Decentralized Enforcement of k-Anonymity for Location Privacy Using Secret Sharing

David Förster
Robert Bosch GmbH
david.foerster@de.bosch.com

Hans Löhr
Robert Bosch GmbH
hans.loehr@de.bosch.com

Frank Kargl
Ulm University, Germany &
University of Twente, NL
frank.kargl@uni-ulm.de

***Abstract*—Protection of location privacy by reducing the accuracy of location data, until a desired level of privacy (e.g., measured as *k-anonymity*) is reached, is a well-known concept that is typically implemented using a privacy proxy. To eliminate the risks associated with a central, trusted party, we propose a generic method to enforce k-anonymity of location data in a decentralized way, using a distributed secret sharing algorithm and the concept of *location and time specific keys*. We describe our method in the context of a system for privacy-friendly traffic flow analysis, in which participants report origin, destination, start and end time of their trips. In order to protect their privacy the accuracy of time and location information is reduced, until it applies to at least k distinct trips. No trusted, central party is required to determine how much the accuracy of each *trip report* must be reduced. The participants establish location and time specific keys via vehicle-to-vehicle (V2V) communication at the beginning and end of their trips. They use these keys to encrypt trip reports with several levels of accuracy, and uploaded them to a central, untrusted database. The keys are published using a secret sharing algorithm that allows their reconstruction, once at least k shares of the same key have been uploaded. Consequently, trip reports become available automatically, after k vehicles have made "the same trip" (same origin, destination, start and end time) with respect to a certain accuracy level.**

I. Introduction

Traffic authorities require information about traffic flows for operational control as well as strategic planning of new infrastructure. Only a few years ago it was hardly feasible to measure traffic flows directly. Instead, the origin-destination (OD) matrices representing the traffic flow were often estimated based on traffic counts [1]. The advent of cellular communication allowed for large-scale collection of traffic flow data. Even without drivers' involvement traffic flows can be derived from the data generated by the regular operation of mobile phone networks [2], [3]. More accurate results can be achieved by explicit collection of *floating car data (FCD)*, containing GPS position and sometimes also speed and other information [4], [5]. Most GPS navigation systems and smartphone navigation apps collect floating car data from their users, in order to incorporate traffic conditions in their routing decisions [6].

Measurement of *local traffic densities* can be done in a fully anonymous manner, by having vehicles submit FCD records in a predefined time interval. If no identifiers are included in the submitted data and different records from the same vehicle cannot be linked, submission of the data does not affect drivers' privacy, because no information about their trips' origin or destination can be inferred. For large-scale traffic analysis and planning, however, knowledge about *traffic flows* (as represented by OD matrices) is required. In contrast to FCD records this information is much more privacy sensitive. It was shown that, even with personal identifiers removed, detailed location traces (or origin/destination pairs) can be used to identify drivers' home location [7] or even their identity [8], [9]. Therefore, additional privacy protection is required when collecting information about trips' origin and destination.

A common approach to protecting location privacy is to deliberately reduce the spatial or temporal accuracy of information until a certain privacy level can be guaranteed [10], e.g., expressed as *k-anonymity* [11]. A user is k-anonymous if he cannot be distinguished from $k-1$ other users based on the information he reveals. This is well-suited for the use case of traffic flow analysis: Information about routes that are taken by many drivers are most important. Those drivers can reveal origin and destination of their trip with a rather high accuracy and still remain k-anonymous. Routes that are only used by few drivers are less important, therefore it is acceptable that the accuracy of those reports must be reduced more in order to achieve the same level of privacy protection.

k-anonymity can easily be enforced when all records are stored in central, trusted database. However, a database containing large quantities of highly accurate trip reports would be an attractive target for hackers. Recent security breaches such as the Sony hack [12] and revelations about state-run surveillance activities [13] have given rise to public concerns about privacy. It may be more attractive for drivers to participate in a system where privacy protection does not depend on the protection of a central database (and its operator's honest behavior), but is verifiably enforced by the participants themselves.

An essential building block of the system we propose is vehicle-to-vehicle (V2V) radio communication. Vehicle-to-x (V2X) communication, comprising vehicle-to-vehicle and vehicle-to-infrastructure (V2I) communication, has been developed and standardized during the last decade. Car manufacturers have announced the first V2X equipped models for model year 2017 [14]. Based on IEEE 802.11p radio communication [15] vehicles can exchange messages in an ad-hoc manner within a range from one hundred to a few hundred meters [16]. The technology is expected to enable a wide variety of safety, comfort, and entertainment functions [17]. Due to the expected contribution to road safety, the U.S. has initiated the process for making V2X-based safety functions a requirement for newly sold cars [18], which is promising with regard to adoption and market penetration.

978-1-4673-9411-6/15/$31.00 ©2015 IEEE

Our contribution

We describe a generic mechanism for enforcing k-anonymity for location data that does not require a central, trusted party and is therefore robust against malicious backend providers and compromised backend systems. As an example for its application we created a system for privacy-preserving traffic flow analysis, in which participants make available origin, destination, start and end time of their trips. Parties that query the system learn the information with highest accuracy possible such that it still applies to at least k trips.

The remainder of this paper is structured as follows: We survey related work in Section II and present our system model and our requirements in Sections III and IV. We describe our system and its building blocks in Section V and evaluate its security properties and performance in Section VI before we conclude with Section VII.

II. Related work

Beresford and Stajano define *location privacy* as "the ability to prevent other parties from learning one's current or past location" [19]. Several publications highlight the requirement for advanced privacy protection beyond simple anonymization: Hoh et al. examine privacy in traffic monitoring systems and were able to identify drivers' home locations from their GPS traces with a success rate of about 85% [7]. Krumm conducted a similar experiment and was able to infer the identity of 5% of the participants using a public internet search engine to look up people living near the identified home locations [8]. Using data from the U.S. Census Bureau, Golle and Partridge demonstrated that the majority of the U.S. working population can be uniquely identified by the combination of their home and work location [9]. Jeske examines the data submitted by the Google Maps and Waze smartphone navigation apps and finds that both apps submit location data with a high accuracy and use unique identifiers to track users even across several trips [6].

An established metric to measure location privacy is *k-anonymity* [11], originally defined for privacy protection of records in a central database. A record is k-anonymous in a given dataset if it cannot be distinguished from at least $k-1$ other records based on the attributes revealed. Gruteser and Grunwald apply k-anonymity to location privacy, suggesting that a user is k-anonymous if he cannot be distinguished from at least $k-1$ other users based on the location data (position and time) he reveals [10]. They propose to use *spatial and temporal cloaking* of location data for privacy protection, i.e., reducing their accuracy until a predefined level of k-anonymity is met. They employ a central, trusted *anonymity server* that acts as a proxy and calculates the required reduction of accuracy, based on its knowledge of all users' exact position. Our approach is based on the same concept of privacy protection, however, we do not require a trusted, central party. Duckham and Kulik propose a graph based approach to obfuscation in order to degrade the quality of location to the level required by a service provider [20]. Their approach does not require a central, trusted server. Instead, each user applies the location obfuscation individually but protection of their users' identities is not a requirement. Krumm gives a general overview of threats to location privacy and strategies for its protection [21].

There are several approaches to privacy-friendly collection of traffic data. However, their focus is to prevent linking of trip segments, and in particular origin and destination of trips, whereas we propose to make exactly this data available in a privacy-preserving way. Hoh and Gruteser describe a path perturbation algorithm (running on a central, trusted server) that protects location privacy while maintain a certain data quality by provoking path confusion for an attacker trying to track vehicles [22]. The PADAVAN scheme uses anonymous credentials and mix cascades for privacy-friendly collection of traffic densities [23]. As the scheme is explicitly designed to prevent linking of submitted samples, an end-to-end analysis of trips is not possible. Rass et al. describe the privacy-friendly collection of floating car data [24]. They use *sample identifiers* (for individual samples submitted to the server) and *trip identifiers* constructed in such a way that only certain entities can determine which samples belong to the same trip. These entities, however, can reconstruct the trip with full accuracy. Hoh et al. propose a privacy-friendly traffic monitoring system using *virtual trip lines*, where vehicles report to a central database, whenever they cross a virtual trip line, similar to a virtual inductive loop [25]. k-anonymity can be achieved by reducing the temporal accuracy of trip line crossings. Privacy protection is based on a segregation of responsibilities between several central components. Therefore, no single entity can subvert the privacy guarantees. If multiple entities are compromised (or collaborate), though, position updates can be obtained with full accuracy.

In the SOKEN protocol, due to Achenbach et al. [26], mobile users exchange and forward key material in an ad-hoc manner via Bluetooth. Later, two users who wish to communicate can derive a shared secret from their common keys. While the purpose of our system is different, we use a similar mechanism of ad-hoc key exchanges and key forwarding. We also share the authors' assumption that large-scale surveillance of ad-hoc key exchanges via short-range radio is difficult to achieve for an attacker.

III. System model and scenario

We assume a traffic scenario with participating vehicles V_i that are all equipped with V2X communication devices and mobile internet access. They report information about their trips to the trip database. The traffic authority (TA) queries the trip database in order to obtain traffic flow information. We assume that the V2X system is protected by a standard privacy-friendly authentication mechanism [27]. Figure 1 shows an overview of our system model.

A. Attacker model

The attacker's goal is to learn the participants' exact location traces, i.e., who traveled where and when. We consider different types of attackers: The *malicious backend provider* can access all central databases deployed in our scheme, but is unable to eavesdrop on local V2X communication. We argue that this a realistic attacker model as backend providers have full access to the data they store. Ubiquitous surveillance of V2X communication, in contrast, is very hard to achieve as it would require the attacker to be in transmission range whenever two vehicles exchange messages. The *active insider attacker* possesses valid credentials for the V2X system and

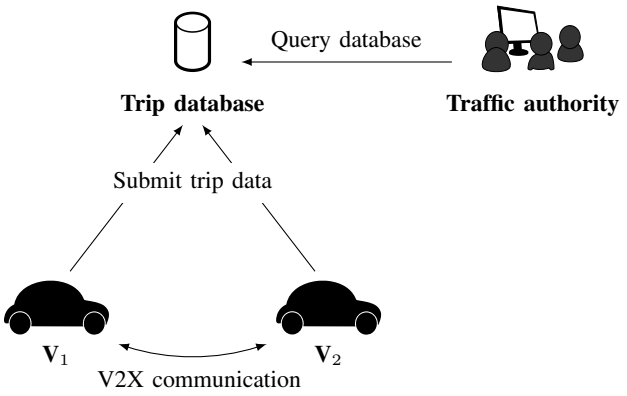

Figure 1: Participating vehicles can exchange information via V2X communication. They also have a mobile data connection to connect to the trip database via internet. Traffic authorities can query the database to obtain information about traffic flows.

actively participates in our system in order to subvert other users' privacy. The *passive insider attacker* has valid credentials, too, but only eavesdrops on communication taking place in his vicinity, without actively participating in our system. The *outsider attacker* is equipped with a V2X communication device, but does not posses valid credentials. (This is a very weak attacker, merely listed for completeness.)

IV. Requirements

We define the following requirements to capture the interests of traffic authorities on the one hand and participating drivers on the other hand:

R.1 Traffic centers require information about traffic flows for the purpose of operational traffic control and assessment of requirements for infrastructure. We assume that while the information does not have to be totally accurate, the higher its accuracy the more useful it is. In particular, origin and destination of trips must be reported together in order to enable macroscopic traffic analysis.

R.2 Drivers require protection of their privacy, quantified by the concept of k-anonymity. They will be reluctant to participate in data collection, if the information they report can be used to create individual mobility profiles. For maximum protection we put forward the requirement of verifiable privacy, i.e., technical protection that augments organizational controls, but has the added benefit that it can be verified by technical means.

V. Privacy-friendly traffic analysis

We first describe the idea behind our approach. Participants upload encrypted reports about their trips to a *trip database*. Multiple copies with different accuracy levels are uploaded and encrypted with different keys. The keys are chosen such that all users that made "the same trip" will use the same key (same trip means same origin, destination and time with respect to the selected accuracy level). The keys are split up using a secret sharing scheme and uploaded, too. A key can be reconstructed when at least k shares of it were uploaded, and the corresponding trip reports can be decrypted. Consequently, the accuracy of each trip report that can be obtained from the database will be such, that it applies to at least k trips. If many participants travel from A to B at the same time, their reports will be revealed with a high accuracy. If somebody travels to a far-off location, on the other hand, only the trip report with very low accuracy will be revealed.

The scheme consists of three phases:

1) Participants establish *location and time-specific keys*, both at the start and destination of their trips.
2) Participants upload copies of their trip reports with different accuracy levels, encrypted with different keys, to the trip database. They apply a secret sharing scheme and upload their shares of the keys, too.
3) Traffic authorities query the trip database. They reconstruct the keys for which enough shares are available and decrypt the corresponding reports. If several reports exist for one trip, all but the one with the highest accuracy are discarded.

Several parameters need to be set system-wide and are valid for all participants:

k – Required size of the anonymity set for trip reports to be revealed to the traffic authority.

Accuracy levels made up by levels of spatial and temporal accuracy, e.g., ((100 m, 1 hour), (1 km, 6 hours), (10 km, 24 hours)). In order to avoid inference attacks by partially overlapping levels of accuracy, we require that for any two accuracy levels (sa_1, ta_1) and (sa_2, ta_2): $sa_1 < sa_2 \Rightarrow ta_1 \leq ta_2$.

p – Modulus used for modular arithmetic in the decentralized secret sharing scheme (cf. Section V-C).

$\mathbf{T_{reconcile}}$, $\mathbf{T_{upload}}$ – Timeouts for key reconciliation and key uploads to the key database (cf. Section V-F).

In the following we cover the building blocks used in our scheme, before we give a complete description of our scheme and its different phases in Section V-F.

A. Location obfuscation

A trip is described by *origin*, *destination*, *start time*, and *arrival time*. k-anonymity can be achieved by reducing the accuracy of each of these properties, until there are $k-1$ other indistinguishable trips. Each *accurate location* (or *accurate time*) can be mapped to a corresponding *coarse location* (or *coarse time*) according to a certain accuracy. For simplicity, we assume that a Cartesian coordinate system is in place.[1] We obtain the coarse location by rounding off the x and y components of the accurate location (e.g., *x=3325 m, y=1876 m* with an accuracy of 250 m becomes *x=3250 m, y=1750 m*). Similarly, the coarse location is obtained by rounding off the accurate location (e.g., 17:46 with a desired accuracy of 1h becomes 17:00). The set of all accurate locations that are mapped to the same coarse location are referred to as a *region*; the set of all points in time that are mapped to the same coarse time is referred to as a *time window*.

[1] When using GPS coordinates, rounding requires additional conversion steps, due to the spherical coordinate system, e.g., using a map projection algorithm.

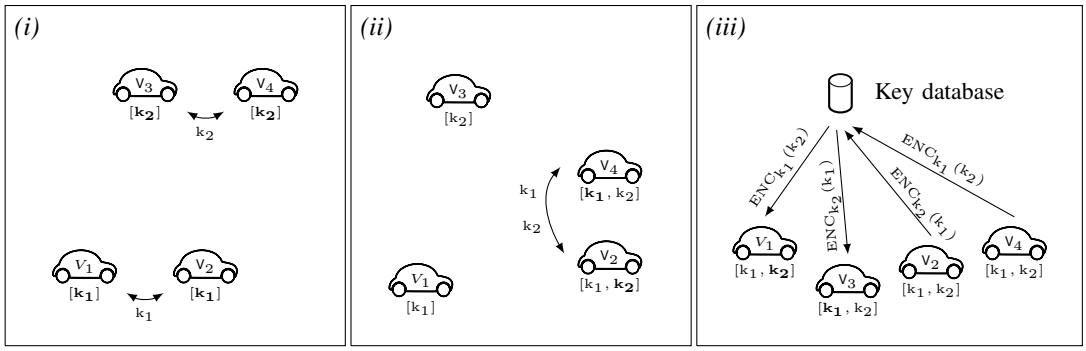

Figure 2: Vehicles generate keys when they meet at the beginning or before the end of their trips *(i)* and forward them while in the respective region and time window *(ii)*. Afterwards keys are synchronized in encrypted form through the key database *(iii)* and an authoritative key can be picked from the common set of keys.

B. Key establishment

We want all participants that were physically present at a certain location at a certain time to share a common *location and time specific key*. With regard to a certain accuracy level, the key should be known to anybody who was present in the *region* that maps to a specific coarse location during the *time window* that maps to a specific coarse time. Several keys (for different accuracy levels) can be established independently and at the same time. Each key record contains the attributes *fingerprint, accuracy level, coarse time, coarse location* and the *cryptographic key* itself. Let $ID(key)$ denote a key's fingerprint and $ENC_{key}(p)$ the symmetric encryption of some plaintext p using the key.

We describe how vehicles establish a key for a specific location and time at a specific accuracy level. The procedure must be run independently for each accuracy level defined in the system parameters (cf. Section V):

1) Map current accurate time and accurate location to coarse location (*region*) and coarse time (*time window*), according to the selected accuracy level.
2) While the vehicle is within the region and time window, indicate readiness to exchange keys, e.g., using a flag in the V2V message sent out. When another participating vehicle comes into communication range, which is ready to exchange keys, forward and receive all preliminary keys (for the current time and location window) that have been obtained before. If no keys were forwarded in either direction, establish a new *preliminary key* (e.g., using Diffie-Hellman). Stop key exchanges and forwarding, once the vehicle leaves the region or the time window.
3) Derive an *authoritative key* from all preliminary keys as follows.
   a) Let S be the set of preliminary keys for the current region and time window. For each pair
   $$sk_i, sk_j \in (S \times S), sk_i \neq sk_j$$
   create the encrypted key record
   $$ID(sk_i), ID(sk_j), ENC_{sk_j}(sk_i)$$
   and upload it to the central *key server*. (The server removes any duplicate uploads.)
   b) Download and decrypt all records of encrypted keys that are not stored locally yet, but for which the encryption key is available. Create and upload records for newly downloaded keys which are not stored on the server yet. Wait some time and repeat until $T_{reconcile}$ elapses.
   c) Sort all keys lexicographically. The first key is the authoritative key for the current time and location window.

The procedure is based on the assumption that all participants present at the given region within the given time window are connected through paths of common and forwarded keys. If this is the case, they will all eventually obtain the same authoritative key, provided step 3 (b) is repeated often enough. If not, the accuracy with which the trips will be revealed later on will degrade, but privacy protection remains intact. For practicality, the reconciliation phase is limited by a timeout $T_{reconcile}$. Figure 2 shows a high-level sketch of the key establishment procedure.

Key exchanges are only conducted among vehicles that posse valid credentials for the V2X system. All V2V communication links are encrypted to protect against local eavesdroppers, e.g., using Diffie-Hellman keys. To prevent identification based on network addresses, all connections to the key database are made through an anonymization network, such as Tor[2].

C. Decentralized, non-interactive secret sharing

Assume a common secret s, shared by an unknown number of parties. We want each party to derive some information from that secret, called a *share*, such that s is revealed only when at least k parties reveal their share.

We base our construction on Shamir's secret sharing [28]. In the original scheme the secret s is only known to a central trusted party which generates the shares and distributes them among the participants. The shares are created by constructing a polynomial $f(x)$ of degree k with random coefficients, such that $f(0) = s$. Each of the n parties ($n > k$) obtains one point of the polynomial $(x_i, f(x_i))$, while the polynomial

[2]https://www.torproject.org/

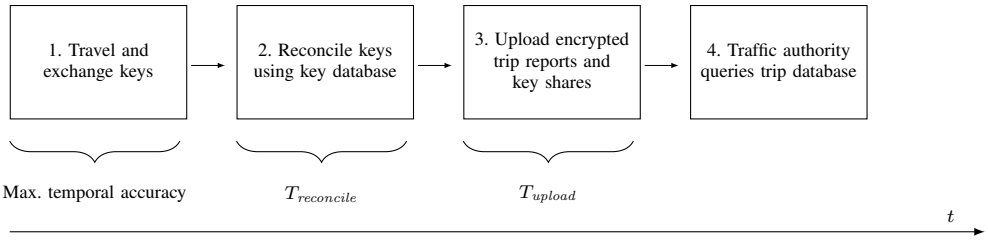

Figure 3: High-level overview of processing steps. The length of each phase (but the last one) is specified and each phase must be completed by all participants, before the next step can begin.

itself is kept secret. Consequently, any k of the n parties can collaborate and reconstruct the full polynomial and reveal the secret. All computations are done using modular arithmetic.

Our setting is slightly different because each party knows the secret s, but must construct its share independently from the others. Using a cryptographic hash function h, each party can (by itself) obtain the coefficients

$$a_i := h(i||s) \text{ for } i \in [1, k]$$

and construct

$$f(x) = s + \sum_{i=1}^{k} a_i x^i \mod p.$$

Note, that all parties will obtain the same polynomial. Then each party chooses x_r at random from a sufficiently large range to avoid collisions and calculates its share $(x_r, f(x_r))$. Like in the original construction, s will be revealed when at least k of the participants make their share available. We use $share(s, k)$ to denote the creation of a share. For a practical implementation the secret s and the output of h must be converted to numbers and the prime p used for modular computation must be larger than any possible value of s.

D. Build and upload trip reports

Assume a participant has completed a trip and the location and time specific keys $origin\_key_i$ and $destination\_key_i$ have been established for each accuracy level AL_i, at the trip's origin and destination respectively. For each accuracy level he creates and uploads a trip report as follows:

1) Create $trip\_key_i := h(origin\_key_i||destination\_key_i)$ using a cryptographic hash function h.
2) Create the trip report rep containing the coarse locations of *origin* and *destination* and coarse *start time* and *arrival time* with respect to the current accuracy level.
3) Create the *encrypted trip record*

   $$ID(trip\_key_i), share(trip\_key_i, k), ENC_{trip\_key_i}(rep)$$

   and upload it to the trip database.

All connections to the trip database are made through an anonymization network.

E. Reconstruction of trip reports

Query the trip database for all trip records that can be decrypted. Specifically, download records for which at least $k - 1$ other records are available which have been encrypted with the same key. Reconstruct the trip keys from the shares included in the records and decrypt the trip reports.

F. Phases of operations

The building blocks described in Sections V-A to V-E are executed sequentially in different, dependent phases (cf. Figure 3).

1) Participants exchange location and time specific keys at the beginning and end of their trips. For each accuracy level keys are exchanged independently, while the vehicle is in the origin or destination region and start or end time window (with respect to that accuracy level). The beginning of a trip can be identified trivially, however, some trigger is required that signals the upcoming end of the trip, e.g., from the navigation system. Alternatively, keys can exchanged continuously during the trip, so that the keys for the end of the trip can be determined retrospectively when the vehicle is turned off. Continuous key exchanges can also improve the connectivity for other participants, if their keys are "carried and forwarded" within their validity regions and time windows.
2) Key reconciliation (which involves uploading encrypted keys to the key database) must only be started, when the time window for which the keys are valid has ended. If keys were uploaded too early, it would be possible to infer the end time of the respective trip more accurately than intended. For practical reasons and in order to execute the phase for all accuracy levels simultaneously, we propose to begin the phase only after the time window for the lowest temporal accuracy (i.e., the longest one) has ended. The length of the reconciliation phase $T_{reconcile}$ must be sufficiently long to allow all involved vehicles (which may not be online all the time) to perform multiple iterations of the reconciliation protocol.
3) Trip uploads must only be performed after the previous phase was completed because the authoritative keys may not be available before. It should be completed within the time T_{upload}.
4) The trip database may be queried at any time. However, the trip reports will only be available after the previous phases were completed.

VI. Evaluation

We evaluate our system with regard to the attacker model described in Section III-A and examine its performance in a specific scenario using simulations.

A. Security analysis

Our scheme is secure against the *malicious backend provider*, i.e., he cannot obtain more information than any honest party, that queries the trip database. Even with full access to the key database and the trip database, he would have to break the secret sharing scheme (which is information-theoretically and even perfectly secure) or the encryption itself. He could delete or alter records in the key database, which would sabotage the establishment of common keys, or manipulate the trip database. These attempts would, in fact, affect the availability of trip reports, but not have any negative effect on participants' privacy. We emphasize that even though the key database and the trip database are central components in our systems, they need not be trustworthy, as all the sensitive data they hold is encrypted.

By regular participation in our scheme, the *active insider attacker* can collect location and time specific keys and reveal them without applying the secret sharing scheme. This would, in fact, subvert the privacy of all participants that used the keys to encrypt their trip reports. However, the attack is quite limited because only those trips can be revealed where the attacker was physically present both at the origin and destination. Yet, an *active insider attacker* with large-scale physical presence, e.g., a malicious operator of a dense network of V2X road-side units that might be deployed in the future, poses a serious threat to our system.

The *passive insider attacker* and the *outsider attacker* are equally weak and cannot interfere with our system in any meaningful way. Even though they can eavesdrop on V2X communication in general, the exchange and forwarding of keys is protected from them by the encrypted communication channel.

k-anonymity towards the traffic authority is guaranteed when only one accuracy level is used. When using different accuracy levels (which makes the scheme more practical) special cases can be constructed in which k-anonymity can be violated by combining information from different accuracy levels: Consider a set of k trips with high accuracy that is contained in a set of $k+1$ trips with lower accuracy. As both sets can be decrypted, some information about the trip that is only in the coarse set can be inferred. If this information is considered sensitive in a specific scenario, the scheme must be deployed with only one accuracy level.

B. Simulation results

We evaluate our system's performance in a specific simulation scenario and focus on two aspects: 1) Is our V2X-based approach for key establishment suitable for deriving common authoritative keys among vehicles within the same region and time window? 2) How does the reduction of accuracy affect the information available to the traffic authority?

To answer the first question, we compare the results from our scheme to the theoretical optimum, that could be reached using a central privacy proxy that has access to all accurate trip data and decides for each trip at which accuracy levels it can be revealed while maintaining k-anonymity. For the second question, we examine how many trips are revealed at different accuracy levels (and for different values for k), both for our scheme and for the theoretical optimum.

Traffic was generated using the SUMO traffic simulator and the *LuST* traffic scenario [29]. The scenario provides 24 hours of synthetically generated, yet realistic, traffic in the city of Luxembourg and covers an area of approximately 156 km^2. We removed the public buses from the scenario, considering only passenger vehicles, and ended up with a total of 218 938 trips. In order to cope with the large number of vehicles and the long simulation time, we generated the traffic traces offline. Then we ran our Python-based implementation on the traces, assuming radio connectivity between two vehicles, whenever they are within a fixed communication range (100 or 200 m). We evaluated two variants of our scheme: In the *start/end* variant vehicles exchange keys only when they are within the origin or destination regions (and within the start or end time windows). In the *whole trip* variant keys are exchanged during the whole trip. Outside the origin or destination regions and start or end time windows, keys are only forwarded in order to increase the connectivity among other participants and discarded after leaving the respective regions.

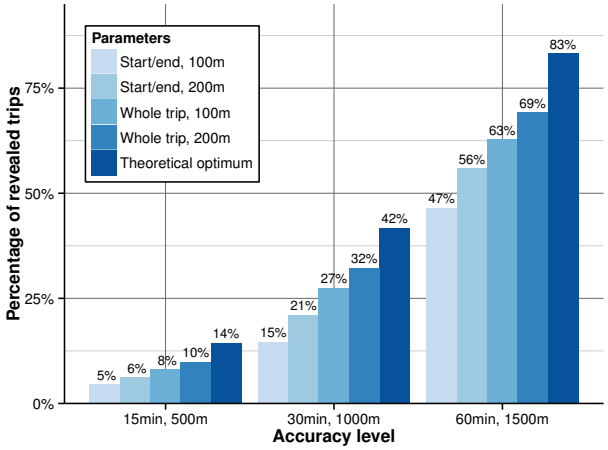

Figure 4: Percentage of revealed trips for $k = 3$, comparing our simulation results with the theoretical optimum at different accuracy levels and for different parameters.

Figure 4 displays the number of revealed trips for $k = 3$ at different accuracy levels for different variants in comparison to the theoretical optimum. At the lowest accuracy level a significant number of trips are revealed (69% for the *whole trip* variant and a communication range of 200 m), which is a significant share of the theoretical optimum of 83%. For higher accuracy levels less trips are revealed. However, the results for our scheme are still relatively close to the theoretical optimum. This suggests that the key exchange mechanism performs well, but that the specific traffic pattern does not allow for trips to be revealed at those accuracy levels without violating the k-anonymity boundary. The communication range has a significant impact on the results. While we were unable to conduct detailed simulations on the physical network layer

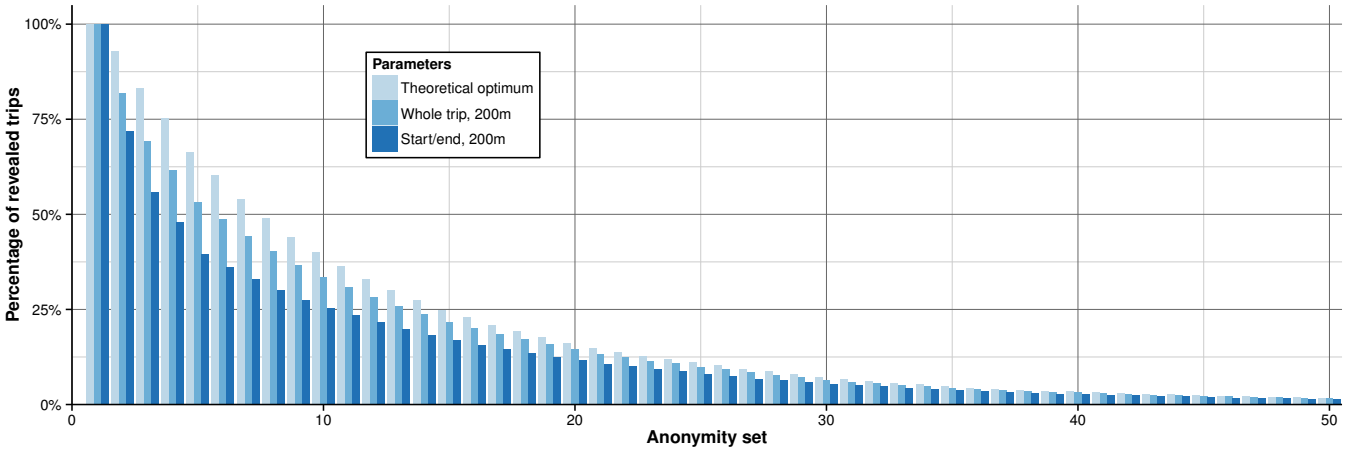

Figure 5: Percentage of trips revealed for a given value of k: Cumulative distribution function (x-axis truncated) of anonymity sets for the theoretical optimum and two simulation scenarios for an accuracy level of 60 min and 1500 m.

due to the size of the scenario, related work [16] suggests that our assumed parameter choices of 100 and 200 m are in fact realistic. Our scheme performs significantly better in the *whole trip* variant, where continuous key exchanges outside of origin and destination regions help other participants establishing common keys.

Figure 5 displays the cumulative distribution function of anonymity sets for an accuracy level of 60 min and 1500 m, i.e., what fraction of trips would be revealed for a given choice of k. The share of revealed trips drops rather quickly for higher values of k. For $k = 10$, in the *whole trip* variant and a communication range of 200 m, 34% of trips are revealed, compared to the theoretical optimum of 40%, while for $k = 20$, only 15% are revealed for the same parameters, compared to the theoretical optimum of 16%. Again, we can can see that our scheme performs reasonably well, but that the k-anonymity constraint severely limits the revelation of information.

Overall, the simulations show that the V2X-based key exchange mechanism works well and that our scheme can provide information about a significant share of traffic at an accuracy level that we expect is still useful practice.

VII. Conclusion

We propose a generic mechanism for enforcing k-anonymity for location privacy based on secret sharing. Using a decentralized version of Shamir's secret sharing [28], participants can make location information available in encrypted form together with a share of the key. It will only be revealed, once $k-1$ other parties made available the same location information. This is particularly useful, when location information is made available with different levels of accuracy, resulting in the information being revealed with the highest possible accuracy such that it still applies to at least k distinct users. Note that when using different accuracy levels, special cases can be constructed in which k-anonymity can be violated by combining information from different levels.

To establish the practicality of our proposal, we describe a traffic monitoring system, where participants make available origin, destination and start and end times of their trips to a traffic authority. For privacy protection the accuracy of time and location information is reduced, such that each report applies to at least k trips. We evaluate our scheme in a simulation scenario with 24 hours of synthetic, but highly realistic traffic in the city of Luxembourg and compare our results with the theoretical optimum, that could be achieved by having a central, trusted party calculate the minimum reduction of accuracy required to satisfy the k-anonymity requirement. Our results show a that significant share of trips is revealed for a rather coarse accuracy level, while less trips are revealed for higher accuracy levels. We conclude that our scheme performs rather well and that the smaller share of trips revealed for higher accuracy levels (and larger values of k) is due to the anonymity requirement itself. It is not surprising that it is much harder to enforce k-anonymity for origin/destination pairs than for single locations. In fact, most related approaches for privacy-friendly collection of traffic data aim for unlinkability of origin/destination pairs for that very reason.

With our work we show that privacy-friendly collection of origin/destination pairs is in fact possible, although a significant loss of accuracy (or share of revealed trips) must be accepted. We expect that the described traffic monitoring system could be deployed and deliver useful information at different scales: In an urban context (as done in our simulation scenario), across several cities, e.g., in order to analyze requirements and efficiency of highway systems, or even across several countries, e.g., to find out where people from certain regions spend their vacation. As the mechanism for decentralized enforcement of k-anonymity is quite generic, we envision its application for location privacy in other scenarios and beyond.

References

[1] T. Abrahamsson, "Estimation of origin-destination matrices using traffic counts – a literature survey", International Institute for Applied Systems Analysis, Tech. Rep. IR-98-021, May 1998.

[2] J. White and I. Wells, "Extracting origin destination information from mobile phone data", in *Eleventh International Conference on Road Transport Information and Control*, IET, Mar. 2002, pp. 30–34.

[3] N. Caceres, J. Wideberg, and F. Benitez, "Deriving origin destination data from a mobile phone network", *Intelligent Transport Systems, IET*, vol. 1, no. 1, pp. 15–26, Mar. 2007.

[4] S. Turksma, "The various uses of floating car data", in *Road Transport Information and Control, 2000. Tenth International Conference on (Conf. Publ. No. 472)*, Apr. 2000, pp. 51–55.

[5] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway", *Transportation Research Record: Journal of the Transportation Research Board*, no. 1855, pp. 49–59, 2003.

[6] T. Jeske, "Floating car data from smartphones: What google and waze know about you and how hackers can control traffic", *Proceedings of the BlackHat Europe*, 2013.

[7] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems", *Pervasive Computing, IEEE*, vol. 5, no. 4, pp. 38–46, 2006.

[8] J. Krumm, "Inference attacks on location tracks", in *Pervasive Computing*, Springer, 2007, pp. 127–143.

[9] P. Golle and K. Partridge, "On the anonymity of home/work location pairs", in *Pervasive Computing*, Springer, 2009, pp. 390–397.

[10] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking", in *Proceedings of the 1st international conference on Mobile systems, applications and services*, ACM, 2003, pp. 31–42.

[11] L. Sweeney, "K-anonymity: A model for protecting privacy", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[12] BBC, "The interview: A guide to the cyber attack on hollywood", Dec. 2014. [Online]. Available: http://www.bbc.com/news/entertainment-arts-30512032 (visited on 09/03/2015).

[13] The Guardian, "Surveillance", Sep. 2015. [Online]. Available: http://www.theguardian.com/world/surveillance (visited on 09/03/2015).

[14] General Motors, *Cadillac to introduce advanced 'intelligent and connected' vehicle technologies on select 2017 models*, Sep. 2014. [Online]. Available: http://media.gm.com/media/us/en/gm/news.detail.html/content/Pages/news/us/en/2014/Sep/0907-its-overview.html.

[15] "IEEE standard for information technology–telecommunications and information exchange between systems–social and metropolitan area networks–specific requirements part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications", *IEEE Std. 802.11-2012*, 2012.

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

[17] H. Hartenstein and K. Laberteaux, *VANET vehicular applications and inter-networking technologies*. John Wiley & Sons, 2009, vol. 1.

[18] U.S. Departement of Transportation – National Highway Traffic Safety Administration, "Federal motor vehicle safety standards: Vehicle-to-vehicle (V2V) communications; advance notice of proposed rulemaking (ANPRM); Docket No. NHTSA–2014–0022", *Federal Register*, vol. 79, no. 161, Aug. 2014.

[19] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing", *IEEE Pervasive computing*, vol. 2, no. 1, pp. 46–55, 2003.

[20] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy", in *Pervasive computing*, Springer, 2005, pp. 152–170.

[21] J. Krumm, "A survey of computational location privacy", *Personal and Ubiquitous Computing*, vol. 13, no. 6, pp. 391–399, 2009.

[22] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion", in *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*, IEEE, 2005, pp. 194–205.

[23] A. Tomandl, D. Herrmann, and H. Federrath, "Padavan: Privacy-aware data accumulation for vehicular ad-hoc networks", in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2014 IEEE 10th International Conference on*, IEEE, 2014, pp. 487–493.

[24] S. Rass, S. Fuchs, M. Schaffer, and K. Kyamakya, "How to protect privacy in floating car data systems", in *Proceedings of the fifth ACM international workshop on VehiculAr Inter-NETworking*, ACM, 2008, pp. 17–22.

[25] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson, "Virtual trip lines for distributed privacy-preserving traffic monitoring", in *Proceedings of the 6th international conference on Mobile systems, applications, and services*, ACM, 2008, pp. 15–28.

[26] D. Achenbach, D. Förster, C. Henrich, D. Kraschewski, and J. Müller-Quade, "Social key exchange network – from ad-hoc key exchanges to a dense key network", in *Tagungsband der INFORMATIK 2011, Lecture Notes in Informatics*, vol. P192, Oct. 2011.

[27] P. Papadimitratos, L. Buttyan, T. Holczer, E. Schoch, J. Freudiger, M. Raya, Z. Ma, F. Kargl, A. Kung, and J.-P. Hubaux, "Secure vehicular communication systems: Design and architecture", *Communications Magazine, IEEE*, vol. 46, no. 11, pp. 100–109, 2008.

[28] A. Shamir, "How to share a secret", *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[29] L. Codeca, R. Frank, and T. Engel, "Lust: A 24-hour scenario of Luxembourg city for SUMO traffic simulations", in *SUMO User Conference 2015-Intermodal Simulation for Intermodal Transport*, 2015.