SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks. 2010; **3**:233–249
Published online 21 September 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sec.143

The impact of key assignment on VANET privacy

Jason J. Haas[1]*,†, Yih-Chun Hu[1] and Kenneth P. Laberteaux[2]
[1] *University of Illinois–Urbana-Champaign, IL, U.S.A.*
[2] *Toyota Research Institute; MI, U.S.A.*

Summary

There are two underlying principles that guide how a vehicular ad hoc network (VANET) is built: there will be misbehavior and the extent to which vehicles can misbehave should be bounded. Additionally, the main use for VANETs currently is to enable safety applications where vehicles' position, velocity, and acceleration are broadcast to other vehicles in the VANET. Combining these guiding principles with this application results in the privacy of vehicles and users being an important concern for VANET design. Safety application messages are signed using keys and therefore linked to vehicles. In this work, we investigate how to assign keys to vehicles in order to preserve privacy and maintain our guiding VANET design principles. Specifically, we investigate the design space where individual keys may be given to multiple vehicles and vehicles may have multiple keys. Through a simple security analysis, we eliminate the case where all keys are common. Through mathematical and logical analysis, we conclude that keys should not be owned by multiple vehicles, that is, keys should be unique to vehicles and vehicles should be given multiple keys. Specifically, we show that it is impossible to provide good privacy and fast revocation when keys are shared among vehicles. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: vehicular networks; privacy; key assignment

1. Introduction

Generally, a VANET consists of vehicles that are assigned keys by a *Certificate Authority* (CA). The CA signs these keys and cryptographically ties them to a vehicle or group of vehicles, creating certificates. Vehicles use these keys to sign messages that they send on a VANET, tying the messages to the vehicle or group of vehicles. Keys may be assigned to more than one vehicle in general, and vehicles may have more than one key. We will provide more motivation for why keys should be assigned to vehicles in VANETs in Section 3.

Privacy is a central requirement for VANET systems. In this paper we investigate how specific key assignment methods affect vehicular privacy in VANETs, and we define privacy in this context as the inability to link a broadcast signature to a vehicle or a usefully-small group of vehicles. Though operating a vehicle already involves significant privacy risks from technologies such as automated toll collection and automatic license plate recognition [1,2], VANETs are unique in that vehicles use relatively long-range, non-line-of-sight, radio communications to very accurately advertise their positions in safety beacon messages.

*Correspondence to: Jason J. Haas, University of Illinois–Urbana-Champaign, IL, U.S.A.
†E-mail: jjhaas2@illinois.edu

Furthermore, vehicles sign messages using the keys assigned to them. Periodically, these messages include either a certificate, certificate digest, or certificate chain, which attests to the validity of the key. Most of the contemplated VANET designs require VANET vehicles to regularly broadcast their certificate(s). Even if a certificate contains only the vehicle's pseudonym, if care is not taken in how the VANET is designed, and specifically, in how keys are assigned to vehicles, it may be possible to remotely track individual vehicles in a VANET using their certificates and signatures. Additionally, in this work, we require an efficient revocation process (see Axiom 3.2 and Corollary 3.4 below). Given this requirement, it may not be possible to provide the same level of privacy with a VANET as what privacy would exist without the VANET. We evaluate the privacy provided by key assignment methods, the robustness of those methods as defined by the properties we describe in Section 4, and the ability of those methods to maintain VANET security goals.

We can divide attempts to preserve privacy into preserving privacy: (1) from the CA or (2) from non-CA entities (e.g., other vehicles). There are many reasons to be interested in providing privacy to vehicles, including protection from big-brother behavior of governments and corporations, and auto manufacturer concerns of acceptability of VANETs to consumers. We will discuss the motivation for providing privacy when we discuss the details of CA privacy in Section 4.1 and non-CA privacy in Section 4.2. However, our primary concern in this work is to analyze the viability of various key assignment methods for a VANET with respect to the privacy and security these methods provide.

Privacy is significantly affected by how keys are assigned in a VANET. However, privacy may also be impacted by other factors outside of key assignment that will affect the level of privacy a vehicle can maintain. Specifically, it is possible to correlate broadcast VANET data with information obtained through other methods, such as cameras, and it may be impossible to defend against determined attackers who use both sources of information. We will omit a detailed discussion of this problem of tracking because the problem exists independent of the key assignment method used, and we are only concerned with privacy and security issues arising from key assignment in this paper.

We organize this paper as follows. In Section 2 we review related work. We present in Section 3 the basic assumptions on which we build our arguments in this paper. We describe in Section 4 the properties that we use to assess various key assignment methods. In Section 5, we will consider what privacy a vehicle can maintain under any key assignment method. Finally, in Section 6 we conclude and summarize our work.

2. Related Work

Generally, privacy for VANETs can be divided into two, not entirely separable, categories: preventing identity information leakage from credentials (e.g., information in certificates that cryptographically bound together) and preventing unrelated third-parties from tracking vehicles and users. These two categories are not entirely separable because if an unrelated third-party can associate a single vehicle with multiple identities or pseudonyms, then it is easier for the third-party to track its target vehicle or user. Our discussion of previous work in this section will reflect this dichotomy.

Raya and Hubaux [3] gave a brief introduction to privacy issues in VANETs. They also propose that vehicles use multiple pseudonyms and that vehicles should change their pseudonyms periodically. These authors and Dötzer [4] suggest that *conditional privacy*, that is, using identifying information that is publicly anonymous but can be linked to a long-term identity (e.g., a VIN) by the CA, should be used in VANETs so that malfunctioning and malicious vehicles can be identified.

Dötzer provided a variety of privacy topics for VANETs from an automobile manufacturer's point-of-view [4]. Specifically, Dötzer notes that privacy is an important topic among vehicle manufacturers because customers may be willing to select a different manufacturer's vehicles based on which vehicle's technology is in-line with the customer's view on privacy matters. In the category of leaking identity information, Dötzer observes that it may be possible to correlate messages over long periods of time such that identity information can be recovered by an interested over-hearing party. He proposes that changing identifiers at multiple layers may be required to thwart privacy-compromising adversaries. Implicit in this proposal is the idea that vehicles will have multiple identifiers, which are likely to be pseudonyms. However, such identifier changes may not be sufficient if an adversary is able to identify a vehicle based on its RF fingerprint. The potential of using analog/RF fingerprinting or radiometric identification to reduce or eliminate user privacy has been investigated by a number of other authors [5–7]. Recently, Brik *et al.* [7] have demonstrated the viability of using radiometric identification for identifying individual, off-the-shelf WiFi cards. For the purposes of this paper, we consider RF fingerprinting simply to

DOI: 10.1002/sec

be a potential mechanism that an adversary could use to link multiple pseudonyms (or keys below). Dötzer also gives example situations of when various entities might be interested in compromising privacy and proposes that mix-zones may be useful in trying to reduce an adversary's ability to link multiple identifiers together.

A number of other authors have proposed and studied using mix-zones to enhance user privacy in VANETs and in other networks where users have high mobility [8–10]. Mix-zones try to prevent a privacy-compromising adversary from linking two pseudonyms to a single vehicle by having vehicles stop transmitting for a period of time in a busy location, that is where there are many vehicles or nodes. Mix zones can work without vehicles ceasing to transmit, but this provides little privacy because vehicles' movements are highly predictable over the period between safety message beacons. The privacy provided is often measured in terms of the size of the *anonymity set*, that is, the number of vehicles in the mix-zone at the time when pseudonyms are changed, among which the adversary can only probabilistically link pseudonyms. The main problem with this approach to mix-zones is that they defeat the purpose of deploying the VANET, that is, vehicles are not able to use the VANET for safety enhancement in critical areas of high road-traffic density, such as, intersections, interchanges, and rush-hour traffic. For the purposes of our investigation in this paper, we are less concerned on how a privacy-compromising adversary can link pseudonyms (or keys in our discussion below) than we are with the effects of such linking on the privacy provided by various key assignment methods.

Sampigethaya *et al.* [11,12] and Jiang *et al.* [13] investigated using silent periods. However, like mix-zones, silent periods interfere with the safety-related goals of deploying a VANET, and silent periods are intended to decrease an adversary's ability to track vehicles using the VANET. Again, for our investigation in this paper, we are less concerned with how identity information might be linked, than we are with investigating the effects of linking on privacy. Thus, we will not discuss silent periods further.

Group Signatures: Others have proposed the use of group signatures for obtaining privacy while maintaining the goal of binding safety beacon information to vehicles, that is, maintaining conditional privacy [14,15]. Parno and Perrig have noted that using group signatures can come at the cost of not being able to attribute misbehavior to a vehicle, thus failing to support conditional privacy [14]. However, they also note that there are group signature schemes that allow a group manager to link a signature to the individual group member. Such a group manager could take one of two forms: an online CA, or another vehicle in the VANET. Important to this discussion is the topic of Road-Side Units (RSUs), which are fixed-position infrastructure DSRC radios that the deploying entity can use to spread and gather information pertinent to the VANET (e.g., distribute traffic and road condition information or gather traffic data for traffic reports). Using an online CA as the group manager is untenable for either or both of the following reasons: roadways will not support group signatures during incremental deployment, that is, when RSUs are being deployed, and/or roadways may not have sufficient RSU coverage due to the cost of deploying such a large number of RSUs. This latter view was shared by Dötzer [4] and more recently by Resendes [16]. Using another vehicle in the VANET as the group manager may open the door to a number of additional privacy-compromising attacks. First, the group manager could be the privacy-compromising adversary, thus, the privacy of the vehicles in the adversary's group is trivially compromised. Second, the mechanism for choosing the group leader may come under attack, for example, if an election algorithm is used, the adversary may pretend to be multiple vehicles to win the election.

Raya and Hubaux have also noted that group signatures are computationally expensive and therefore may not be suitable for VANETs where vehicles have insufficient computing power [15]. Lin *et al.* [17] develop a protocol to use anonymous group signatures to bind heartbeats to a long-term identity that is only known by the CA. In their system, every vehicle can verify the correctness of signatures of every other vehicle's signatures, but only the CA can recover the identity from a vehicle's signatures. Thus, vehicles cannot differentiate other vehicle's broadcasts based on signature information. The authors propose that high-powered processors be used on vehicles to reduce the time required for verifying signatures. Specifically, the authors give 7.2 ms as the time required to verify a signature in their scheme. However, this means that during a heartbeat period (100 ms), fewer than 14 signatures can be verified by any vehicle.

Because of the computational delay imposed by group signature verification, we will not consider further the use of group signatures in our discussion of key assignment and privacy in VANETs.

Mathematical Analysis: Xi *et al.* [18] propose sharing keys among groups of vehicles for the purpose of

DOI: 10.1002/sec

increasing privacy. The authors acknowledge that revocation may be slow (i.e., take many revocation events, which we discuss below) when vehicles have many keys and keys are shared among groups of vehicles. Consequently, they propose that the CA could require vehicles to authenticate messages using multiple keys, thus allowing multiple keys to be revoked per revocation event. The authors begin a probabilistic analysis of the effects of revocation on vehicles that are not the desired target of the revocation event. We expand significantly on this analysis and show that it is not possible to maintain privacy, fast revocation, and protect innocent vehicles from falsely being revoked because of revocation events arising from vehicles that should be revoked.

In previous work [19], we briefly overviewed privacy issues in VANETs, and we began a general mathematical analysis of privacy in relation to the ease of revocation. We continue this work specifically investigating key assignment in this paper.

3. First Principles

In this section, we discuss some of the basic underlying assumptions we make about the operation of a VANET, and the behavior of users and vehicles in a VANET. We will build our arguments about privacy in the latter sections of this paper on the foundation laid in this section. These assumptions are similar to the privacy related requirements given by Dötzer [4].

3.1. Malfunction and Misbehavior

Safety beacons will contain precise information about vehicle positions, velocities, and accelerations. Vehicles will present warnings to drivers based on information gathered by the vehicle from safety beacons in the VANET. These warnings will inform the driver about potentially dangerous situations, such as, hazardous road conditions, excessive speed approaching curves, and emergency braking behavior by other vehicles [20]. If these warnings are presented when there are no hazardous situations, then users may become desensitized to the warnings, or the warnings themselves may pose a safety threat. If an attacker can inject falsified packets into the VANET causing this desensitization or causing accidents because drivers do react to the falsified warnings, significant harm could be done as a result of the VANET instead of resulting in the VANET helping reduce vehicle crashes or the severity of crashes. These undesirable outcomes could also result from erroneous information generated by malfunctioning hardware. It is likely that the perverse attractiveness of these undesirable outcomes will cause some users to intentionally generate falsified safety beacons. It is also likely that hardware will fail. We state these assumptions in the following axiom.

Axiom 3.1. *Some users of a VANET will misbehave or will have equipment that malfunctions.*

We mathematically codify this axiom in our use of f, the fraction of the total vehicles that misbehave or have malfunctioning equipment.

Since vehicles may malfunction and users may misbehave, we want to limit the amount of damage such behavior can cause in a VANET. We state this formally in the following axiom.

Axiom 3.2. *Users that misbehave or vehicles that have malfunctioning equipment should be excluded from the network in order to limit the damage caused by these entities.*

Specifically, we would like to remove such vehicles and users from the network.

3.2. Implementation

As a consequence of these axioms, we need to bind safety beacon information to a certificate. Binding this information in a certificate allows vehicles to differentiate between correctly behaving vehicles that should be trusted and malfunctioning or misbehaving vehicles that should not be trusted. This binding also provides a mechanism for excluding misbehaving vehicles from the VANET. We express this requirement in the following corollary.

Corollary 3.3. *In order to differentiate between trusted and untrusted vehicles and to exclude misbehaving vehicles, safety beacon information should be bound to certificates, belonging to vehicles.*

Vehicles may be assigned multiple certificates so that long-term vehicle behavior, that is, positions, cannot be correlated to a single vehicle. By changing signing keys, and correspondingly certificates, a vehicle may achieve greater privacy. Additionally, vehicles may share certificates so that certificates do not correspond to vehicles in a one-to-one manner. In other words, observing the use of one certificate multiple times does not equate to observing a vehicle multiple times. Similarly, observing a vehicle

DOI: 10.1002/sec

multiple times does not equate to observing the use of the same certificate multiple times.

Asymmetric cryptography should be used to create certificates and to bind safety beacon information to vehicles' certificates. Certificates are a standard way of binding information to an entity. Using asymmetric cryptography makes key distribution easier as compared to symmetric cryptography. If a VANET designer used symmetric cryptography, then vehicles would need to exchange keys with all other vehicles. This could be done either through distribution during manufacturing or through *ad hoc* mechanisms. Vehicles are not a static population; new vehicles will be added to the VANET, and old vehicles will be removed from the VANET. Thus, preloading the certificates of other vehicles onto a new vehicle during manufacturing is not a complete solution. Using *ad hoc* mechanisms to distribute symmetric keys without an online trusted third party introduces security concerns. Consequently, we choose to assume that asymmetric cryptography is used to bind safety beacon information to vehicles in the form of certificates. Additionally, using asymmetric cryptography makes removing vehicles from the network easier.

To remove a vehicle from the network, all of a vehicle's certificates need to be invalidated. To invalidate a certificate, the certificate must be revoked.

Corollary 3.4. *A VANET invalidates a certificate in order to remove or start to remove a malfunctioning vehicle or malicious user's radio from the VANET. This process is called revocation.*

Since vehicles do not know all other vehicles initially, a central authority needs to be established to both attest to the validity of normal vehicles and revoke malfunctioning vehicles or malicious users. This central authority generally takes the form of a CA and signs vehicles' certificates, thus creating a Public Key Infrastructure (PKI) to organize key information.

4. Properties

Using the axioms and the corollaries presented in Section 3, we present properties relating to vehicle privacy, some desirable, some undesirable. A VANET may possess these properties, or they may arise from employing privacy enhancing mechanisms. We assume that vehicles are assigned a number of keys, *d*. This number, *d*, may be larger than 1, allowing vehicles to change or rotate the keys they use to prevent long-term tracking from correlating safety beacons signed with the same key. Each key may be shared among a number of vehicles, *g*, which likewise may be larger than 1. Allowing *g* to be larger than 1 has been proposed as another mechanism for increasing vehicular privacy, making vehicles indistinguishable from other vehicles that have been assigned that same key [18]. We also will use the result from Axiom 3.1 that there is a fraction of revoked vehicles, *f*. Generally, we will apply these properties to key assignment methods, which we will discuss in Section 5.

4.1. CA Privacy

A CA is a centralized organization that signs vehicles' keys for the purpose of generating certificates.[‡] Since a vehicle's key information must pass through some CA, the CA often has privileged information about the identity of a key owner. Because the CA has this information, the CA may become a useful tool for law enforcement. One specific concern with the CA having this information is that government law enforcement may subpoena the CA to enforce the law, including driving violations (e.g., speeding). Auto manufacturers are concerned that without CA privacy, a VANET system will not gain acceptance among buyers for this reason. Other reasons to retain privacy from a CA include possible misuse by an observing government agency or employee for political or personal reasons, and the possibility for unintentional leakage. For the latter reason, consider the following scenario. Privacy from CA(s) is not maintained, and law enforcement uses VANET information to track vehicles to gather evidence for prosecution. This information will need to be retained to possibly be presented in a court of law. A side-effect of having to store this information is the government agency that stores this information will be a centralized target for hackers who want the VANET information. User privacy can be reduced through another mechanism if this information is stored. There have been many cases of government agencies or their employees leaking privacy-sensitive information in unintended ways, e.g., a lost USB flash drive [21,22], a misconfigured web server, or a misplaced organizational laptop [23]. Thus, it may be better not to deploy a VANET where privacy depends on the ability of the CA to keep users' data private. It may be possible for the CA to perform its duties (e.g., revoking vehicles

[‡] In general, it is possible to have multiple CAs for a VANET. For example, one CA might be a national government and another a state or provincial government. CAs might even be non-government organizations such as auto manufacturers.

DOI: 10.1002/sec

and assigning keys to new vehicles) without retaining information sufficient for compromising privacy; however, the CA initially has access to this information since the CA assigns keys, and therefore, we must consider that the CA has access to this privacy-sensitive information.

One way to describe privacy in this context is whether evidence gathered in the form of vehicle position or speed data signed by a valid key would be definitive evidence in a court. If it can be demonstrated that this evidence suffers from an unacceptably high *false positive rate*, such VANET evidence should not be definitive in a court, no more than establishing guilt of the driver because he owns a 'blue car'. Such a false positive rate is an unacceptably high probability that one vehicle could be mistaken for another if the distinction between the two is based on only which keys the vehicles hold. For a VANET, having privacy from a CA means that a CA, even if subpoenaed, would only know imprecise or unreliable information about the owner of any given certificate. We describe the condition of a CA having poor ability to link a key to a VANET vehicle as having 'privacy from a CA'. If the VANET designer's goal is to provide maximum privacy to VANET vehicles (above all other considerations), then privacy from a CA is an attractive attribute. As we show below, maintaining privacy from the CA comes with significant compromises to other design goals.

Increasing g leads to increasing the number of vehicles among which vehicles are indistinguishable to the CA. Thus, vehicle privacy is increased by increasing g.

4.2. Non-CA Privacy

Maintaining privacy from a non-CA entity may be important for other reasons. If it is possible to remotely track vehicles through their VANET messages, a corporation (a non-CA entity) may be able to track a user's shopping habits and correlate that to the user's home address. Thus, the corporation may specifically target VANET users with advertisements. Additionally, VANET messages could be used by private investigators to track the people they are observing.

Generally, by increasing d, we increase privacy from non-CA entities. By assigning a large number of keys to vehicles, non-CA entities will not know if broadcasts signed with two different keys came from the same or two different vehicles based on key information alone. Care must be taken in how certificates are constructed so that vehicles cannot be identified by information included in their certificates. Clearly, the use of other information, even the information signed in the broadcast can be used to reduce a vehicle's privacy. Further discussion of using safety beacon information to track vehicles is beyond the scope of this paper. Increasing d, however, increases the cost of revocation, either in the size of the Certificate Revocation List (CRL) or in the computational cost of revoking the certificate. We will see below that each of the considered methods of key assignment provide non-CA privacy.

4.3. Revocation

Malfunctioning or malicious vehicles' false information may harm correct and innocent vehicles. For example, if a malfunctioning vehicle broadcasts an incorrect position, another vehicle may display a false warning to its driver. A malicious vehicle might also cause other vehicles to think the roadway the malicious vehicle is on is more congested by broadcasting incorrect roadway congestion information. These false reports may cause deceived vehicles to take a different route, leaving the malicious vehicle's roadway less congested. In our following discussions, we do not distinguish between malfunctioning and malicious vehicles, describing them collectively as *untrusted vehicles*.

Our discussion of the basic principles which we assume hold for a VANET, which we gave in Section 3, resulted in our deducing that a VANET should use a PKI structure and asymmetric cryptography to sign packets in order to protect users from unlimited damage from vehicles that are untrusted. The CA assigns keys to vehicles, and vehicles use keys to sign messages. Receiving vehicles assume that messages are valid once they verify the correctness of the message signature and the validity of the associated key[§].

Revocation is a mechanism for protecting correct and innocent vehicles from the effects of untrusted vehicles. Stated more concretely, a CA revokes a key by publicly announcing that the key is no longer valid. Receivers thus distrust any information signed by a key once it learns that a CA has revoked the same key. As we will discuss in detail below, it is possible for a vehicle to have multiple keys. Therefore, it is also possible that a specific vehicle will have some, but not all, of its keys revoked. An untrusted vehicle, if left with at least one unrevoked key, can continue to operate with its valid key. Thus, we stress that a vehicle is not revoked until all d of its keys are revoked.

[§] Vehicles may also filter messages based on message content, if they determine the content to be inconsistent or invalid, as proposed by Golle *et al.* [24]. However, this is orthogonal to our discussion in this paper.

DOI: 10.1002/sec

We define a *revocation event* as the following sequence of events:

I. One or more entities observe and report to the CA that a vehicle, using a specific key , acted in an untrustworthy manner.
II. The CA revokes the reported key, and perhaps other keys assumed to be associated with the reported vehicle (details discussed below). The CA creates an updated CRL containing these newly revoked keys.
III. The CA uses some method to disseminate this new CRL to all vehicles in its area of responsibility.

Roughly speaking, a revocation event occurs when a vehicle 'gets caught' acting in an untrustworthy manner while using one of its keys, causing the CA to revoke one or more of the keys.

We now consider the speed of a revocation process. As we will show below, there is a tradeoff between revocation speed and privacy. While all three steps above contribute to the speed of a revocation process, the first and third steps have previously received consideration in the literature [24–26]. However, to the best of our knowledge, the impact of the second step on privacy and revocation speed has not been considered. Thus, we focus on the second step in our discussion below. Since we focus on the second step, instead of specifying revocation speed in minutes, we define a *faster* revocation process as one that requires fewer revocation events.

Intuitively, when a CA receives a report of a vehicle's use of a specific key linked to untrustworthy behavior, depending on the information the CA has, the CA can respond by revoking only one of the vehicle's keys (i.e., the reported key), revoking all of the vehicle's keys, or some fraction of the vehicle's keys. If the mechanism of revocation is to be useful for protecting correct and innocent vehicles, revocation should be fast. If an untrusted vehicle's keys cannot be revoked quickly, vehicles cannot fully trust the PKI to perform its core mission, i.e., enabling vehicles to identify an untrusted vehicle at the time of contact. The slower the revocation, the larger the window of opportunity for untrusted vehicles to damage to the VANET. Since the creation and maintenance of any PKI is non-trivial and often expensive, it would be unwise to create a PKI with known slow revocation properties.

We restate our assumption that the number of untrusted vehicles is proportional to n, the total population of vehicles. Therefore, after a settling time, it can be assumed that $f \cdot n$ cars are fully revoked.

4.4. Brittleness

If vehicles share keys, i.e., $g > 1$, then each revocation event affects innocent vehicles as well as the target vehicle. If we assume that g is large, that is, a large number of vehicles share any single key, then revoking all keys of an untrusted vehicle impacts a large number of users that shared keys with the now revoked vehicle. As a result, the privacy retained by non-targeted vehicles is reduced i.e., the number of pseudonyms or keys for these non-targeted vehicles is reduced.

For example, suppose there is a VANET consisting of 5 vehicles ($n = 5$), each of which are assigned two keys ($d = 2$). Each key is shared by two vehicles ($g = 2$). Now suppose that one of the vehicles is revoked; that is, all of its keys are revoked. Assuming that a single vehicle does not share both keys with the revoked vehicle, then there are two vehicles that have only a single valid key remaining after the revoked vehicle is revoked. Before the revocation, for each key, each vehicle could hide among a group of two vehicles, itself and the other vehicle that shares the key with it. However, after the revocation, the size of this group is reduced to 1. Thus, the privacy of the vehicles who share keys with the revoked vehicle is reduced because after revocation no other vehicle shares the keys held by the revoked vehicle. If d were increased to 3, then the size of the group among which a vehicle is indistinguishable would also be larger, thus providing more privacy to vehicles. Increasing g to 3 would increase privacy by increasing the size of the group a vehicle hides among, but it also would increase the *brittleness*, that is, the loss of privacy experienced by innocent vehicles affected by the revocation.

Brittleness can be decreased by increasing d. If a large number of vehicles lose a key due to a revocation, holding more keys reduces the adverse affect to non-revoked vehicles.

4.5. Collapsibility

The property of *collapsibility* reflects how resilient a key assignment method's security properties are to key compromise. If a vehicle's hardware and correspondingly keys are compromised, a significant number of vehicles may be unable to use their keys to sign safety messages with the security properties intended by Axiom 3.1 and Corollary 3.3.

Collapsibility can be decreased by decreasing g. Intuitively, the smaller the number of vehicles that share a key, the smaller the number of vehicles that will be affected by a hardware and associated key compromise.

DOI: 10.1002/sec

Collapsibility is similar to brittleness in effect, however the two differ in mechanism.

4.6. Sybil Attack Resilience

Assigning identities to vehicles raises the concern that misbehaving vehicles may perform Sybil attacks. If each vehicle holds multiple keys, a malicious vehicle could use multiple keys simultaneously, giving the attacker an advantage. This advantage could take many different forms. The attacker could use multiple identities to artificially congest a roadway and make other drivers think they should take an alternative path to avoid the congestion. If revocation is based on voting for to determine malfunctioning or malicious vehicles, then performing a Sybil attack gives the attacker a greater number of votes. However, if keys are shared (i.e., $g \neq 1$), then Sybil attacks may also be performed even if $d = 1$, giving attackers the same advantage.

4.7. Key Collisions

A key collision occurs when two or more vehicles use the same key simultaneously and within a 2-hop radio range of each other. When a key collision occurs, a non-colliding vehicle that overhears both colliding vehicles' transmissions may think that a single vehicle is claiming multiple locations and is either malfunctioning or malicious. By increasing g, the probability of having key collisions increases. Key collisions will increase the number of vehicles wrongfully revoked, and present an opportunity for malicious vehicles to use a shared key to revoke another vehicle's keys. When $g = n$, key collisions may be expected and therefore ignored. Increasing d will decrease the probability of a key collision occurring, but it also increases the number of identities a Sybil attacker can immediately use.

5. Key Assignment

In this section, we consider what performance various key assignment methods can provide in terms of the properties given in Section 4. The CA is responsible for assigning keys to vehicles, therefore the CA will be the holder of privacy-sensitive vehicle data, such as keys and relationships between keys and vehicles. The CA keeps these keys so that the CA can revoke untrusted vehicles. How keys are assigned affects vehicle privacy and the usability of a VANET.

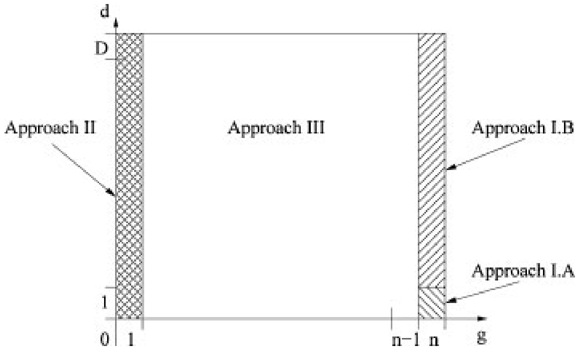

Fig. 1. The *g versus d* design space explored in relation to CA privacy.

Key Assignment Design Space: For our discussion in this section, a key assignment method describes how many keys each vehicle will own (d), as well as how many vehicles are assigned the same key (g). To fully analyze the g–d design space, we divide the space up into four different regions, which are illustrated in Figure 1. We will explore the properties of each of these regions below. As we will show, the choice of g and d will determine the amount of privacy available from the CA and from non-CA entities. We specifically investigate the impact of each key assignment region on the level of privacy a vehicle can maintain with respect to the CA and with respect to non-CAs entities. We will also show below that there is a trade-off between how much privacy a vehicle can maintain from the CA and how quickly a vehicle can be revoked from the network.

There are three main approaches for assigning a vehicle's signing keys which cover the g–d design space, as shown in Figure 1:

I.(A) There is only one key in the VANET. Each vehicle is supplied with this same key. ($g = n$, $d = 1$)
(B) There are many keys in the VANET, and each vehicle has every key. ($g = n$, $D \geq d > 1$)
II. Each vehicle is loaded with a set of keys. Keys are not shared between vehicles. ($g = 1$)
III. Each key is shared among a group of vehicles. ($n > g > 1$, $D \geq d \geq 1$)

We denote the maximum number of keys possible in a VANET as D, and the total number of vehicles as n. We assume a fixed D, though D may be very large. Without making this assumption, the problem of choosing a key assignment method becomes much harder due to trying to assign keys to vehicles when $g \neq 1$ and still being able to provide privacy to vehicles using newly added keys. We evaluate a static n as an assessment of VANET privacy properties at any given instance. The result of our analysis below will lead us to the conclusion that Approach II is the best of the three key assignment methods.

5.1. Key Assignment Approach I

We divide Approach I ($g = n$) into two sub-approaches: Approach I(A) ($d = 1$) and Approach I(B) ($d > 1$). In Approach I(A) all vehicles are provided with a copy of the same key. Under this approach, signing a message can be a simple, symmetric cryptographic operation since the key distribution problem is trivially solved, and non-repudiation is not possible. Since all vehicles have the same key, the CA cannot identify which vehicle signed an individual message based on the message's signing information. Similarly, no non-CA entity can tell two signatures apart based on key information alone. Thus, CA and non-CA privacy is complete with regard to key information. Unfortunately, a single hardware failure or successful attack would result in the complete compromise of such a solution, since no message could be trusted following the compromise of a single shared secret key. Thus, Approach I(A) is extremely collapsible.

One difference between Approach I(A) and Approach I(B) is that instead of having only a single key for the entire network, there are many keys, though each vehicle still shares all of the keys. Approach I(B) suffers from the same problem of collapsibility as Approach I(A) : a single hardware compromise results in all of the keys being compromised, which means no message could be trusted. Similarly, the CA still cannot identify the transmitter of a message for revocation purposes. Approach I(B) can also suffer from key collisions, where Approach I(A) did not have this problem. However, Approach I(B) possesses the same level of privacy from CAs and a non-CAs as Approach I(A) did.

Approach I(A) and I(B) share some common performance in terms of brittleness and Sybil attack resilience. Both Approach I(A) and I(B) are extremely non-brittle, in that the privacy of systems using Approach I(A) or I(B) is not reduced by the revocation of untrusted vehicles; however, the revocation of a single vehicle would result in the collapse of the VANET. Both Approach I(A) and I(B) are highly susceptible to Sybil attacks since correct vehicle behavior and a vehicle misusing the common key(s) is indistinguishable from safety message information alone.

5.2. Key Assignment Approach II

Approach II provides each key to only one vehicle ($g = 1$), i.e., no single key is held by more than one vehicle. This type of distribution solves Approach I's problem with key compromise and collapsibility. The primary advantage of the second approach, where each key is known only to one vehicle, is that revocation is efficient: once a single misbehavior is matched to a key, all the keys belonging to that vehicle can be revoked in a single revocation event, thus resulting in the exclusion of that vehicle from the VANET. In this approach, fast revocation can be the same as complete revocation, since each key is held by a single vehicle. A second advantage of this approach is that it can provide the property of non-repudiation if public keys are used for signing. The disadvantage of this approach is that each key uniquely identifies a vehicle, raising privacy concerns.

The CA will need to be able to correlate a vehicle's keys in order to enable fast revocation. For example, the CA may keep a list of all keys for each vehicle. These lists may be subpoenaed by a law enforcement agency that wishes to track certain VANET users. More generally, consider a design that attempts to protect users' privacy from the CA by intentionally keeping incomplete information about users keys at the CA. In such a case, the CA may not be able to revoke a vehicle in a single revocation event. However, as discussed in Section 4.3, if revocation is not fast (i.e., achieved in a small number of revocation events), the PKI has diminished value because it is not performing its function. Thus, even in such a case, there must be a process to keep revocations fast. In this approach, the same mechanism that is used to enable fast revocation can also be used for privacy compromise. In other words, if there is a mechanism that is useful for quickly revoking a vehicle's certificates, the same mechanism can be used to compromise a vehicle's privacy since the CA must be able to revoke all of the keys of that vehicle and therefore all the keys must be known to the CA. This linkage between fost revocation and privacy compromise holds independent of whether a single CA holds all of the information necessary for revocation or the information is dispersed among several CAs. Thus, no system based on having a single vehicle per key ($g = 1$) can provide both fast revocation and privacy from the CA. If a VANET designer is willing to sacrifice maintaining complete privacy from the CA, then this assignment method is tenable.

Since vehicles can be loaded with a large number of keys, d, and keys may never need to be reused, the

DOI: 10.1002/sec

non-CA privacy provided by Approach II is high. Only the CA can know the keys assigned to a vehicle, not non-CA entities if $d \neq 1$. Additionally, if $d \neq 1$ then Sybil attacks are possible. However, since all keys are unique to their vehicles, there are no key collisions.

Similarly, since vehicles do not share keys with other vehicles (i.e., $g = 1$), then Approach II is not brittle. When a vehicle's key is revoked, the privacies of other vehicles in the VANET are not reduced because no other vehicle has been assigned the revoked key.

5.3. Key Assignment Approach III

Approach III provides each key to several vehicles. This approach solves the problem of a CA definitively knowing which vehicle possesses each key, since each key is shared among a group of vehicles. However, as shown below, this improved CA-privacy comes at the expense of slower revocation. We mathematically explore this trade-off between revocation speed and CA-privacy below.

Approach III solves the issue of key compromise and collapsibility that Approach I had since not all vehicles are assigned the same keys in the VANET. Since $d \neq 1$ and $g \neq 1$, Approach III provides non-CA privacy because signatures are not attributable to individual vehicles ($g \neq 1$) and non-CA entities do not know which keys have been assigned to individual vehicles. Sybil attacks are again possible in Approach III because vehicles are assigned multiple keys. We will include a discussion of the brittleness of Approach III in our mathematical assessment of Approach III below.

Approach III attempts to avoid the trade-off of maintaining privacy from the CA and enabling fast revocation that Approach II had to make. Again, one goal of sharing keys is that, when law enforcement detects that a vehicle is using a certain key, they cannot affirmatively link that key back to a single vehicle. In particular, because other vehicles share the same key, law enforcement cannot affirmatively prove that a single vehicle was the violator. It may be desirable to set an even tougher goal for vehicle privacy in this scenario. Since evidence in a court of law builds a case, it may be desirable that the information obtained from a VANET be even less probative such that it cannot be efficiently used to build a case. A casual inspection of Figure 1 may lead one to think that the design space of Approach III is large. However, practical considerations greatly constrain the design of key management of Approach III. Here we list four primary constraints:

Constraint 1 — Assuming the CA can only revoke a single key for a single reported infraction (single reported key), revoking a vehicle requires that each of its keys be individually revoked, which upper bounds the number of keys per vehicle (d), if the efficacy of revocation is to be preserved. Generally this constraint is applicable, but we will discuss below the case when this constraint does not hold. (At the time a key is revoked, it may be possible to infer which other keys a vehicle holds besides the key being revoked. We will discuss this possibility in greater depth below.)

Constraint 2 — If a constant fraction f of all vehicles have had all of their keys revoked as results from Axiom 3.1, then excessive sharing will result in all keys within the system being revoked. (Intuitively, if each key is shared by g nodes, and $g \approx 1/f$, then all of a vehicle's keys will be revoked with high probability. We will investigate this outcome in mathematical detail below.)

Constraint 3 — A privacy compromiser might observe *transitions* between keys. For example, a law enforcement officer may observe a single vehicle speeding, and during this observation, the speeding vehicle transitions from one key to another. With each additional key observed by the law enforcement officer, the pool of suspects shrinks. If the number of keys that the law enforcement officer observes in this way (ρ) is sufficiently large, but the degree to which each key is shared (g) is sufficiently small, then the probability that more than one vehicle has all such keys approaches zero, contravening the objective of providing anonymity from the CA.

Constraint 4 — Key collisions will cause additional revocations in a VANET. A designer must mitigate these additional unnecessary revocations by either increasing d or decreasing g.

Considering Constraint 1, it may be possible for the CA to infer which additional keys are held by a vehicle given a single key from the vehicle. This situation might arise and be of significance during revocation. Using the single key reported for revocation, a CA may be able to infer additional keys held by the offending vehicle, and the CA may use this knowledge to revoke more than one key at a time. However, this option is not considered by current approaches. We will show below that for other reasons, mathematically described, Approach III does not allow for both privacy and fast revocation. If the CA can infer more than one key given a single key, Approach III simply provides even less privacy. Additionally, any inference method useful to the CA

DOI: 10.1002/sec

Table I. Privacy design parameters.

g	Number of vehicles sharing each key (cars/key)
d	Number of keys held by each vehicle (keys/car)
n	Number of vehicles in the VANET
ε	Probability of false positive
f	Fraction of vehicles with all certificates revoked
ρ	Number of keys from one vehicle required by a single observer to break privacy
w	Wrongful revocation rate
σ	Vehicle encounter rate
C	Key collision rate

for revocation purposes will be useful to other agencies for their various purposes (e.g., law enforcement using inference to more completely track and ticket speeding vehicles). Consequently, we will not consider the use of inference by the CA in our discussion below, and we will consider Constraint 1 to apply.

Designers of a VANET must be careful in how they choose the parameters mentioned above. Failing to do so can result in consequences that may not be immediately apparent. Table I shows the notation we will use in the following discussion. To illustrate these design decisions and the constraints discussed above, we consider two bounding cases under two opposite assumptions: complete independence in terms of key assignment, that is, keys are assigned completely at random, and complete dependence, that is, any two cars that share a single key in common will also share all of their keys in common. In the following discussion, we will assume g and d to be constants, that is, keys are shared among groups of equal size and each vehicle is loaded with the same number of keys. A vehicle that has had some keys revoked will have less than d keys remaining that it can still use. We will justify these assumptions after presenting our mathematical analysis of privacy for Approach III.

5.3.1. Independent distribution

Assuming keys are distributed independently, in Approach III, each key is held by multiple vehicles. Therefore, the CA does not know which vehicle among the group that holds a certain key is the vehicle being reported for revocation. In other words, the CA cannot know from a single revocation report more than the single associated key, and only one key can be revoked at a time. Thus, the number of keys per vehicle d must be limited so that revocation is still an effective mechanism.

Design: Let us assume that all keys are distributed independently in a VANET that has n vehicles. Let us also only concern ourselves for now with revocation effects due to the fraction of vehicles, f, that have had all of their keys revoked. The probability that an arbitrary key is not revoked is equal to the probability that no vehicle in the group of revoked vehicles holds that key. Quantitatively, the probability that an arbitrary key is not revoked is $(1-f)^g$. We call w the fraction of vehicles wrongfully revoked [‖]. We define an innocent vehicle as a vehicle outside the group of justifiably revoked $f \cdot n$ vehicles. Under the independent key distribution assumption, each innocent vehicle may or may not share a key with one of the $f \cdot n$ untrusted vehicles. Consequently, the probability that an innocent vehicle has all of its keys revoked is,

$$w = \left(1 - (1-f)^g\right)^d \tag{1}$$

Equation (1) states that for each of a vehicle's d keys, at least one of the $g-1$ other holders of those keys are in the completely revoked, untrusted vehicle group.

Above, we mentioned that one metric of privacy is the likelihood of being correctly or incorrectly identified after an observation. If a vehicle blends into its surroundings and enjoys high privacy, then the likelihood of it being misidentified is high. Consider, for example, identification based on hair-color (high likelihood of misidentification, since many people share the same hair color), DNA matching (low likelihood of misidentification) and blood type[¶]. To aid in the intuition, imagine cases where a court of law tries to identify a defendant based on some identifier, e.g., hair color, DNA matching, blood type. Essentially, the higher the likelihood for misidentification, the less likely the court will treat the evidence as definitive. Thus, a person who is identified by some highly shared characteristic, e.g., brown hair, retains a higher level of privacy than the person who is identified with some unique characteristic, e.g., certain combinations of DNA markers. This method of measuring privacy is appropriate for discussing the privacy VANET users maintain from the CA and those (such as law enforcement officers) that can subpoena the CA. Over the past few years,

[‖] Wrongful revocation may be highly unacceptable to users. The average consumer may not accept or understand when a vehicle service provider (e.g., repair shop) explains that their vehicle or VANET safety enhancements are not functioning because some other vehicles are misbehaving and the network is designed knowing that this could happen.

[¶] Some current resources list blood type as having at least 0.6% likelihood of misidentification. The least common blood type in the United States is AB−, which is present in 0.6% of the population.

DOI: 10.1002/sec

protecting privacy has become a more pervasive issue in society. Thus, VANETs may be unacceptable to users if governments or police can employ VANET data to issue driving violations, such as speeding.

Continuing this line of thought, a vehicle that broadcasts identifiers that are widely shared maintains more privacy from the CA than if it broadcasts identifiers shared by a small group of vehicles. To illustrate, consider a scenario where a law-enforcement-controlled listening-station receives a VANET packet from a speeding vehicle. If that packet contains a non-reputable signature created using key held by only one vehicle ($g = 1$), then that vehicle has essentially no privacy. Via a subpoena to the CA, law enforcement could use the unique identifier to determine the identity of the vehicle.

If, on the other hand, the identifying key is held by exactly three vehicles ($g = 3$), then the vehicle has more privacy than before. Then consider that law enforcement finds a vehicle shown to hold the offending key and arrest the vehicle owner. In this case, if law enforcement has no more evidence, they must admit that there is only a 1/3 chance that they have the correct driver, i.e., the misidentification rate (given that the defendant is shown to have the key) is still 2/3. Courts should not convict based solely on such evidence since it has such a high misidentification rate. Of course, if the key is held by even more vehicles, the misidentification rate, and thus privacy, increases.

If instead the law-enforcement-controlled listening-station is able to observe the same vehicle using two different identifiers, e.g., the speeding vehicle switches from one signing key to a different signing key while being observed by the listening-station, then the situation changes: assuming as we do throughout this subsection that the keys were assigned independently, the pool of possible vehicles (i.e., vehicles which hold both observed keys) shrinks significantly. We now develop a general mathematical analysis of privacy when an arbitrary number of certificates are known to come from a single vehicle.

Suppose a vehicle behaving maliciously is observed by a law enforcement officer, and suppose that the officer observes the vehicle using ρ keys. The probability that a second arbitrary vehicle other than the observed vehicle shares ρ keys[#] with the observed vehicle is,

$$\varepsilon = \left(\frac{g-1}{n-1}\right)^{\rho} \tag{2}$$

ε is the probability of *mistaken identity* or of a *false positive*. We will refer to ε as the false positive rate below. Again, the false positive rate is the probability that one vehicle is mistaken for another vehicle based on key information. Here, the two vehicles can be mistaken for each other because they have at least ρ keys in common. Consider the following scenario. One vehicle is observed using ρ specific keys. A second vehicle is compelled to admit that it also possesses the same ρ keys. In this scenario, the second vehicle has a probability of being mistakenly identified as the first with probability ε. One analog of this false positive rate is the false positive rate for DNA matching between two random people. Under the assumption of independent key distribution, Equations (1) and (2) result in a trade-off between the false positive rate and the wrongful revocation rate, which we will illustrate below.

Initially, one might think that the false positive rate should be minimized. However, a high enough false positive rate (e.g., from using keys to identify rule-breakers) implies that such evidence would not be definitive. In other words, those who would like to discourage law enforcement from using keys for identifying suspects would want ε to be so large that such use would be widely discredited. This argument is similar to the fact that law enforcement cannot convict a driver of a 'green car' simply because they observed a green car breaking the law; the likelihood of mistaken identification is too high. Essentially, the larger the ε, the less privacy is sacrificed by VANET users to the CA.

A false positive may occur if the innocent vehicle and the observed vehicle share at least ρ keys. If we specify a lower bound, ε'_{ρ}, which provides an acceptable amount of privacy for a given ρ, then, solving Equation (2) for g, we get,

$$g(\varepsilon'_{\rho}) \geq 1 + (n-1)\varepsilon'^{\frac{1}{\rho}}_{\rho} \tag{3}$$

Thus, $g()$ is $\Theta(n)$, that is, given a number of observed keys ρ and a privacy bound, ε'_{ρ}, as the number of cars n increases, the number of keys held by each car, $g()$ would also need to increase linearly with n. Additionally, since $g()$ is $\Theta(n)$, $(1-f)^g$ goes to 0 and $w = (1-(1-f)^{g}(\varepsilon'_{\rho}))^d$ goes to 1 as n goes to ∞. w going to 1 implies that once all the $f.n$ vehicles are completely revoked, the remaining $(1-f).n$ innocent vehicles are also revoked. Thus, Approach III with independent key distribution does not scale

[#] In a privacy compromising situation ρ would need to be the number of keys observed by the privacy attacker.

DOI: 10.1002/sec

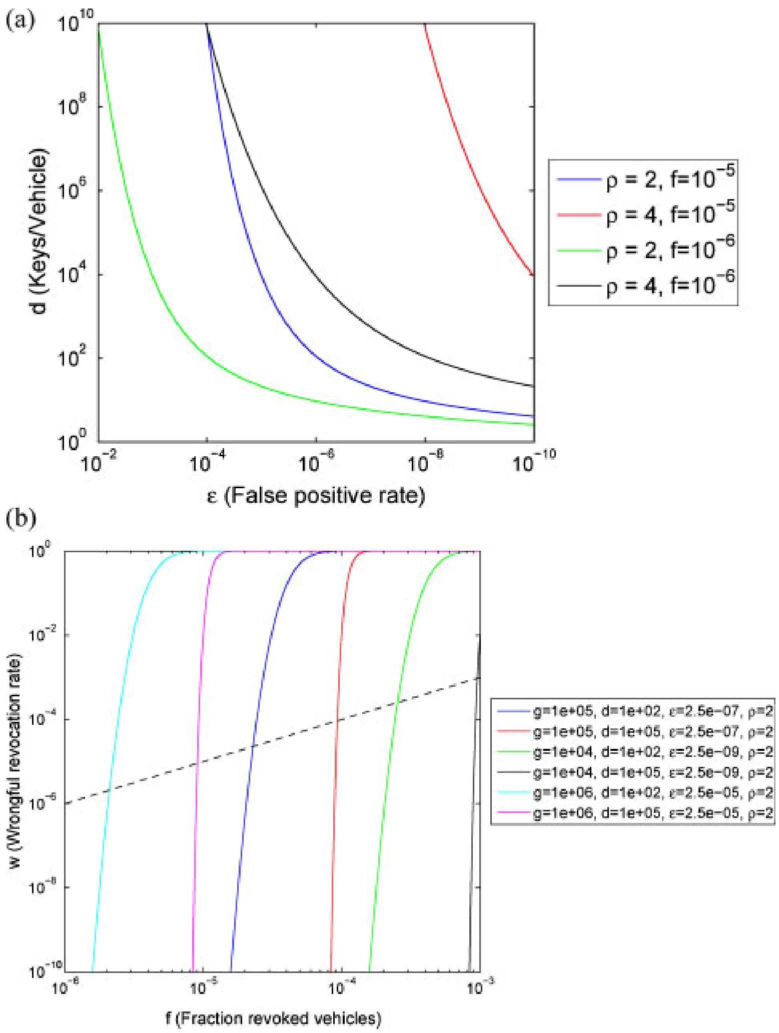

Fig. 2. Privacy performance of independent key distribution. (a) Independent distribution—Number of keys per car *versus* false positive rate; (b) Independent distribution—Wrongful revocation rate *versus* fraction revoked vehicles.

with increasing network size. When we fix the false positive rate (ε), the maximum fraction of cars wrongfully revoked (w), number of cars (n), and number of connected keys (ρ), we find a bound on d that satisfies Constraints 2 and 3 by combining Equations (1) and (2)

$$d \geq \frac{\log(w)}{\log\left(1-(1-f)^{1+(n-1){\varepsilon'_\rho}^{1/\rho}}\right)} \qquad (4)$$

We now explore realistic designs that keep ε sufficiently large and the impact of that strategy on revocation effectiveness, under the framework of Approach III and independent key distribution.

Examples: Consider the situation in the United States where $n \approx 200$ million. Figure 2 illustrates some of the design space for the United States. Figure 2(a) shows the number of keys, d, that must be distributed to each vehicle in the VANET for a given false positive rate, ε (or level of privacy desired), and various choices of ρ and f. This graph shows that to increase ε and achieve more privacy, the number of keys assigned to each vehicle d must increase dramatically or ρ must be decreased for a fixed fraction of completely revoked vehicles f. Figure 2(b) shows the wrongful revocation rate as a function of the fraction of completely revoked vehicles f for various choices of g, d, ε and ρ. The dashed line shows the case of $w = f$. This figure compares the fractional population size of two groups, vehicles that have been completely and intentionally revoked

DOI: 10.1002/sec

(*f*) and vehicles that are wrongfully revoked (w). The graph shows that w increases rapidly with increasing *f*, and increases rapidly with increasing levels of privacy, ε. Since w increases rapidly with increasing ε, this shows the trade-off between wrongful revocation and privacy. To achieve a plausible level of privacy, that is, a reasonably large number of vehicles to hide among, e.g., $\varepsilon = \frac{1}{4000}$ in the graph, and $w < f$, a large number of vehicles must share each key, i.e., g must be large (greater than 1 million in the figure) and the fraction of revoked vehicles must be small (less than 1 in 100 000 in the figure). To illustrate these results further, when $\varepsilon = 10^{-7}, w = 10^{-4}, f = 10^{-5}$, and $\rho = 2$, then $d \approx 110$. When ρ increases, d grows quickly; at $\rho = 4$, $d > 2.7 \cdot 10^{16}$. To reduce d to $\approx$320, f cannot exceed 10^{-6}, but this results in 100 times more innocent revoked vehicles than rightfully revoked vehicles.

5.3.2. Dependent distribution

Design: Now, let us assume keys are no longer independently distributed but are distributed in a completely dependent manner, that is, if any two vehicles share a single key, then they share each of their keys. At a high level, this type of distribution will suffer increasingly from wrongful revocation as *f* and *g* increase. The larger the groups of vehicles are, that is, *g*, the more wrongfully revoked vehicles there will be.

Consider again Constraint 1 from above in this new manner of distribution. When a vehicle is reported for revocation purposes, the CA can do one of two things: the CA can revoke a number δ of a vehicle's keys or the CA can revoke all of a vehicle's keys. A dependent distribution is essentially Approach II where the CA could keep a list of unique keys held by each vehicle, except now the CA can keep a list of unique keys held by a single group of vehicles. Thus, revocation can be fast for this manner of distribution for the same reasons as revocation could be fast for Approach II above. This results in d not needing to be bounded for fast revocation purposes.

Since keys are no longer independently distributed, the false positive rate is, $\varepsilon = \frac{g-1}{n-1}$, and after fixing ε, $g \geq 1 + (n-1)\varepsilon$. With a dependent distribution, ε is no longer a function of ρ. In this design, the false positive rate cannot be decreased by increasing the number of keys a vehicle holds. The probability that a vehicle is wrongfully revoked has become $w = 1 - (1-f)^g$. Again, the wrongful revocation rate cannot be reduced by increasing the number of keys held by a vehicle.

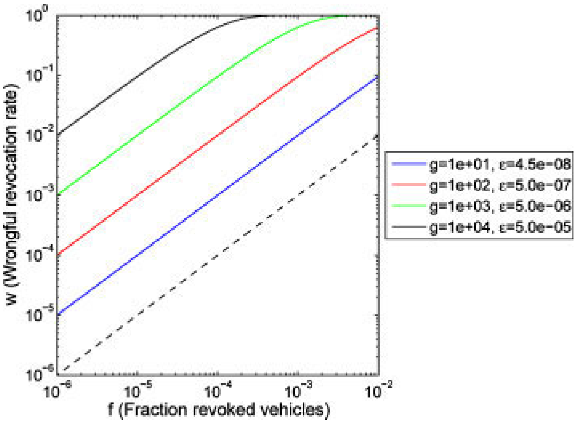

Fig. 3. Dependent distribution—wrongful revocation rate *versus* fraction revoked vehicles.

Examples: Using the same assumption for the vehicle population in the United States, Figure 3 shows the resulting wrongful revocation rate as a function of the fraction of completely revoked vehicles, and for various choices of g and ε. Again, the dashed line shows the case of $w = f$. In the dependent case, the resulting privacy ε is solely dependent on g for a fixed n. This figure shows that for greater privacy (e.g., $\varepsilon = \frac{1}{20\,000}$), the fraction of vehicles wrongfully revoked is many orders of magnitude higher than the fraction of revoked vehicles. Thus, keys distributed in a completely dependent manner is untenable.

5.3.3. Key collisions

Under either distribution, independent or dependent, because more than one vehicle can be using the same key at the same time in the same location, key collisions will occur. When an observer hears two other 'vehicles' claiming two separate locations, each using the same key, the observer has no way to differentiate between the case of two innocent vehicles using the same key by coincidence, or one untrustworthy vehicle pretending to be multiple vehicles. In such cases, an observer may report the common key as untrusted, with the goal of triggering its revocation. There is a rate at which keys will be removed from the VANET due to unnecessary revocations from key collisions. These revocations will come from vehicles that overhear other well-behaved vehicles that happen to be using the same key within the radio range of the overhearing vehicle (the vehicles whose keys are colliding need not be within radio range of each other for this to occur). We define σ to be the rate at which a vehicle encounters other vehicles (e.g., 100 vehicles per day).

Intuitively, the expected number of revocations due to key collisions is proportional to the rate at which a vehicle encounters other vehicles, σ, and the number of vehicles that share each key, g. The expected number of revocations due to key collisions is inversely proportional to the number of keys held by each vehicle, d, and the total number of vehicles in the VANET, n. That is, the more vehicles another vehicle encounters, or the more vehicles that share keys with that vehicle, the higher the probability key sharing vehicles will encounter each other and a key collision revocation will occur. Conversely, by increasing d it becomes less likely that vehicles that share keys will use their shared keys concurrently.

Let us consider a single vehicle. This vehicle encounters σ other vehicles during some time period. During this period, suppose that no vehicle changes the key that it is using. Let C be a random variable that denotes the number of vehicles encountered by the first vehicle and that have been assigned the key that the first vehicle is using during this time period.

We can calculate the probability that a vehicle encounters any number of other vehicles, $i \leq \sigma$, that hold the key the first vehicle is using during a period as

$$P(C = i) = \frac{\binom{\min(g-1,\sigma)}{i}\binom{n-g}{\sigma-i}}{\binom{n-1}{\sigma}} \tag{5}$$

We make use of the property that $\binom{a}{b} = 0, \forall b > a$ to handle the case where $\sigma > g - 1$. However, not all vehicles that share the key the first vehicle is using are using that key during this period. K is a random variable describing how many vehicles are using the same key as the observed vehicle when the observed vehicle encounters those other vehicles. Thus, the probability that $C = i$ and that $K = j$ of the encountered sharing cars are using that key is

$$P(C = i \cap K = j) = \frac{\binom{\min(g-1,\sigma)}{i}\binom{n-g}{\sigma-i}}{\binom{n-1}{\sigma}} \left(\frac{1}{d}\right)^{j} \times \left(1 - \frac{1}{d}\right)^{i-j} \tag{6}$$

In order for a designer to be able to choose system parameters such as d and g, a useful quantity to know is the expected number of revocations due to key collisions over some period. We can calculate this using the following

$$E[C \cap K] = \sum_{i=1}^{\sigma}\sum_{j=1}^{i} \frac{\binom{\min(g-1,\sigma)}{i}\binom{n-g}{\sigma-i}}{\binom{n-1}{\sigma}} \left(\frac{1}{d}\right)^{j} \times \left(1 - \frac{1}{d}\right)^{i-j} \tag{7}$$

First, notice that the terms introduced to capture which key a vehicle is using during a given period, terms that include d and j, are binomial and not dependent on i. Thus, the expected value contributed by this part is easily calculated as $\frac{i}{d}$. Again, making use of a well-known binomial identity, we achieve

$$\kappa = E[C \cap K] = \frac{\sigma(g-1)}{d(n-1)} \tag{8}$$

This expected value, κ is the expected number of key-collision-induced revocations over a given period of time per vehicle.

If we bound κ from above as κ', holding d constant, then we bound g from above

$$g \leq 1 + \frac{\kappa' d}{\sigma}(n-1) \tag{9}$$

Conversely, we can choose a constant g and achieve a lower bound on d. Our analysis, however, is for the initial static situation in a VANET. As the VANET operates, key collisions and revocations will occur, thus keys will be removed from the VANET. This key evaporation leads to an increased probability that two vehicles will be using the same key simultaneously, and will lead to an accelerating rate of wrongful revocations due to key collisions.

Ignoring the accelerating key evaporation, imagine that vehicles are constrained to a fixed area of operation, then the steady-state behavior of the VANET is a complete lack of privacy. Because of the mixing of vehicles in their overlapping areas, the certificates that are shared by the overlapping vehicles will all be revoked due to key collisions after a sufficiently long period of time. Thus, for VANET traffic that is recorded at a single location, there is no privacy provided to vehicles. In reality, vehicles are not constrained to a fixed area. However, it is likely that most of the time, a vehicle operates in the same area, that is, trips made outside some area are relatively infrequent as compared to 'normal' operation.

The result of key collisions is that a designer must increase d to combat the extra unnecessary revocations, However, increasing d makes revocation in response to malfunction or malice slower when keys are not assigned in a completely independent manner, as was

DOI: 10.1002/sec

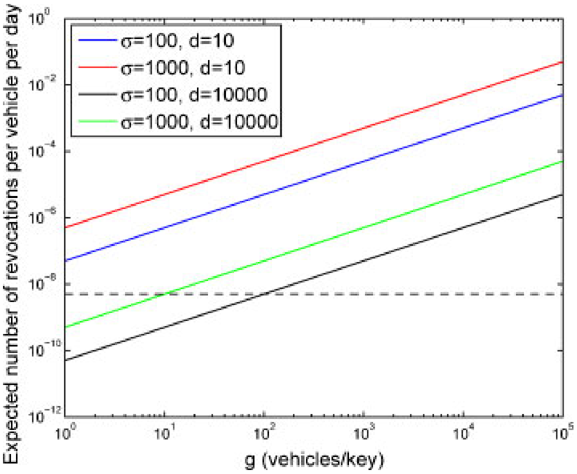

Fig. 4. Expected number of revocations due to key collisions per vehicle per day. σ (vehicles/day) d (keys/vehicle).

the case when we were trying to preserve privacy and reduce wrongful revocation.

Examples: Let us assume the same vehicle population as above ($n = 200$ million), and that vehicles encounter each other at a rate of σ (e.g., 100 vehicles per day). Figure 4 shows the expected number of revocations per vehicle per day with σ (the number of vehicles encountered per day) measured in vehicles/day and d being the number of keys assigned to each vehicle. The dashed line shows $\kappa = \frac{1}{n}$. Intuitively, one might expect the network-wide number of revocations per day to be on the order of $n\kappa$. Taking $g = 100$, $d = 10\,000$, and $\sigma = 10$, $n\kappa \approx 1$. Increasing σ to 1000 gives $n\kappa \approx 10$.

5.3.4. Discussion

In reality, designers of a VANET may choose a system with key distribution lying somewhere between complete independence and complete dependence. The above discussion illustrates that privacy from the CA may be unattainable while maintaining reasonable network performance (e.g., more malicious vehicles are revoked than legitimate vehicles, i.e., $w < f$) using Approach III.

Let us reconsider our assumptions that g and d are constants, and allow g to be a function of the individual key under the mathematical framework developed above.[**] Suppose g is chosen such that $g_{lo} \leq g \leq g_{hi}$. For the case of independent key distribution, the keys that are shared by g_{lo} vehicles would be considered *low-privacy keys* since there is a smaller population of vehicles for an individual vehicle holding one of these keys to hide among. The low-privacy keys provide a smaller ε to vehicles than choosing the constant g would have provided. Consider also that keys shared among g_{hi} vehicles result in a larger w, that is, more vehicles are revoked wrongfully. Thus, these keys have *higher risk* for revocation than keys distributed with constant g. The same outcome is obtained if keys are distributed in a completely independent or in a completely dependent manner.

Consider allowing d to vary from vehicle to vehicle. Let $d_{lo} \leq d \leq d_{hi}$. If keys are distributed completely independently, vehicles with d_{lo} keys will have a smaller false positive rate, and thus are *low-privacy vehicles*. Similarly, vehicles with d_{hi} keys have a higher false positive rate and are *high-privacy vehicles*. This type of distribution implies different classes of service in terms of privacy for different vehicles. When keys are distributed in a completely dependent manner, having a non-constant d does not affect either the false positive rate or the wrongful revocation rate.

Thus, allowing g and d to vary results in both poorer privacy and more wrongful revocations. Considering the above arguments for how to assign keys to vehicles, we can eliminate designs that are infeasible, given our goals of maintaining privacy, enabling fast revocation, and building a robust system. Key assignment Approach I is infeasible because a single CA or vehicle compromise results in complete key compromise for the VANET. Approach III is infeasible because it cannot provide both CA privacy (large ε) and fast revocation (small d). As a result, we are left with Approach II, where limited CA privacy is possible, but is not able to be retained for vehicles that have keys revoked. We believe Approach II to be superior for key assignment because the safety properties of the VANET are preserved (e.g., vehicles are removed quickly for infractions through revocation), it provides a highly non-brittle VANET, and it results in no wrongful revocation, unlike Approach III.

6. Conclusion

If a VANET designer wants to preserve the principles laid out in Section 3, there is only one key assignment method that is viable, that being Approach II. Choosing Approach II may come with the sacrifice

[**] The mathematics of a scheme where g and d vary are not actually the same. In fact, having non-constant g and d may make the mathematics of an analysis, such as we have carried out above, intractable.

of CA privacy, but it is the only approach that results in a viable VANET and provides fast revocation. We have shown that due to collapsibility, Approach I is not viable. Similarly, we have shown that due to the inherent trade-off between fast revocation and the false positive rate (i.e., privacy), plus the brittleness and potential evaporation of legitimate keys due to collisions, Approach III is not viable.

Approach III is not totally dismissible, though from our analysis, Approach II provides a superior VANET design. We have investigated two methods for distributing keys under Approach III, completely independent and completely dependent distributions. There may be other methods for distributing keys under Approach III. For example, a geographic-based distribution, where certain keys are assigned to vehicles only within a certain geographic area, may provide superior performance to the distributions we have discussed. Geographic distribution may suffer from increased key collisions, due to higher densities of vehicles having been assigned a specific key, and may be restrictive in the sense that vehicles privacy can only be maintained while they are within the geographic area in which the vehicle is designed to operate. This latter problem may be important to, for example, college students or military families who move large distances and take their vehicles with them. Whether a type of distribution exists that makes Approach III viable is an open problem for research.

References

1. Naito T, Tsukada T, Yamada K, Kozuka K, Yamamoto S. Robust license-plate recognition method for passing vehicles under outside environment. *Vehicular Technology, IEEE Transactions on* 2000; **49**: 2309–2319.
2. Chang S-L, Chen L-S, Chung Y-C, Chen S-W. Automatic license plate recognition. *Intelligent Transportation Systems, IEEE Transactions on* 2004; **5**: 42–53.
3. Raya M, Hubaux J. The security of vehicular ad hoc networks. In *Workshop on Security of ad hoc and Sensor Networks*, 2005.
4. Dötzer F. Privacy issues in vehicular ad hoc networks. In *Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks*, ACM Press, 2005.
5. Gerdes RM, Daniels TE, Mina M, Russell SF. Device identification via analog signal fingerprinting: a matched filter approach. In *Proceedings of the 13th Annual Network and Distributed System Security Symposium, NDSS'06*, February 2006.
6. Desmond LCC, Yuan CC, Pheng TC, Lee RS. Identifying unique devices through wireless fingerprinting. In *WiSec '08: Proceedings of the First ACM Conference on Wireless Network Security*, ACM: New York, NY, USA, 2008; pp. 46–55.
7. Brik V, Banerjee S, Gruteser M, Oh S. Wireless device identification with radiometric signatures. In *MobiCom '08: Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, ACM: New York, NY, USA, 2008; pp. 116–127.
8. Beresford AR, Stajano F. Mix zones: User privacy in location-aware services. In *PERCOMW '04: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops*, IEEE Computer Society: Washington, DC, USA, 2004, p. 127.
9. Feudiger J, Raya M, Félegyházi M, Papadimitratos P, Hubaux J-P. Mix-zones for location privacy in vehicular networks. In *WiN-ITS 2007*, August 2007.
10. Buttyn L, Holczer T, Vajda I. On the effectiveness of changing pseudonyms to provide location privacy in vanets. In *In Proceedings of ESAS*, 2007.
11. Sampigethaya K, Huang L, Li M, Poovendran R, Matsuura K, Sezaki K. CARAVAN: providing location privacy for VANET. *Proceedings of the 3rd Annual Conference on Embedded Security in Cars(escar 2005)*, November 2005.
12. Sampigethaya K, Li M, Huang L, Poovendran R. Amoeba: robust location privacy scheme for vanet. *Selected Areas in Communications, IEEE Journal on* 2007; **25**: 1569–1589.
13. Jiang T, Wang HJ, Hu Y-C. Preserving location privacy in wireless lans. In *MobiSys '07: Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*, ACM: New York, NY, USA, 2007; pp. 246–257.
14. Parno B, Perrig A. Challenges in securing vehicular networks. In *Workshop on Hot Topics in Networks (HotNets-IV)*, 2005.
15. Raya M, Hubaux J-P. Securing vehicular ad hoc networks. *Journal of Computer Security* 2007; **15**(1): 39–68.
16. Resendes R. The new 'Grand Challenge'—deploying vehicle communications, Keynote Address—*The Fifth ACM International Workshop on VehiculAr InterNETworking (VANET 2008)*. September 2008.
17. Lin X, Sun X, Ho P-H, Shen X. Gsis: a secure and privacy-preserving protocol for vehicular communications. *Vehicular Technology, IEEE Transactions on* 2007; **56**: 3442–3456.
18. Xi Y, Sha K, Shi W, Schwiebert L, Zhang T. Enforcing privacy using symmetric random key-set in vehicular networks. *Autonomous Decentralized Systems, International Symposium on* 2007; **0**: 344–351.
19. Hu Y-C, Laberteaux KP. Strong VANET security on a budget. *Proceedings of the 4th Annual Conference on Embedded Security in Cars(escar 2006)*, November 2006.
20. Robinson CL, Caminiti L, Caveney D, Laberteaux K. Efficient coordination and transmission of data for cooperative vehicular safety applications. In *VANET '06: Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks*, ACM: New York, NY, USA, 2006; pp. 10–19.
21. BBC News. Probe into data left in car park. http://news.bbc.co.uk/2/hi/uk_news/7704611.stm, November 2008.
22. BBC News. Firm 'broke rules' over data loss. http://news.bbc.co.uk/2/hi/uk_news/politics/7575989.stm, August 2008.
23. TSA Public Affairs. TSA Suspends Verified Identity Pass, Inc. Clear Registered Traveler Enrollment. http://www.tsa.gov/press/releases/2008/0804.shtm, August 2008.
24. Golle P, Greene D, Staddon J. Detecting and correcting malicious data in vanets. In *VANET '04: Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks*, ACM: New York, NY, USA, 2004; pp. 29–37.
25. Laberteaux KP, Haas JJ, Hu Y-C. Security certificate revocation list distribution for vanet. In *VANET '08: Proceedings of the fifth ACM International Workshop on VehiculAr Inter- NETworking*, ACM: New York, NY, USA, 2008; pp. 88–89.
26. Papadimitratos PP, Mezzour G, Hubaux J-P. Certificate revocation list distribution in vehicular communication systems. In *VANET '08: Proceedings of the Fifth ACM International Workshop on VehiculAr Inter-NETworking*, ACM: New York, NY, USA, 2008; pp. 86–87.

DOI: 10.1002/sec