bounded to less than 1 (less than $3.205 \cdot 10^{-6}\%$) even for pretty high values of k-anonymity, which may satisfy most of the anonymity policies.

Figure 3 shows the evolution during time (rounds) of the number of false positives (mean and standard deviation) for 20 concurrent attackers and k-anonymity ranging from 30,000 to 150,000. The conclusion here is clear: the average number of false positives decreases in almost two orders of magnitude in every single round; and this is very promising result. Obviously, more rounds mean more time to detect attackers, but just a few rounds make negligible the probability of flagging an honest user as an attacker.

VI. Conclusions

In a VANET, every vehicle must own valid credentials issued by a trusted third party or CA in order to allow users to authenticate data. However, the use of credentials linked to vehicles may violate users' privacy as long as it facilitates the vehicle tracking. This is a risk that can be overcome by means of anonymity schemes.

The use of anonymity schemes can mitigate the risk of vehicles tracking; however, it comes at the expenses of: on the one hand, preventing VANET authorities (CA and TAs) from identifying malicious users and revoking them from the

network; or on the other hand, to discard the anonymity of users in front of the authorities thus to allow their revocation.

In this paper we have presented a method based on k-anonymity that preserves the vehicles' anonymity both against other vehicles/users of the system and the authorities while still being able to track malicious users and revoke them.

For the evaluation of the proposal, we have derived an analytical model for the number of false positives and negatives in several scenario conditions, and we have validated the model by simulation. Then we have analyzed the performance of our proposal with a real country vehicle fleet (the Spanish one) leading to quite promising results in terms of malicious tracking efficiency while providing good levels of k-anonymity. Provided method can effectively identify malicious users whenever they misbehave a given number of times with almost negligible rates of false positives.

Other directions for future work include the development and evaluation of possible attacks to the system, in parallel with an investigation of more efficient and secure schemes.

ACKNOWLEDGMENT

This work was partially supported by the Spanish *Ministerio de Economía y Competitividad*, the Spanish *Comisión Interministerial de Ciencia y Tecnología*, the Spanish *Ministerio de Industria, Energía y Turismo*, the *Generalitat de Catalunya* and European FEDER funds under the projects: MINECO TUERI (TIN2011-25452), CICYT TAMESIS (TEC2011-22746), INNPACTO DEPHISIT (IPT-2012-0585-370000), CONSOLIDER ARES (CSD2007-00004), as well as the grant 2009 SGR-1362 to consolidated research groups, the funding of which is gratefully acknowledged.

REFERENCES

[1] TomTom user data sold to Dutch police, used to determine ideal locations for speed traps. http://www.engadget.com/2011/04/27/tomtom-user-data-sold-to-danish-police-used-to-determine-ideal/.

[2] P. Caballero-Gil, C. Caballero-Gil, and J. Molina-Gil. How to build vehicular ad-hoc networks on smartphones. *Journal of Systems Architecture*, 59(10, Part B):996 – 1004, 2013.

[3] E. C. CRYPTOGRAPHY. The advantages of elliptic curve cryptography for wireless security. *IEEE Wireless Communications*, page 63, 2004.

[4] V. Daza, J. Domingo-Ferrer, F. Sebe, and A. Viejo. Trustworthy privacy-preserving car-generated announcements in vehicular ad hoc networks. *Vehicular Technology, IEEE Transactions on*, 58(4):1876–1886, May 2009.

[5] M. de Fomento Gobierno de España. Encuesta de movilidad de las personas residentes en España (movilia 2006/2007). http://www.fomento.es/MFOM/LANG_CASTELLANO/ESTADISTICAS_Y_PUBLICACIONES/INFORMACION_ESTADISTICA/Movilidad/Movilia2006_2007/default.htm.

[6] D. G. de Tráfico (DGT). Parque de vehículos por ccaa, provincias y tipos, 2012.

[7] Google. Google Maps, traffic option. https://support.google.com/maps/answer/ 61454?hl=en.

[8] D. Huang, S. Misra, M. Verma, and G. Xue. Pacp: An efficient pseudonymous authentication-based conditional privacy protocol for vanets. *Intelligent Transportation Systems, IEEE Transactions on*, 12(3):736–746, Sept 2011.

[9] R. Lu, X. Lin, H. Zhu, P.-H. Ho, and X. Shen. Ecpp: Efficient conditional privacy preservation protocol for secure vehicular communications. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages –, April 2008.

[10] K. Oishi, M. Mambo, and E. Okamoto. Anonymous public key certificates and their applications. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 81(1):56–64, 1998.

[11] M. Raya, P. Papadimitratos, I. Aad, D. Jungels, and J.-P. Hubaux. Eviction of misbehaving and faulty nodes in vehicular networks. *Selected Areas in Communications, IEEE Journal on*, 25(8):1557–1568, Oct 2007.

[12] F. Schaub, F. Kargl, Z. Ma, and M. Weber. V-tokens for conditional pseudonymity in vanets. In *IEEE Wireless Communications and Networking Conference (WCNC 2010)*, pages 1–6, Los Alamitos, April 2010. IEEE Computer Society Press.

[13] F. Sebé-Feixas. Privacy in vehicular networks and location-based services. URV Chairs - Summer Courses, June 2007.

[14] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, Oct. 2002.

[15] Sygic. Traffic service. http://www.sygic.com/en/android:traffic.

[16] TomTom. HD traffic. http://www.tomtom.com/en_us/services/live/hd-traffic/.

[17] A. Wasef and X. Shen. Edr: Efficient decentralized revocation protocol for vehicular ad hoc networks. *Vehicular Technology, IEEE Transactions on*, 58(9):5214–5224, Nov 2009.

[18] WAZE. Real-Time maps and traffic information based on the wisdom of the crowd. http://www.waze.com/.

[19] S. Zakhary, M. Radenkovic, and A. Benslimane. The quest for location-privacy in opportunistic mobile social networks. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International*, pages 667–673, 2013.

[20] L. Zhang, Q. Wu, A. Solanas, and J. Domingo-Ferrer. A scalable robust authentication protocol for secure vehicular communications. *Vehicular Technology, IEEE Transactions on*, 59(4):1606–1617, May 2010.