Protecting Vehicular Networks Privacy in the Presence of a Single Adversarial Authority

Chang-Wu Chen*, Sang-Yoon Chang†, Yih-Chun Hu‡ and Yen-Wen Chen§
*Ethereum Asia Pacific, †University of Colorado Colorado Springs,
‡University of Illinois at Urbana-Champaign, §National Central University
Email: *changwu.me@gmail.com*, *schang2@uccs.edu*, *yihchun@illinois.edu*, *ywchen@ce.ncu.edu.tw*

***Abstract*—In vehicular networks, each message is signed by the generating node to ensure accountability for the contents of that message. For privacy reasons, each vehicle uses a collection of certificates, which for accountability reasons are linked at a central authority. One such design is the Security Credential Management System (SCMS) [1], which is the leading credential management system in the US. The SCMS is composed of multiple components, each of which has a different task for key management, which are logically separated. The SCMS is designed to ensure privacy against a single insider compromise, or against outside adversaries. In this paper, we demonstrate that the current SCMS design fails to achieve its design goal, showing that a compromised authority can gain substantial information about certificate linkages. We propose a solution that accommodates threshold-based detection, but uses relabeling and noise to limit the information that can be learned from a single insider adversary. We also analyze our solution using techniques from differential privacy and validate it using traffic-simulator based experiments. Our results show that our proposed solution prevents privacy information leakage against the compromised authority in collusion with outsider attackers.**

I. Introduction

Real-world deployment of Vehicular Ad Hoc Networks (VANETs) [2]–[5] is becoming a reality; for example, the National Highway Traffic Safety Administration (NHTSA) has announced its plan to deploy VANET by 2020 [6]. To ensure message integrity and message authentication, IEEE Std 1609.2-2013 [7] and ETSI TS 102 941 v1.1.1 [8] specifies the deployment of a Public Key Infrastructure (PKI) based on asymmetric cryptography. In 2013, the Security Credential Management System (SCMS) [1] was proposed by the United States Department of Transportation (USDoT) and the Crash Avoidance Metrics Partnership (CAMP) and is the leading candidate design for the V2V security infrastructure in the US. Unlike traditional PKIs, the SCMS needs to support 300 million vehicles and efficiently revoke certificates. The design of the SCMS is to provide pseudonym certificates [9]–[14] to support Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications. In VANET, vehicles communicate with each other by exchanging Basic Safety Messages (BSM), which include the vehicle's information such as its position, velocity, acceleration and so on to enhance transportation safety and efficiency. Each BSM transmission is signed by a pseudonym certificate to prevent attackers from disseminating false messages that mislead benign vehicles. Once the vehicle receives the BSM, it verifies the validity of the signature and ensures the certificate has not been placed on the Certificate Revocation List (CRL). To avoid being tracked, the vehicle switches pseudonym certificates frequently, perhaps every 5 minutes, to protect users' privacy.

The goal of the SCMS is to manage pseudonym certificates in a way that balances privacy with accountability. Outsider attacks are addressed using pseudonym certificates, while insider attacks are limited by dividing the SCMS into multiple components such that no single component has complete information for certificate linkage. To ensure accountability, however, the SCMS must be able to perform efficient certificate revocation to eradicate misbehaving or malicious vehicle from vehicular networks. A vehicle is considered misbehaving if it sends information identified as false by the detection mechanism [15]–[17]. The revocation process requires a device's pseudonyms to be linked in order to hold the misbehaving vehicles accountable. In this paper, we identify a privacy vulnerability of the SCMS when the Misbehavior Authority (MA) is compromised; the MA is in charge of *misbehavior detection* and *certificate revocation*, but as a single point of failure in the SCMS design, violates the design principle of privacy in the presence of insider compromise.

In VANET, safety messages are classified as *Beacon* and *Alert messages* [18], [19]. Some messages are obviously wrong, (for example, driving at 600 miles per hour), and the MA can revoke such a vehicle simply by observing the Beacon. Other attacks are not as readily ascertained, but the MA can aggregate information across multiple reports and revoke the offending vehicle. Depending on the ease of attribution, the procedure to process misbehavior reports is different. For obvious violations, the MA directly revokes the offender's certificate. For violations that require more information, the MA has to record and calculate the number of accusers and misbehaviors, then determine which vehicle is misbehaving and revoke its certificate. Unfortunately, the current design does not specify the procedure for revoking vehicles in the latter case.

In our work, we consider both cases: 1) misbehavior that is readily caught; and 2) misbehavior for which evidence is gathered over time. We now illustrate how an MA can compromise a user's privacy using mechanisms for catching the latter type of misbehavior. Figure 1(a) represents the misbehavior reports received by the MA. Each misbehavior report includes a signed BSM from a misbehaving vehicle.

978-1-5386-0683-4/17/$31.00

These reports are aggregated as a directed graph that captures the accusers and accused; e.g., Alice accuses Bob in the leftmost pair. This graph is a pseudonym-level graph; because a vehicle can switch pseudonyms, the MA cannot determine the mapping from the pseudonym-level graph to the node-level graph. For illustrative purposes, we label each pseudonym in Figure 1(a) such that the first letter of the pseudonym is vehicle's real identity. In this case, the misbehavior reports could be converted into the vehicle identity graph shown in Figure 1(b). Once a vehicle is accused by a threshold number of vehicles, it is considered to be misbehaving; for example, when the threshold is 2, only node C's certificates will be revoked. Unfortunately, in the present SCMS design, the process to resolve a pseudonym-level graph to a node-level graph is not specifically regulated, and consequently leaks information about innocent vehicle identity to the MA.

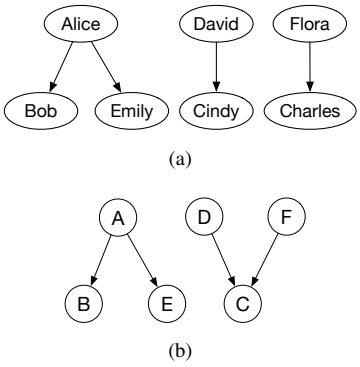


Figure 1: (a) Misbehavior reports. (b) Vehicle identity graph.

A well-defined revocation process for vehicular networks requires a delicate balance between privacy and accountability; a vehicle remains anonymous until misbehavior is detected, and only the misbehaving vehicles' pseudonyms certificates are linked and revoked. In this paper, we analyze the privacy information leakage under two threat models: a passive MA (honest-but-curious) and an active MA (injecting accusations). We show that in the current design, the MA can learn substantial linkage information through the pseudonym resolution process. We then design a revised scheme that offloads certain components of misbehavior detection to another entity. Our revised scheme is privacy-preserving under a single malicious authority, and uses re-labeling and differential privacy to protect privacy. Our scheme also preserves *utility*, a critical metric that drives misbehavior detection, and thus does not affect the outcome of the MA's operations of misbehavior detection and certificate revocation. We define utility to be the correct revocation of misbehaving vehicle certificates.

Our contributions in this paper can be summarized as follows:

- Explore a privacy vulnerability of the SCMS
- Present a privacy-preserving scheme compatible with the SCMS
- Conduct a simulation analysis to validate the effectiveness of our scheme

The remainder of the paper is organized as follows. In Section II, we provide a brief description of the SCMS and describe misbehavior revocation. Section III defines the threat model. In Section IV, we propose our approach and present the simulation results in V. Finally, Section VI concludes the paper.

II. Background

Security and privacy protection in VANET is an active area of research. In currently envisioned VANET protocols, cryptographic signatures and certificates provide accountability and authorization. To limit the ability to track vehicles through their VANET communications, *pseudonym certificates* have been proposed for user privacy protection [13]. Strategies for changing pseudonyms is also necessary [11], as attackers can link multiple pseudonyms to the real identity and track the vehicle [9], [20]. Current VANET specifications require each vehicle to maintain a pool of pseudonyms, and frequently change the pseudonym in use. One challenging problem is to efficiently manage these pseudonyms such that they can be revoked if the node misbehaves; the task of such management falls on a credential management system.

The Security Credential Management System (SCMS) has been proposed by the USDoT and the CAMP, an industry consortium. The interactions between the components in the SCMS architecture are performed automatically, without human intervention. The SCMS supports all basic PKI functions, while adopting a logical, administrative separation that defends against a single insider adversary. The SCMS design relies on the existing pseudonym certificate scheme to ensure privacy from outsider attackers. Furthermore, the SCMS is responsible for misbehavior detection and certificate revocation.

A. SCMS components

Figure 2 describes the SCMS architecture. For brevity, we focus our discussions on the SCMS components that are involved in misbehavior detection and certificate revocation, since our work focuses on these aspects; the interested reader can learn about other components from the paper [1].

- **Misbehavior Authority (MA)** – processes misbehavior reports to determine misbehaving vehicles based on a global misbehavior detection mechanism (still being developed). Once a vehicle is determined as misbehaving, the MA begins the revocation process. First, it revokes the enrollment certificate[1] by adding the enrollment certificate to the blacklist, so the vehicle is not able to obtain new pseudonym certificates. Further, the MA asks for linkage information to update the Certificate Revocation List (CRL); once the CRL is disseminated, legitimate vehicles can ignore the newly revoked vehicle.
- **Linkage Authority (LA)** – maintains linkage values that can link together multiple certificates from the same vehicle. The Linkage Authority functionality is divided

[1]Due to limited storage on the vehicles, each vehicle maintains certificates for a limited time, and uses its enrollment certificate to obtain future certificates.

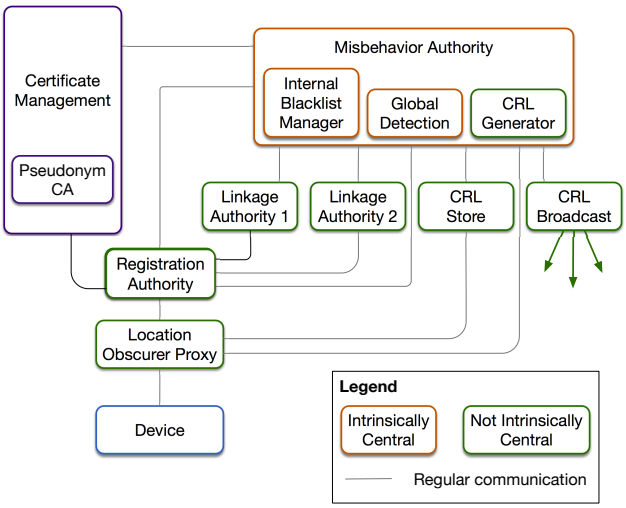


Figure 2: The SCMS architecture adapted from [1].

between two administrative entities, such that no single entity can definitively link two certificates. The LA interacts with the MA for pseudonym resolution.

- **Registration Authority (RA)** – receives and validates requests for vehicle devices' pseudonym and, upon validation, forwards them to the PCA.
- **Pseudonym Certificate Authority (PCA)** – issues short-lived pseudonyms as requested by the RA, but does not know the identity of the corresponding vehicle.
- **Location Obscurer Proxy (LOP)** – all communications between deployed vehicles and the SCMS system transits the LOP, which removes network address information (such as IP addresses) from communications between a vehicle and the SCMS system, and shuffles messages to reduce information leakage from timing.

A design goal of the SCMS is that pseudonyms cannot be linked even in the presence of one malicious authority. For example, in pseudonym generation, the RA does not know which pseudonym certificates correspond to which request and the PCA does not know which request corresponds to which vehicle; in linking, there are multiple LAs. However, for accountability, the MA is able to perform pseudonym resolution and certificate revocation. Each pseudonym certificate includes a linkage value as shown in Figure 3, which is generated from PCA by the pre-linkage value. The pre-linkage value is calculated by the LA through the linkage seed. The MA can submit two individual requests to the PCA and the LA to obtain linkage information (e.g., linkage seed), which allows the MA to link multiple certificates to the same vehicle. The pseudonym resolution is described in detail in Appendix A.

After the MA obtains linkage information for vehicle identity resolution, the MA can run the *Misbehavior Detection Scheme* to decide which vehicle device needs to be revoked. However, no one supervises the MA's requests for linkage information. Furthermore, the MA has sole authority for certificate revocation, which allows the MA revoke a certificate and cause privacy

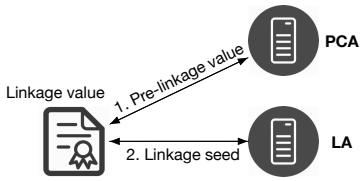


Figure 3: Resolve Linkage value to Linkage seed.

loss. In this paper, we focus on the privacy loss from the MA.

B. Misbehavior Detection Scheme

Since there are no published, complete proposals for a Misbehavior Detection Scheme (MDS), we adopt a simple threshold-based scheme; if the number of accusations against a device exceeds a threshold, then we deem the accused to be misbehaving. Any node can accuse any other node at most one time. In Figure 4(a), Bill is accused once because pseudonyms having the same first letter are from the same vehicle. However, Bill is accused three times in Figure 4(b). By limiting the number of distinct accusations to one, we can limit the slander attack where an attacker falsely accuses some nodes many times. Note that MDS is an active area of research [15]–[17] and beyond the scope of this paper. So MDS here is used to address how the certificates are chosen, not for its performance.

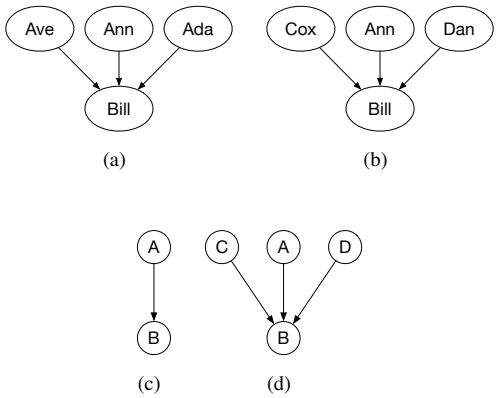


Figure 4: Pseudonym misbehavior counting.

In evaluating our approach, we also consider that outsider attacks can collude to make false accusations. In the MDS that we design, we attempt to minimize the impact of false accusations. Figure 4(c) and Figure 4(d) are the node-level graphs that the LA generates from the pseudonym-level graph sent by the MA, Each node in the node-level graph is the linkage seed. Henceforth, we call the linkage value the *pseudonym* and the linkage seed as the *identity*.

C. Analysis

In our work, we consider the SCMS support for two modes of misbehavior detection and revocation: vehicles where violation is immediately apparent, and vehicles where multiple misbehavior reports are needed to ascertain misbehavior. In the

former case, the MA can examine a misbehavior report and send the misbehaving certificate to the LA for immediate revocation; in such cases, the MA can internally verify the correctness of the misbehavior report. In the latter case, the MA can only learn information about which nodes are accusing which nodes through the LA. This means that the LA must return a node-level graph (or some functional equivalent) in response to a pseudonym-level graph. However, the node-level graph reveals the identity of benign vehicles. In the next sections, we only consider the latter case, since it presents a greater challenge to ensuring vehicle privacy.

III. Adversary Models

In this section, we describe the adversary models used in this paper and identify a privacy vulnerability of the SCMS when the MA is compromised.

A. Honest-But-Curious Adversary

One adversary model is an honest-but-curious attacker; this attacker is a passive attacker that does not violate the protocol, but is interested in inferring more information from the revocation process. Because this attack is passive, the MA does not attempt to insert additional misbehavior reports for learning. The MA simply observes the node-level graph that the LA returns in pseudonym resolution, which a non-attacking MA (or a passive adversary) uses to count the number of accusations as illustrated in section II-B. Because the LA directly returns the node-level graph to the MA, which associates with the linkage seeds, the identity of the well-behaved vehicles are thus revealed.

B. Malicious Adversary

In contrast to the honest-but-curious adversary, an MA could also be an active and malicious attacker. This MA can manipulate the data it sends to the LA to infer more information, and the MA might revoke a node's certificates maliciously, knowing that a few extra revocations will have minimal system impact. We now illustrate a de-anonymization strategy that allows the MA to learn vehicle linkage even if the LA returns linkage information only for misbehaving vehicles. Given a conviction threshold 3, the MA intentionally inserts two fake accusations from David and Flora into the misbehavior reports as shown in Figure 5. The MA can now learn node B's linkage seed, since that linkage seed is needed for revocation; furthermore, the MA can perform this attack an arbitrary number of times. For this attack, our goal is to design a defense mechanism to limit the MA's knowledge while ensuring that misbehaving nodes can still be revoked. The challenges include 1) how to prevent the MA from learning information; and 2) how to be cautious to false accusation from inside attack. Moreover, we also show such an MA may collude with outsiders to perform its attack.

IV. Approach

In this section, we propose some solutions to limit MA's ability and to strengthen system privacy protection.

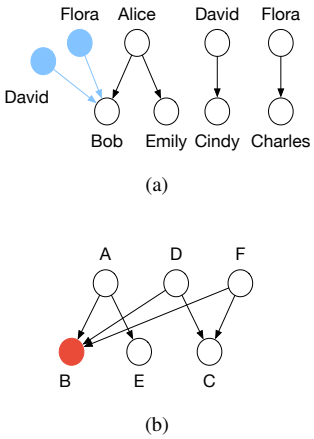


Figure 5: De-anonymization attack.

A. Graph Relabeling

We assume that if an MA sends a pseudonym-level accusation graph to the LA, and receives in response a node-level accusation graph, that the returned graph is relabeled. Specifically, the labels in the node-level accusation graph should be random, so that each label is free of any association with node identity (e.g., linkage seed) or with labels from a previous node-level graph. Furthermore, the order of the nodes in each graph should be randomized, so that information is leaked only through graph structure.

Specifically, the MA sends the pseudonym-level graph to the LA. Then, the LA transforms the pseudonym-level graph to a node-level graph. Each node in the node-level graph, as shown in Figure 6(a) is given with a new identifier at random, and only the LA knows the mapping. The perturbed graph is called a *re-labeled graph* as shown in Figure 6(b). Re-labeled graphs preserves the original graph structure so the MA is still able to run the MDS on it. Node C is required to be revoked if the conviction threshold is set to 2, and so is node 2. Then, the MA will make a second query to the LA for node 2's linkage seed. When the MA requests linkage seeds, the LA or an additional SCMS entity can verify the validity of each request by examining the re-labeled graph. Figure 6(c) illustrates the process of the two-phase query for the MA to obtain linkage information.

B. Differential privacy

Though graph relabeling leaks information only through the graph structure, including changes in the graph structure, the MA can launch a de-anonymization attack by inserting edges to the accusation graph to infer additional information. Fundamentally, the graph structure preserves correlations that allow the MA to infer information.

In addition, common graph perturbation schemes such as k-anonymization, random walk or isomorphism [21]–[24] cannot resist an active MA. Because misbehavior detection continues across time, the MA can insert multiple false edges, observe

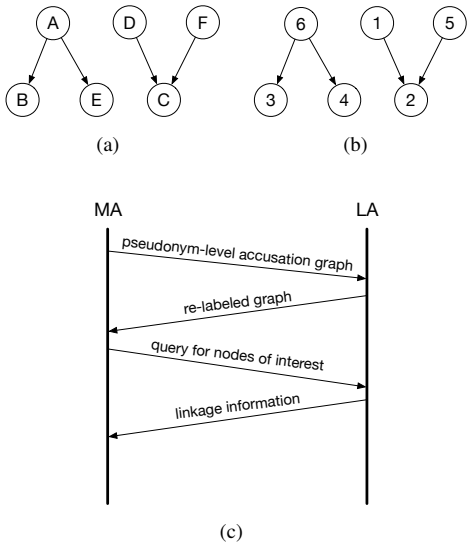


Figure 6: (a) Node-level graph. (b) Re-labeled graph. (c) Two-phase query.

Table I: Misbehavior counting table.

Node	A	B	C	D	E	F
Number of accused	0	1	1	0	1	0

the resulting pseudonym graph, and repeat the process until it deanonymizes the pseudonyms being observed.

We propose that the entire Misbehavior Detection Scheme is executed at the LA, which returns the tabular data rather than the graph data to the MA. The tabular data should include a random identifier for each node, and the number of times that node has been accused, as shown in Table I. Furthermore, to limit an active MA's ability to learn based on changes in this table from query to query, we perturb the data by *differential privacy*. Then, the LA returns the tabular data to the MA, which determines the nodes of interest. Finally, the MA requests linkage seeds to finish the revocation process. Differential privacy gives a strong guarantee that the statistical information of the released database is nearly the same, whether a single record is in the database or not. Therefore, the participant's privacy is protected and its information can not be learned.

Definition 1 (Differential privacy [25], [26]): A randomized mechanism $\mathcal{A}$ maintains ϵ-differential privacy if for any datasets $\mathcal{D}_1$ and $\mathcal{D}_2$ differing on a single record, and for any possible sanitized datasets $\mathcal{S} \in Range(\mathcal{A})$,

$$Pr[\mathcal{A}(\mathcal{D}_1) = \mathcal{S}] \leq e^{\epsilon} \times Pr[\mathcal{A}(\mathcal{D}_2) = \mathcal{S}] \tag{1}$$

where the probability is taken over the randomness of $\mathcal{A}$. The parameter ϵ is called the privacy budget.

Definition 2 (Global sensitivity): Let f be a query function which maps database $\mathcal{D}$ to the statistical information in real numbers. For any function $f : D \rightarrow \mathbb{R}^d$, the sensitivity of f is

$$S_G(f) = \max_{\mathcal{D}_1 \Delta \mathcal{D}_2 = 1} \|f(\mathcal{D}_1) - f(\mathcal{D}_2)\|_1 \tag{2}$$

where $\mathcal{D}_1 \Delta \mathcal{D}_2 = 1$ represents for all adjacent databases differing in at most one record. Smaller sensitivities result in less distortion. For example, if the query is a counting function, such as how many records in the database have property p, the sensitivity $S_G(f)$ is 1 because the removal or addition of a single record only affects the result by 1. The sensitivity depends on the query function. One of the techniques that provides differential privacy is the use of the *Laplace mechanism*. Noise is generated by Laplace distribution and is added to the value of the query function f. The probability density function of Laplace distribution is

$$p(x; \lambda) = \frac{1}{2\lambda} \exp\left(-\frac{|x|}{\lambda}\right) \tag{3}$$

where $\lambda = \frac{S_G(f)}{\epsilon}$.

Theorem 1 (Laplace mechanism): For any function $f : D \rightarrow \mathbb{R}^d$, the computation is

$$\mathcal{A}(\mathcal{D}) = f(\mathcal{D}) + Lap(\frac{S_G(f)}{\epsilon})^d \tag{4}$$

and maintains differential privacy. Note that if λ is increasing, based on the Laplace distribution the $Lap(\frac{S_G(f)}{\epsilon})$ curve is flatter, which results in higher noise.

C. Detection of collusion attacks

In Section IV-B, we introduce differential privacy to use noise in the tabular data to reduce the MA's ability to learn. However, this scheme does not satisfy the Misbehavior Detection Scheme design requirements discussed in Section II-B, because simply counting misbehaviors is not reliable; in particular, the identities of the accusing nodes may impact collusion detection at the MDS. The relabeled graph preserves the graph structure and correlations, which is useful for detecting collusion attacks. When the LA returns the tabular data, it should filter out false accusations from colluding attackers so the MA can ignore such false accusations.

Now we discuss our scheme to detect collusion attacks. Figure 7 is the out-degree distribution of the accusers from our simulation, which will be described in detail in Section V. The level of correlation represents the attacker's collusion group size and the level of accusation represents the frequency of attacker accusations. In this example, the attacker does not collude with other vehicles and the probability of accusation is 0.05. The malicious vehicle can generate a large number of false accusations; however, by considering the out-degree of each node, we can limit the number of false accusations that the attackers can make.

The accuser is suspicious if its accusation rate is unreasonable. We use the mean value of the out-degree as the reference to filter out accusations from suspicious vehicles. In our scheme, first, the MA provides the accusation graph to the LA. Next, the LA analyzes the accusation graph and removes accusations from suspicious vehicles. Then the LA calculates the value of accusation counting with differential privacy, and returns the results in tabular form to the MA. The MA can compare

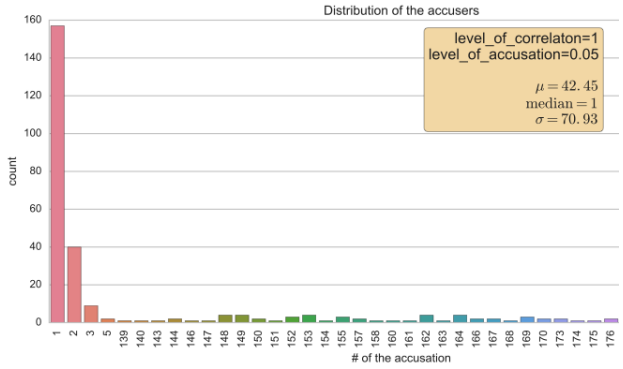


Figure 7: The out-degree distribution of the accusers.

the number of *non-suspicious accusations* to any detection threshold. When the MA wishes to revoke a table entry, the MA provides the LA with the index in the table, and the LA returns the linkage seed corresponding to that table entry. The MA then proceeds with revocation as before.

V. Experimental Evaluation

To analyze the utility of our proposed scheme, we need a model of the accusation graph structure. We also evaluate the performance of the collusion attack detection scheme. However, it is difficult to get the data from actual traffic, especially because misbehavior detection operates over long periods of time, on a time scale from weeks to years. For our evaluation, we use the SUMO [27] road traffic simulation to generate traffic patterns. Such traces are designed to model human activity, as opposed to other forms of synthetic traffic generation which simply randomly generate traffic. Most drivers drive back and forth from home to office every day, so the probability that two vehicles will encounter each other is non-uniform. In addition, each driver also participates in some random activities, such as visiting, shopping, dining out and so on. Compared to random traffic, SUMO generates more realistic driver behavior.

A. Simulation settings

We simulate an area with 10000 inhabitants, 2000 households, and an average of 2.28 cars per household. We run the simulation for 3 weeks to generate trips. In the first experiment, we generate the accusation graphs for observation. We deploy 5% malicious vehicles in our simulation scenario. Each malicious vehicle misbehaves with 1% probability. While a vehicle misbehaves, it can be detected by each of its neighbors with 1% probability. Furthermore, we consider a small false positive rate of 10^{-5}. These parameter settings are summarized in Table II. Based on this traffic and these probabilities, we generate an accusation graph. There are a total of 248 accusations as shown in Figure 8 and the simulation results are summarized in Table III.

B. Differential Privacy's Impact on Utility

In this section, we examine the impact of differential privacy on utility based on the accusation graph we developed through

Table II: Simulation parameters.

Parameter	Value
Percentage of malicious vehicle	5%
Probability of misbehavior	1%
Probability of detection	1%
Probability of false accusation	1e-3%

Table III: Results of Experiment 1.

Number of vehicles	14337
Number of trips	66234
Number of malicious vehicle	716
Number of total accusation	248

experiment 1. Differential privacy gives a strong privacy guarantee, using noise to reduce the amount of information an attacker can infer, while minimizing the impact on the released result. Based on the accusation graph in Figure 8, we apply the differential privacy method on the number of accusations against each node. As mentioned before, the sensitivity of a counting query is 1, which results in less noise, or equivalently, a greater level of privacy for the same amount of noise. The use of differential privacy ensures that a malicious MA learns only a bounded amount, even if the MA inserts false accusations. In order to measure the utility under differential privacy, we calculate True Positive Rate (TP) and False Positive Rate (FP) under varying privacy budgets ϵ with fixed conviction threshold $\tau = 10$ and present the results in Table IV. TP is defined as the number of correctly revoked certificate divided by the total number of revoked certificates. FP is defined as the number of incorrectly revoked certificates divided by the total number of unrevoked certificates. As ϵ increases, λ decreases, reducing the noise magnitude. Because larger privacy budgets reduce noise, the true positive rate is getting improves and the query result is more accurate.

Although differential privacy is a powerful tool, the answer might leak privacy if queries are too frequent. The total privacy budget is given by $\epsilon = \sum_i \epsilon_i$. Normally, when the privacy budget is exhausted, the user can not make further queries. However, in the SCMS, each set of pseudonym certificates has a limited life (usually a few years), and no queries need to be made regarding a set of pseudonym certificates that have already expired. Furthermore, our data set varies over time, and only a single entity (the MA) performs queries; both of these factors result in a slower privacy leakage than the worst-case model adopted in differential privacy.

Table IV: Utility vs. ϵ.

ϵ	TP	FP
0.1	100.00%	32.66%
0.2	100.00%	14.11%
0.3	100.00%	4.03%
0.4	100.00%	2.02%
0.5	100.00%	1.61%
0.6	100.00%	0.81%
0.7	100.00%	2.02%
0.8	100.00%	0.40%
0.9	100.00%	1.61%
1.0	100.00%	0.00%

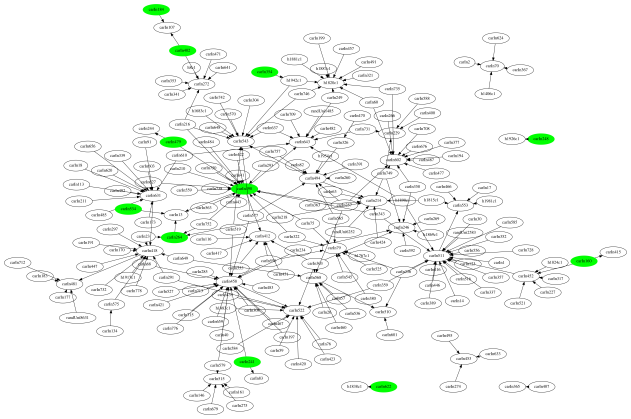


Figure 8: 237 positive accusations and 11 false accusations in green.

C. Performance of collusion attack detection

In experiment 2, we evaluate the performance of our scheme for detecting collusion attack under various attacker behavior. We use the same parameter settings as our previous experiment (Table II), and introduce 0.5% (71) colluding attackers. We explore the attack space across levels of correlation and levels of accusation. A correlation of k represents that whenever the attackers wish to accuse a particular vehicle, k attackers work together on that accusation. The accusation level reflects the probability of accusation; if the accusation level is 0.01, then whenever any attacker encounters any vehicle, the attacker has a 1% chance of accusing the other vehicle. Accusation probability is measured on a per-attacker, per-encounter basis; the number of BSMs received during any encounter is independent of the probability of accusation. An attacker that has a higher accusation probability could potentially cause a larger number of revocations; however, it also means the colluding attackers will send many more misbehavior reports and will consequently be more likely to be caught.

Our simulation shows the performance of attackers that makes false accusations against innocent vehicles in an attempt to have them revoked from the system. Figure 9 shows the results of our simulation runs across various levels of correlation and accusation. The result shows that at increasing levels of collusion, more legitimate vehicles are revoked. We varied the probability with which the legitimate vehicles detected the misbehaving vehicle, and re-ran various levels of correlation and accusation, and found similar results as shown in Figure 10; in particular, higher levels of collusion always results in a more powerful attack. Table V shows the performance of our proposed scheme by showing how filtering false accusations varies as the level of accusation varies from 10^{-5} to 0.15 and the level of correlation from 1 to 30. For this experiment, we choose a conviction threshold $\tau = 5$ because attacking nodes are rarely accused (due to our low probability of detection), and because our simulation covers only three weeks, limiting the number of times that malicious vehicles are actually accused. In general, τ should be set to balance true positives and false positives. Because an attacker that wants to revoke many innocent vehicles will have a abnormally large out-degree, our scheme readily detects malicious attackers, dramatically reducing the number of incorrectly revoked certificates. In addition to maintaining high accuracy, our scheme effectively reduces the threat from collusion attackers. Note that the true positive rate is low as the level of accusation is low. The reason is the sophisticated attacker tries to avoid the detection from the system. In other words, the system cannot distinguish the malicious attacker from the out-degree distribution. The misbehavior revocation procedure, however, is executed periodically. The attacker cannot be caught at this time, but still be possible to be detected next time.

Our defense relies on detecting such attacks through the use of out-degree; unlike other applications users cannot create many false accounts to launch Sybil attacks Thus, an attacker that accuses too frequently will be an outlier, and an attacker that does not accuse very frequently can only affect a small number of legitimate vehicles.

VI. Conclusion

In this paper, we have studied the privacy of the SCMS and mitigated the problem of an insider attack by a single adversarial authority. In the existing SCMS design, the MA has extensive power and knowledge. Furthermore, outside attackers can collude to cause revocations of legitimate nodes. We propose an approach that offloads the graph analysis part of misbehavior detection to the LA so the LA can minimize privacy leakage, and enforce that the MA does not arbitrarily revoke nodes. In our scheme, we adopt differential privacy to perturb the data so that the MA can infer only bounded information from each query. Our evaluation shows that differential privacy can preserve privacy while retaining the detection power of the underlying misbehavior detection scheme. In addition, our scheme is able to prevent the collusion attack where attackers collude to revoke legitimate nodes.

VANET is close to actual deployment. It is expected to save lives from car accidents and provide comfortable and convenient applications. However, to ensure consumer acceptance, the certificate authority design must ensure that no single misbehaving authority can compromise vehicle privacy. Our paper reflects some of the design problems in the current SCMS proposal, which we hope will be addressed before deployment.

References

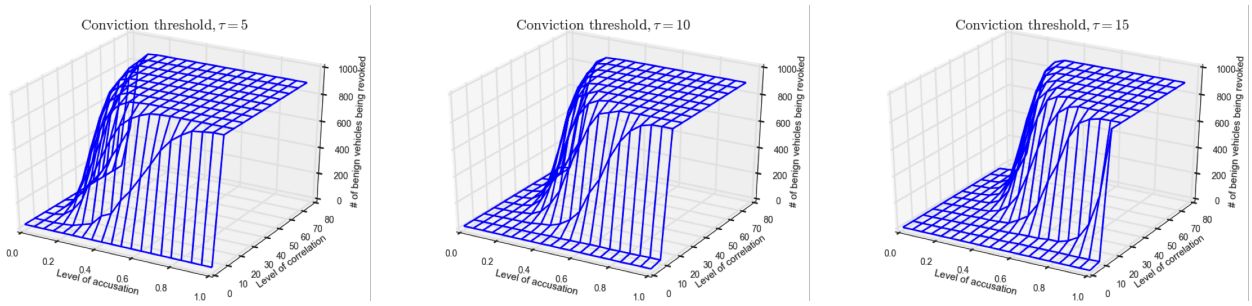
[1] W. Whyte, A. Weimerskirch, V. Kumar, and T. Hehn, "A security credential management system for v2v communications," in *Vehicular Networking Conference (VNC), 2013 IEEE*, Dec 2013, pp. 1–8.

[2] F. Li and Y. Wang, "Routing in vehicular ad hoc networks: A survey," *Vehicular Technology Magazine, IEEE*, vol. 2, no. 2, pp. 12–22, June 2007.

[3] H. Hartenstein and L. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, June 2008.

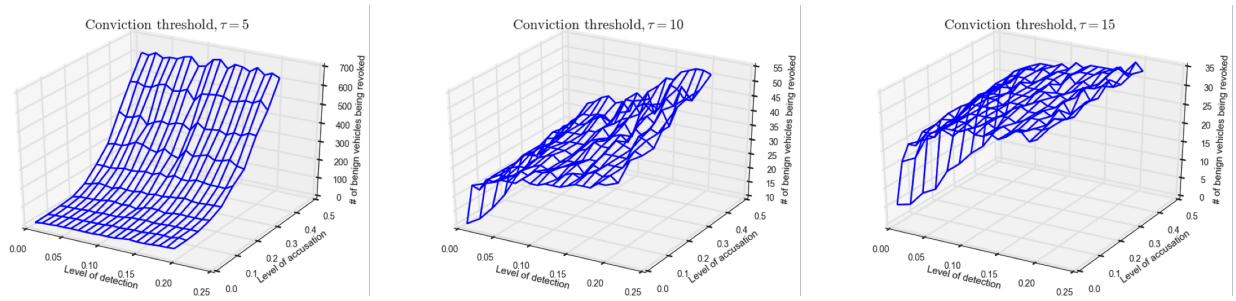
[4] S. Zeadally, R. Hunt, Y.-S. Chen, A. Irwin, and A. Hassan, "Vehicular ad hoc networks (vanets): status, results, and challenges," *Telecommunication Systems*, vol. 50, no. 4, pp. 217–241, 2010.

[5] M. Khodaei, H. Jin, and P. Papadimitratos, "Towards deploying a scalable amp; robust vehicular identity and credential management infrastructure," in *2014 IEEE Vehicular Networking Conference (VNC)*, Dec 2014, pp. 33–40.



(a) (b) (c)

Figure 9: Level of threat with various level of correlations and accusations.



(a) (b) (c)

Figure 10: Level of threat with various level of detections and accusations. (Under the level of correlation is 10.)

Table V: The performance of the collusion attack detection scheme.

	Level of accusation													
Correlation	1e-5		1e-4		1e-3		0.01		0.05		0.1		0.15	
	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
1	0.00%	0.38%	5.63%	0.38%	84.51%	0.38%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
4	0.00%	0.38%	35.21%	0.38%	98.59%	0.01%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
8	4.23%	0.38%	81.69%	0.38%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
12	2.82%	0.38%	92.96%	0.38%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
15	9.86%	0.38%	84.51%	0.08%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
19	8.45%	0.38%	95.77%	0.08%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
23	18.31%	0.38%	98.59%	0.08%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
26	22.54%	0.38%	98.59%	0.01%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
30	22.54%	0.38%	100.00%	0.01%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%

[6] C. Aubernon, "NHTSA Unveils Plan Instituting New V2V Technology By 2020," http://www.thetruthaboutcars.com/2014/08/nhtsa-unveils-plan-instituting-new-v2v-technology-by-2020/, 2014, [2014-08-19].

[7] "Ieee standard for wireless access in vehicular environments - security services for applications and management messages," *IEEE Std. 1609.2-2013*.

[8] "Intelligent transport systems (its); security; trust and privacy management," *ETSI TS 102 940 V1.1.1 (2012-06)*.

[9] E. Fonseca, A. Festag, R. Baldessari, and R. L. Aguiar, "Support of anonymity in vanets - putting pseudonymity into practice," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, March 2007, pp. 3400–3405.

[10] A. Studer, E. Shi, F. Bai, and A. Perrig, "Tacking together efficient authentication, revocation, and privacy in vanets," in *2009 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, June 2009, pp. 1–9.

[11] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, "Privacy in inter-vehicular networks: Why simple pseudonym change is not enough," in *Wireless On-demand Network Systems and Services (WONS), 2010 Seventh International Conference on*, Feb 2010, pp. 176–183.

[12] J. Haas, Y.-C. Hu, and K. Laberteaux, "Efficient certificate revocation list organization and distribution," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 3, pp. 595–604, March 2011.

[13] J. Petit, F. Schaub, M. Feiri, and F. Kargl, "Pseudonym schemes in vehicular networks: A survey," *Communications Surveys Tutorials, IEEE*, vol. 17, no. 1, pp. 228–255, Firstquarter 2015.

[14] M. Khodaei and P. Papadimitratos, "The key to intelligent transportation: Identity and credential management in vehicular communication systems," *IEEE Vehicular Technology Magazine*, vol. 10, no. 4, pp. 63–69, Dec 2015.

[15] 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*, vol. 25, no. 8, pp. 1557–1568, Oct 2007.

[16] S. Reidt, M. Srivatsa, and S. Balfe, "The fable of the bees: Incentivizing robust revocation decision making in ad hoc networks," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 291–302.

[17] S. Ruj, M. Cavenaghi, Z. Huang, A. Nayak, and I. Stojmenovic, "On data-centric misbehavior detection in vanets," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, Sept 2011, pp. 1–5.

[18] F. Bai, H. Krishnan, V. Sadekar, G. Holland, and T. ElBatt, "Towards

characterizing and classifying communication-based automotive applications from a wireless networking perspective," in *in Proceedings of IEEE Workshop on Automative Networking and Applications (AUTONET)*, 2006.
[19] R. Chen, W. L. Jin, and A. Regan, "Broadcasting safety information in vehicular networks: issues and approaches," *IEEE Network*, vol. 24, no. 1, pp. 20–25, Jan 2010.
[20] M. Wernke, P. Skvortsov, F. Dürr, and K. Rothermel, "A classification of location privacy attacks and approaches," *Personal Ubiquitous Comput.*, vol. 18, no. 1, pp. 163–175, Jan. 2014.
[21] K. Liu and E. Terzi, "Towards identity anonymization on graphs," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '08. New York, NY, USA: ACM, 2008, pp. 93–106.
[22] B. Zhou and J. Pei, "The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks," *Knowledge and Information Systems*, vol. 28, no. 1, pp. 47–77, 2010.
[23] J. Cheng, A. W.-c. Fu, and J. Liu, "K-isomorphism: Privacy preserving network publication against structural attacks," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 459–470.
[24] P. Mittal, C. Papamanthou, and D. X. Song, "Preserving link privacy in social network based systems," in *NDSS*, 2013.
[25] C. Dwork, "Differential privacy," in *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)*, ser. Lecture Notes in Computer Science, vol. 4052. Venice, Italy: Springer Verlag, July 2006, pp. 1–12.
[26] R. Sarathy and K. Muralidhar, "Evaluating laplace noise addition to satisfy differential privacy for numeric data," *Trans. Data Privacy*, vol. 4, no. 1, pp. 1–17, Apr. 2011.
[27] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012.

APPENDIX

A. Pseudonym resolution

Pseudonym resolution is an important part involved in linking pseudonym certificates to the vehicle devices. Since no single component in the SCMS has enough information to track the certificate, the MA works with the PCA and the LAs to obtain linkage information for vehicle identity resolution.

The SCMS uses a called *linkage value* to protect privacy. When a new pseudonym certificate is created, the linkage value is associated with the pseudonym certificate for efficient revocation. In the SCMS, the LA is in charge of generating linkage values. There are at least two LAs in the system to limit the information leakage to any single LA. Here we use two Linkage Authorities LA_1 and LA_2 as an example to show how to generate the linkage value. First, before beginning to generate linkage value, each LA calculates the *linkage seed* as follows:

$$ls_i(t) = H(la\_id_i \parallel ls_i(t-1)) \tag{5}$$

where i represents the index of the LA, t represents current time period, $H(.)$ is a one-way non-invertible hash function and $w \parallel v$ denotes concatenation of strings w and v. la_id_i is a 32-bit identifier of LA_i. $ls_i(0)$ is a random 128-bits string as the initial linkage seed picked by LA_i and used to calculate $ls_i(t)$. Then the LA_i calculates the *pre-linkage value* as:

$$plv_i(t,j) = E(ls_i(t), la\_id_i \parallel j) \tag{6}$$

where j represents the index of a set of pseudonym certificates within that period of time and $E(k,m)$ denotes the encryption function with key k on message m. The SCMS suggests that each vehicle has 20 fresh certificates per week to prevent trip tracking, so t represents the week and the value of j ranges from 1 to 20. LAs generate pre-linkage values which, in turn, are relayed to the PCA for linkage value generation; similarly to the certificates themselves, the PCA generates the linkage values while the RA associates them with the pseudonym certificates. Last, *linkage value* is computed by XORing the pre-linkage values from LA_1 and LA_2 as:

$$lv(t,j) = plv_1(t,j) \oplus plv_2(t,j) \tag{7}$$

The last step of generating linkage value is done by the PCA, not by LAs. Because each LA has the initial linkage seed and is able to compute pre-linkage value, the LA can link pseudonym certificates to the same vehicle if the LA can compute the linkage value by itself. As mentioned before, at least two components have to cooperate together to compromise the users' privacy.

B. Revocation processes

The revocation process is the interactions between the MA and other components (i.e., the PCA and the LAs) as shown in Figure 11. Note that the value in the solid rectangle is originally obtained or stored and the value in the dotted rounded rectangle is learned from others. When the MA receives misbehavior reports from vehicles via the LOP, the MA needs to get the corresponding linkage information for resolution. As mentioned previously, at least two components have to collaborate to obtain required information to map the pseudonym certificate to the vehicle device; in this case, the LA and the PCA are the two components. First, the MA communicates with the PCA for pre-linkage values. The MA sends linkage values which are included in pseudonym certificates to the PCA. Because the PCA is in charge of generating linkage values, the PCA can map linkage values to the corresponding pre-linkage values as defined in (7). Second, the MA communicates with the LA for the linkage seeds. Since the MA already has the pre-linkage values from the first step, the MA can make a request to the individual LA for looking up (ls_1, ls_2) from stored information of (plv_1, plv_2) as defined in (6) and thus get the linkage seeds. Finally, the MA adds linkage seeds to the CRL to finish certificate revocation. When the MA obtains a linkage seed, it can identify all future certificates used by the vehicle.

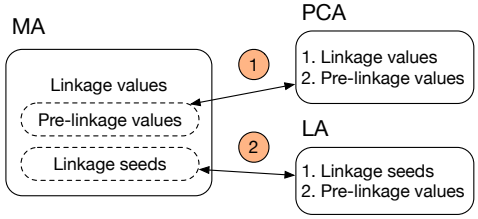


Figure 11: Revocation process.