Differential Privacy in Intelligent Transportation Systems

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

Arik Friedman, Roksana Boreli
NICTA
Sydney, Australia
givenname.surname@nicta.com.au

ABSTRACT

In this paper, we investigate how the concept of differential privacy can be applied to Intelligent Transportation Systems (ITS), focusing on protection of Floating Car Data (FCD) stored and processed in central Traffic Data Centers (TDC). We illustrate an integration of differential privacy with privacy policy languages and policy-enforcement frameworks like the PRECIOSA PeRA architecture. Next, we identify differential privacy mechanisms to be integrated within the policy-enforcement framework and provide guidelines for the calibration of parameters to ensure specific privacy guarantees, while still supporting the level of accuracy required for ITS applications. We also discuss the challenges that the support of user-level differential privacy presents and outline a potential solution. As a result, we show that differential privacy could be put to practical use in ITS to enable strong protection of users' personal data.

Categories and Subject Descriptors

C.2.1 [**Computer-Communications Networks**]: Network Architecture and Design—*Wireless communication*

Keywords

Differential Privacy; Intelligent Transportation Systems; ITS; Privacy

1. INTRODUCTION

Intelligent Transportation Systems (ITS), i.e., the introduction of information and communication technology into transportation systems, and especially vehicles, are generally considered as means to achieve safer, more efficient, and greener road traffic. While some approaches like Car-to-Car communication are still experimental, use of Floating Car Data (FCD) is a more mature ITS technology that is already deployed in the field in many (proprietary) applications. The idea is to turn a vehicle into a mobile sensor that periodically reports its status to a central backend, like a Traffic Control Center (TCC), by means of a standardized data set, the FCD record. FCD data includes at minimum a timestamp and the vehicle position, but may also include additional data like speed or on-board information from ABS and ESC sensors to detect, e.g., icy roads.

FCD records are used in a variety of applications ranging from fleet management to insurance and tolling applications. Early adopters of FCD include taxi fleets, e.g., in the city of Vienna, where about 2,100 taxis submit FCD records[1], which are then used by the TCC to gain a fine-grained picture of traffic situation on all major roads.

Despite the benefits of ITS and FCD applications, their use also brings concerns that drivers' privacy may be negatively affected. Therefore, FCD records are anonymized in many applications so that they do not contain information that would allow direct identification of specific drivers or vehicles. While this may be a first step towards privacy protection, some identifiers (at least pseudonymous) must still be retained to enable attribution of two successive FCDs to the same car. Otherwise, car counts will not be reliable. As was proposed in previous works [9], a privacy protection mechanism such as k-anonymity may be applied to prevent disclosure of private information. However, this protection can be circumvented, and detailed mining of the FCD database might still reveal a lot of private information about drivers and driving behavior, as shown, e.g., in [15].

The question we want to investigate in this paper is how privacy can be protected more reliably and provably in the context of such data collections in ITS, while still allowing reasonable use for traffic analysis or dedicated applications like road tolling. To this end we focus on differential privacy [6], a formal definition of privacy that allows aggregate analysis while limiting the influence of any particular record on the outcome, typically through the introduction of noise.

Throughout the paper we focus on the following motivating scenarios in ITS. **Scenario 1: identification of traffic conditions** – assessment of traffic conditions, e.g., by calculating the average speed of cars in a certain road segment. Tasks that rely on aggregate information represent the key scenario we would like to accomplish with differential privacy. **Scenario 2: detection of speeding vehicles** – law enforcement agencies who are granted access to FCD databases may be tempted to leverage this access to track and monitor individual drivers. However, this could deter individuals from participating in such schemes. We will show how differential privacy in ITS can mitigate such

[1] `http://www.wien.gv.at/verkehr/verkehrsmanagement/verkehrslage/projekt.html`

WiSec'13, April 17-19, 2013, Budapest, Hungary.

privacy breaches. **Scenario 3: eTolling fee calculation** – some applications may nevertheless require access to detailed FCD records, for example, to calculate a road toll based on tracks of journeys. Such applications could be addressed by complementing security mechanisms, beyond differential privacy.

In this paper, we address the challenges in applying differential privacy in practical ITS applications and provide the following contributions:

1) We propose an architecture that integrates differential privacy and additional security mechanisms to provide a comprehensive solution to privacy in ITS.

2) We demonstrate how differentially private mechanisms can be utilized in ITS applications, addressing the accuracy requirements of these applications.

3) We investigate how the privacy parameters can be calibrated within application accuracy requirements, while also considering long-term privacy consequences for the end-user.

2. BACKGROUND AND RELATED WORK

2.1 Privacy Enhancing Technologies in ITS

Protection of private data in ITS has been addressed in the past, often focusing on singular applications and scenarios. As one example, Troncoso et. al. [14] addressed the challenge of privacy-preserving Pay-As-You-Drive (PAYD). Instead of submitting FCD records to the insurance company and having the insurance company calculate the resulting fee, the PriPAYD scheme foresees a trustworthy hardware box installed in the vehicle, which calculates the fee and submits it to the insurance company but without revealing any FCD data. The FCD records are instead given to the driver on USB stick in encrypted form together with a share of the secret key. The second half of the key is given to the insurance company. In case of dispute, both key shares can be combined and the FCD data can be accessed. This way, the driver has full control of the data and can explicitly agree to reveal it to the insurance company.

While many of these approaches achieve the goals of the individual scenario, they have the drawback that they are highly specific and cannot easily be generalized to arbitrary data and arbitrary data processing. Furthermore, the privacy protection relies on the fact that all data processing happens in one On-Board Unit (OBU) and that data leaking from this OBU can be controlled and monitored by the driver. Processing that requires combination of FCD data from different vehicles (e.g., average speed of all vehicles in a given road segment) does not fit into this architecture.

The EU FP7 project PRECIOSA proposed a different approach to privacy preserving data processing in ITS [10, 11]. The *PRECIOSA Privacy-enforcing Runtime Architecture* (PeRA) foresees protection of personal data by augmenting these data with privacy policies and mandatory enforcement of these policies in a distributed system. Whenever personal data are used or communicated, there should also be a policy expressed in the *PRECIOSA Privacy Policy Language* (P3L) that describes the operations allowed on these data. Applications access the data via a dedicated query interface using a SQL-like language called *PRECIOSA Privacy aware Query Language* (PPQL). The *Policy Control Monitor* (PCM) checks the compliance of queries with policies of affected data and either grants or denies access. PeRA is designed to work locally or in a distributed system, the latter case creating a policy enforcement perimeter that can span multiple systems. Within the boundaries of the perimeter, data subjects can rest assured that their personal data are only used in a policy compliant way.

In PeRA, a vehicle transmits data like FCD records together with policies through a confidential communication channel to the importer of a Traffic Control Center. Both data and policy are stored in an encrypted way in the repository and are only accessible via the PCM. PPQL queries can be issued by applications via the Query-API. This approach provides a generic solution to support arbitrary ITS applications, data formats, and operations. It could easily be combined with schemes like PriPAYD to ensure policy compliant data processing in the OBUs and backends.

The concept of differential privacy promises to set hard limits to privacy loss when contributing personal data to a database. However, it has not yet been applied to ITS and its specific applications. In this paper, we will explore how the concept of differential privacy can practically be integrated into the PRECIOSA PeRA framework to provide stronger privacy guarantees for FCD-like applications.

2.2 Differential Privacy

Differential Privacy [6] is a formal definition of privacy that allows computing fairly accurate statistical queries over a database while limiting what can be learned about single records. The privacy protection is obtained by constraining the effect that any single record could have on the outcome of the computation.

DEFINITION 2.1 ((ϵ, δ)-DIFFERENTIAL PRIVACY [5]). *A randomized computation M maintains (ϵ, δ)-differential privacy if for any two multisets A and B with symmetric difference of a single record (i.e., $|A \Delta B| = 1$), and for any possible set of outcomes $S \subseteq Range(M)$,*

$$Pr[M(A) \in S] \leq Pr[M(B) \in S] \cdot \exp(\epsilon) + \delta \ ,$$

where the probabilities are taken over the randomness of M.

Setting $\delta = 0$ amounts to ϵ-differential privacy.

The ϵ parameter controls the privacy/accuracy tradeoff, as it determines the influence that any particular record in the input could have on the outcome. The δ parameter allows ϵ-differential privacy to be breached in some rare cases.

Differentially private computations can be composed, as shown in [5]: a series of n computations, where computation i is (ϵ_i, δ_i)-differentially private, will result in the worst case in a computation that is $(\sum \epsilon_i, \sum \delta_i)$-differentially private. Therefore, when records enter and leave the database frequently, it is possible to ensure (ϵ, δ)-differential privacy for each record by monitoring the computations performed over the database while the record was in it, and ensuring that the sum of privacy parameters for these computations does not exceed the ϵ and δ bounds.

In this work we focus on *event-level privacy* [8], where the privacy protection is with respect to single records in the database, as in Definition 2.1. In contrast, *user-level privacy* [8] considers the combined effect of all records in the database that pertain to a specific user (or vehicle, in our case). When the number of these records is bounded by c, ϵ-differential event-level privacy amounts to $c \cdot \epsilon$-differential user-level privacy due to composability. In Section 4 we further discuss the user-level privacy.

2.2.1 Privacy Through Perturbation

One of the prevalent methods to achieve differential privacy is the Laplace mechanism [6], in which noise sampled from Laplace distribution is added to the value of a computed function. The probability density function of the Laplace distribution with zero mean and scale b is $f(x) = \frac{1}{2b}e^{-\frac{|x|}{b}}$, and its variance is $2b^2$. The noise is calibrated to the global sensitivity of the function, which is the maximal possible change in the value of the function when a record is added to the database or removed from it.

THEOREM 2.1 (LAPLACE MECHANISM [6]). *Let $f : D \to \mathbb{R}^d$ be a function over an arbitrary domain D. Then the computation $M(X) = f(X) + (Laplace(S_G(f)/\epsilon))^d$, where $S_G(f) = \max_{A \Delta B=1} \|f(A) - f(B)\|_1$, maintains ϵ-differential privacy.*

EXAMPLE 2.1. *Consider a database of FCD records, where each record includes the speed of a car in km/h. The speed is a number between 0 and 120, and any reported speed outside this range is clamped. Then the following approximations maintain ϵ-differential privacy: 1) Calculating the number of FCD records in the database:* `Count(*)` + Laplace$(1/\epsilon)$*; 2) Calculating the sum of reported speeds:* `Sum(speed)` + Laplace$(120/\epsilon)$*; 3) Calculating the average speed of cars:* $\frac{\texttt{Sum(speed)}+\text{Laplace}(240/\epsilon)}{\texttt{Count(*)}+\text{Laplace}(2/\epsilon)}$*. In the last example, we combine two queries, where each query maintains $\frac{\epsilon}{2}$-differential privacy.*

3. CHALLENGES IN THE APPLICATION OF DIFFERENTIAL PRIVACY TO ITS

While differential privacy allows to reason formally on the privacy guarantees, it also poses some challenges that may hinder its application in practical systems like ITS.

Computing global-sensitive functions: The `Count`, `Sum` and `Average` functions capture many of the calculations utilized in ITS, and can be evaluated accurately with differential privacy, enabling, e.g., Scenario 1. However, `Max` and `Min` are also valuable functions (e.g., evaluate the speed of the slowest and fastest vehicles in a road section), but have high global sensitivity. Consequently, applying the Laplace mechanism as in Theorem 2.1 to evaluate these functions would provide useless results. We discuss in Section 4.2.1 how techniques relying on local sensitivity [13] can be adapted to overcome this limitation in typical scenarios.

Supporting applications that require precise information: Some applications of ITS require access to precise information. For example, calculating eTolling fees (Scenario 3) is an application, where introduction of noise may be unacceptable as it may result in wrong bills[2]. Noise may also be unacceptable in other applications, such as some safety applications that may have life-and-death consequences. In the scope of this work we focus mainly on applications where noise is acceptable, and even desirable for privacy protection. Other scenarios may be handled through the Controlled Application Environment (CAE), which is part of the existing PRECIOSA framework [4].

Processing time-series data: Differential privacy limits the privacy loss in each query. However, as additional queries are answered by the database, the privacy loss may accumulate. Since differential privacy maintains composability, it is possible to monitor the overall privacy loss (a worst-case evaluation) and bound it. To address the risk incurred by continuous queries, we describe in Section 4.3.2 an expiry mechanism that ensures that FCD records are removed from the database after participating in a certain amount of queries.

Obtaining user-level privacy: While the privacy loss per FCD record can be monitored and bounded, and thus event-level privacy can be obtained, ensuring user-level privacy is a much more difficult problem. At any point in time, it is possible that multiple FCD records pertaining to the same vehicle (and driver) would be retained in the system and new records that correspond to the same vehicle may be added to the database. Consequently, while differential privacy may prevent an adversary from learning of a specific FCD record that indicates speeding, it does not necessarily prevent from learning that a specific vehicle is frequently speeding. There are theoretical bounds [7] that indicate that such leaks cannot be prevented while still keeping the system usable. However, we use similar arguments in Section 4.3.4 to motivate the choice of ϵ in a way that would quantify this inherent risk.

[2]Though Danezis et. al. [3] proposed a private method for billing, where rebates are issued periodically to compensate for billing errors introduced by differentially private noise.

4. DIFFERENTIAL PRIVACY FOR ITS

In this section, we detail our proposal for a system that enables differentially private use of FCD data for selected ITS applications and services, through an extension of the PRECIOSA PeRA policy enforcement framework.

4.1 System Architecture

The proposed Differential Privacy-enhanced PeRA architecture is shown in Figure 1. For the sake of clarity, we only show the main components relevant to this discussion.

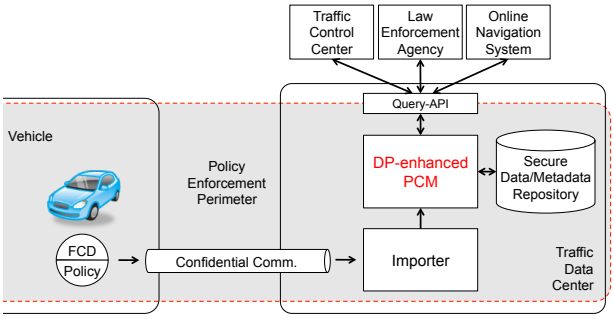

Figure 1: Architecture for enabling the differentially private aggregation of data collected from vehicles in ITS applications.

In line with the existing PeRA architecture, the collection of users' FCD records from the corresponding vehicles is done using a confidential communication channel between the vehicle and the Traffic Data Center (TDC). Collected records are stored in the secure data repository within the TDC. All applications access the FCD data via the Query interface using a set of PPQL queries. As discussed in Section 2, the PRECIOSA P3L policy language already includes the means for expressing, e.g., k-anonymity as a requirement. PPQL enables the formulation of data access queries

and the Policy-Control-Monitor (PCM) acts as an enforcement point for privacy control.

The enhancements required to enable differential privacy include the introduction of a DP-Enhanced Policy Control Monitor (DP-enhanced PCM in Figure 1) and the extension of the P3L policy language to enable specifying a set of selected differential privacy parameters, for every FCD or other data record (or set of data records referring to the same event, e.g., position)[3]. These would reflect the level of privacy loss acceptable to the data subject, or as defined by the applicable data protection regulation.

4.2 The Differential-Privacy-enhanced PCM

Differential privacy is suitable for applications that operate on aggregated data, such as the task of assessing traffic conditions outlined in Scenario 1. Such applications access the Traffic Data Center through the Query-API.

In a simple solution, the PCM can use the Laplace mechanism to estimate `Count`, `Sum` and `Average` queries based on their global sensitivity, as was described in Section 2.2. In the next section we demonstrate how additional techniques from the differential privacy literature [13] can be leveraged to evaluate with reasonable accuracy also functions such as `Max` and `Min`, which are frequently used in ITS applications.

4.2.1 Smooth Sensitivity

For some differentially-private computations, the global sensitivity may be too large, and consequently, introducing noise proportional to the global sensitivity would destroy the utility of the computation. For example, the global sensitivity of the max and min functions, computed over values in the range $[0, \Lambda]$, is Λ, and the Laplace mechanism would require adding noise of magnitude Λ/ϵ, consequently destroying utility. To counter this problem, Nissim et. al. [13] proposed adding data-dependent noise. To this end, they defined the *local sensitivity* of a function.

DEFINITION 4.1 (LOCAL SENSITIVITY [13]). *Let $f : D \to \mathbb{R}^d$ be a function over an arbitrary domain D. The local sensitivity of f at point x is*

$$LS_f(X) = \max_{Y:d(X,Y)=1} \|f(X) - f(Y)\|_1 \quad , \tag{1}$$

where $d(X, Y)$ is the distance between datasets.

Unfortunately, adding noise calibrated to the local sensitivity may still compromise privacy – since the magnitude of noise depends on the data, it becomes a leak channel. To ensure that the magnitude of noise also maintains differential privacy, the concept of *smooth sensitivity* is introduced. While local sensitivity may vary significantly between neigboring datasets, smooth sensitivity changes gradually, and the difference in sensitivity between neighboring datasets is controlled by a parameter β.

DEFINITION 4.2 (SMOOTH SENSITIVITY [13]). *For $\beta > 0$, the β-smooth sensitivity of f at point x is*

$$S^*_{f,\beta}(X) = \max_{Y \in D} (LS_f(Y) \cdot exp(-\beta \cdot d(X, Y)) \quad . \tag{2}$$

EXAMPLE 4.1. *Let $X = \{x_1, \ldots, x_n\}$, where $0 \le x_1 \le \cdots \le x_n \le \Lambda$. The local sensitivity of the function $f_{\min}(X) = \min(x_1, \ldots, x_n)$ at point X is $LS_{f_{\min}}(X) = \max(x_1, x_2 - x_1)$.*

[3]For readability, we will continue our discussion referring just to one FCD record, however other data records or sets of records could be treated the same way.

Nissim et. al. [13] show that the β-smooth sensitivity of $f_{\min}$ at point X is:

$$S^*_{f_{\min},\beta}(X) = \max_{k=0,1,\ldots,n} [\exp(-k\beta) \cdot \max(x_{k+1}, x_{k+2} - x_1)] \quad , \tag{3}$$

where $x_k = \Lambda$ for $k \ge n$. Similarly, for $X = \{x_1, \ldots, x_n\}$, where $\Lambda \ge x_1 \ge \cdots \ge x_n \ge 0$, the β-smooth sensitivity of $f_{\max}$ at point X is:

$$S^*_{f_{\max},\beta}(X) = \max_{k=0,1,\ldots,n} [\exp(-k\beta) \cdot \max(\Lambda - x_{k+1}, x_{k+2} - x_1)] \quad , \tag{4}$$

where $x_k = 0$ for $k \ge n$.

Given the β-smooth sensitivity of a function, it is possible to calibrate the noise to obtain a (ϵ, δ)-differentially private output. The following theorem follows from [13]:

THEOREM 4.1 ([13]). *Given ϵ and δ, set $\alpha = \epsilon/2$ and $\beta = \frac{\epsilon}{2} \cdot \ln(\frac{1}{\delta})$. Then the computation:*

$$M(X) = f(X) + \text{Laplace}\left(\frac{S^*_{f,\beta}(X)}{\alpha}\right) \tag{5}$$

maintains (ϵ, δ)-differential privacy.

EXAMPLE 4.2. *Assume that six cars are stuck in a traffic jam in a road segment, where the speed limit is 90 km/h. Speeds in the FCD database are in the range [0,120]. The cars report the speeds $\{3, 6, 10, 13, 16, 17\}$. Evaluating minimum speed with the Laplace mechanism for 1-differential privacy, would require computing $\min'(X) = 3 + \text{Laplace}(120)$. In contrast, relaxing the privacy requirement with $\delta = 0.01$, for $(1, 0.01)$-differential privacy we set $\alpha = 0.5$ and $\beta = 2.3$. According to Eq. 3, $S^*_{f_{\min},2.3} = 3$, hence $\min'(X) = 3 + \text{Laplace}(6)$ would still convey that the speed of the slowest car is much lower than expected.*

4.3 Calibrating Privacy Parameters

In this section, we address the calibration of the differential privacy parameters and tracking of privacy loss.

4.3.1 Factors in Parameter Calibration

When a query is executed against the FCD repository, the PCM is required to enforce the privacy policies stated for the affected records. In this process, the following factors should be considered.

Per-application accuracy requirements: ITS applications typically have defined accuracy standards for reporting of selected values. E.g., the Data Quality White Paper [1] published by the U.S. Department of Transport defines the required accuracy of speed reporting for traveller information applications to be in the range of 5-20%. The application requirements represent an upper bound on the variance of the noise introduced by the privacy mechanism for each query, and consequently a lower bound to acceptable values for ϵ and δ.

User-driven privacy settings: The privacy policy attached to each FCD record implies an upper bound on the privacy loss that could be incurred due to participation in queries and correspondingly on the acceptable values for ϵ and δ. As privacy requirements are subjective, acceptable levels of privacy may vary between users. Moreover, future ITS regulations could mandate the default values applicable to all users and all uses of FCD data, e.g., within a specific geographical region.

Affected records: In many functions, the amount of Laplace noise depends only on the privacy parameters, and is not affected by the number of records in the database. Consequently, the relative error may vary depending on the number of queried records. Therefore, to guarantee the required level of data accuracy, the PCM should first verify that enough records participate in the query. In scenarios where a limited number of FCD records are available and / or a lot of queries are issued by applications, there are a number of possible strategies to avoid service disruption due to unavailability of relevant records. These include adapting ϵ to the number of records and based on accuracy demands [16]. In Section 4.3.3 we describe a different approach based on sampling, which is suitable for evaluating average queries.

4.3.2 Managing FCD Lifetimes

The FCD record is the elementary piece of information to which a privacy policy is attached. As noted in Section 2.2, the differential privacy parameter ϵ is composable. If an FCD record participates in a series of queries, where each query q_i is ϵ_i-differentially private, then the overall privacy loss for the FCD record is constrained by $\sum \epsilon_i$. While accuracy requirements imply the acceptable value for ϵ_i in a single query q_i, user-driven privacy settings set a limit on the overall privacy loss $\epsilon = \sum \epsilon_i$ over a period of time.

We assume that FCD records are generated at a constant rate for all vehicles, as is the case with today's systems [1], and that queries are issued at random intervals. We further assume that there is only a limited number of queries during an update interval. To maintain differential privacy for any FCD record in this setting, we rely on two *FCD retention* parameters: *privacy budget* and *expiration time*.

Privacy budget: monitoring a privacy budget is an easy way to ensure that differential privacy requirements are maintained, and was used in frameworks such as PINQ [12] and PDDP [2]. In our architecture, the DP-PCM monitors the privacy budget at the FCD level. Each FCD j has a privacy budget b_j, initially set in the privacy policy attached to the record. For each query q_i, which incurs a privacy loss of at most ϵ_i, the FCD record would participate in the query only if $\epsilon_i \leq b_j$, and consequently the budget will be updated to $b_j \leftarrow (b_j - \epsilon_i)$. If the privacy budget of an FCD record reaches 0, it is removed from the repository.

Expiration time: the privacy policy attached to the FCD record can also state an expiration time, after which the FCD is removed from the repository. Since each vehicle generates new FCD records at a constant rate, the expiration time is critical to ensure that only a limited number of FCD records that originated from the same vehicle reside in the repository at the same time. We will discuss the impact of expiration time on user-level privacy in section 4.3.4.

4.3.3 Example: Evaluating Traffic Conditions

To demonstrate how the PCM can address the accuracy requirements of an ITS application while maintaining privacy constraints, we focus on Scenario 1.

Consider a route guidance application that queries FCD records to determine the average speed on a stretch of road, and accepts a 10 % deviation in the resulting speed. The PCM can use the Laplace mechanism as described in Section 2.2.1, adding Laplace noise to the result up to the acceptable inaccuracy. In addition, the application could also specify a minimum set size for a query. E.g., FCD records from at least $n = 50$ vehicles on a 1 km road segment would be sufficient to represent the average speed in an accurate way. Then, the PCM can verify before executing the query that enough records are available to answer the query.

Evaluating the number of records: Given a positive number α, sampling a Laplace distribution with scale b would return a number $-\alpha$ or lower with probability at most $0.5\exp(-\alpha/b)$ (one-sided error). Therefore, to verify that the number of FCD records in a differentially-private count query is at least n, we can set a safety margin α_c, and set $\epsilon_c = \frac{1}{\alpha_c}\ln\frac{1}{2\zeta}$. With probability at least $1-\zeta$, if the noisy count returns a number greater than $n + \alpha_c$, then there are at least n records in the dataset.

EXAMPLE 4.3. *Assume a safety margin of $\alpha_c = 10$, and set $\zeta = 0.05$. Then, executing a differentially private query with $\epsilon_c = \frac{1}{\alpha_c}\ln\frac{1}{2\zeta} = 0.23$, and obtaining a result of 60 or greater, guarantees with probability at least 0.95 that there are at least 50 FCD records in the database. If any smaller number of records is returned, we abort the query evaluation.*

Executing the *average* query: Once the PCM verifies that there are enough records in the dataset, the actual query can be issued, based on a sample of records with the required size[4]. With probability at most ζ, the two-sided error induced by the Laplace noise with scale b is bounded by $b\ln\frac{1}{\zeta}$. Therefore, the accuracy requirement and the records number bound can be used to derive a bound on the ϵ_s used to evaluate the average speed.

EXAMPLE 4.4. *Assume that there are more than 50 FCD records in the repository, and we would like to evaluate the average speed within 10 % deviation based on a sample of 50 records, where each record holds a value in the range $[0, 120]$. A differentially-private sum query would require Laplace noise of scale $120/\epsilon_s$, and over 50 records, the magnitude of noise added to the sum query should be at most 500. Therefore the PCM should set $\epsilon_s = \frac{120}{500}\ln\frac{1}{\zeta}$. For example, to ensure the bounded deviation with probability 0.95, ϵ_s should be set to at least 0.72.*

Algorithm 1 summarizes the process. For the count evaluation, we take a safety margin α that amounts to 10% of the minimum required record-set size, and the same probability bound ζ as the one used for speed accuracy, but any other reasonable values could be used instead.

4.3.4 Implications for User-Level Privacy Loss

User level privacy, as discussed in Section 2.2, is in general difficult to guarantee when many records are associated with each user, due to the level of noise that would be required in the differentially private functions. However, possible privacy threats can be considered when determining the privacy budget for each FCD record.

As an example, in line with Scenario 2, assume that the police tries to use the system to track down reckless drivers who consistently drive 20 km/h over the speed limit, and

[4]In the low-probability case where the noisy evaluation determines there are enough records although their number is below the limit, the query can either be executed on the smaller set, or dummy records with random values can be generated to reach the limit. In either case accuracy will suffer, but privacy would still be maintained.

Algorithm 1: AverageSpeed(P, Λ, n, α_s, ζ)

Input :
P – a road segment for which the average speed should be evaluated,
Λ – upper bound for speed values,
n – lower bound on number of vehicles to aggregate,
α_s – accuracy bound for speed,
ζ – probability bound for accuracy.

1: $\alpha_c = 0.1n$; $\epsilon_c = \frac{1}{\alpha_c} \ln \frac{1}{2\zeta}$. ; $\epsilon_s = \frac{1}{\alpha_s} \ln \frac{1}{\zeta}$.
2: Let RS be the set of all FCD records (one record per vehicle) reported in road segment P, such that for each record r_i with privacy budget b_i, we have $b_i \geq \epsilon_c + \epsilon_s$.
3: $count \leftarrow |RS| + \text{Laplace}(1/\epsilon_c)$.
4: $\forall i \in RS$: $b_i \leftarrow b_i - \epsilon_c$.
5: **if** $count \leq n + \alpha_c$ **then** abort query.
6: Let RS_n be a sample of n records from RS.
7: $avg \leftarrow (\texttt{SumSpeed}(RS_n) + \text{Laplace}(\Lambda/\epsilon_S))\,/n$.
8: $\forall i \in RS_n$: $b_i \leftarrow b_i - \epsilon_s$.
9: **return** avg.

that 2 % of the drivers fall into this category[5]. By querying the system the police aims to conclude that a certain driver is reckless with probability 0.99. From a user u's perspective, it may be desirable to stay "below the radar." Denoting the predicate "u is a reckless driver" with R_u, in differential privacy terms, this could be formulated as follows:

$$\Pr(R_u|DB \cup \texttt{FCD}_u) \leq \Pr(R_u|DB) \cdot \exp(\epsilon) \quad . \qquad (6)$$

For any series of queries that maintains ϵ-differential privacy with $\epsilon \leq \ln \frac{\Pr(R_u|DB \cup \texttt{FCD}_u)}{\Pr(R_u|DB)} \approx 3.9$, the user can avoid being detected by the police.

With respect to this benchmark, it is now possible to interpret the implications of the privacy parameters in terms of the susceptibility of the user to such inferences. For example, if the ϵ per query is 0.01, a new FCD record is generated every 5 minutes and deleted after 5 minutes (so at any time there is only one FCD record in the database per vehicle), an average driving time of one hour each day means that the police would need to monitor the FCD database for more than a month ($\frac{3.9}{0.01 \cdot 12} = 32.5$ days) before it can infer that a certain driver is reckless with high level of confidence. However, the interpretation of the privacy settings in terms of "monitoring period prior to breach" should serve only as a way to roughly judge the implications of different privacy settings in a very restricted scenario, and should not be assumed to reflect a privacy guarantee for a concrete user.

5. CONCLUSION AND FUTURE WORK

In this paper, we have discussed the application of differential privacy to the field of Intelligent Transportation Systems, especially considering the protection of Floating Car Data. As we have shown, event-level differential privacy can be integrated into a policy-enforcement framework like PRECIOSA PeRA in a straightforward way. We have illustrated how policies could be extended by expiration time and privacy budget parameters to specify and enforce a certain level of differential privacy. Implementing user-level privacy is more challenging and may involve limits to how much data can be stored about any specific vehicle at any time.

[5]According to a report from the U.S. Department of Transport (`http://www.nhtsa.gov/staticfiles/nti/pdf/811647.pdf`), on limited access highways in the U.S., 20% of drivers exceed the speed limit by more than 10 mph. Although we are not aware of numbers reflecting consistent severe speeding, for the sake of the example we believe our assumptions to be reasonable.

6. REFERENCES

[1] Ahn, K., Rakha, H., and Hill, D. Data quality white paper. Tech. Rep. FHWA-HOP-08-038, U.S. Department of Transportation, Federal Highway Administration, June 2008. Accessed on August 2012.

[2] Chen, R., Reznichenko, A., Francis, P., and Gehrke, J. Towards statistical queries over distributed private user data. In *NSDI* (2012).

[3] Danezis, G., Kohlweiss, M., and Rial, A. Differentially private billing with rebates. In *Information Hiding* (2011), pp. 148–162.

[4] Dietzel, S., Kost, M., Schaub, F., and Kargl, F. CANE: A Controlled Application Environment for Privacy Protection in ITS. In *ITST* (2012).

[5] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT* (2006), pp. 486–503.

[6] Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *TCC* (2006), pp. 265–284.

[7] Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. Differential privacy under continual observation. In *STOC* (2010), pp. 715–724.

[8] Dwork, C., Naor, M., Pitassi, T., Rothblum, G. N., and Yekhanin, S. Pan-private streaming algorithms. In *ICS* (2010), pp. 66–80.

[9] Gruteser, M., and Grunwald, D. Anonymous usage of location-based services through spatial and temporal cloaking. In *MobiSys* (2003), USENIX.

[10] Kargl, F., Dietzel, S., Schaub, F., and Freytag, J.-C. Enforcing privacy policies in cooperative intelligent transportation systems. In *Mobicom 2009 (Poster Session)* (September 2009).

[11] Kargl, F., Schaub, F., and Dietzel, S. Mandatory enforcement of privacy policies using trusted computing principles. In *Privacy 2010* (March 2010).

[12] McSherry, F. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. *Commun. ACM 53*, 9 (2010), 89–97.

[13] Nissim, K., Raskhodnikova, S., and Smith, A. Smooth sensitivity and sampling in private data analysis. In *STOC* (2007), pp. 75–84.

[14] Troncoso, C., Danezis, G., Kosta, E., Balasch, J., and Preneel, B. PriPAYD: Privacy-friendly pay-as-you-drive insurance. *IEEE Trans. Dependable Sec. Comput. 8*, 5 (2011), 742–755.

[15] Wiedersheim, B., Kargl, F., Ma, Z., and Papadimitratos, P. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. In *WONS* (February 2010).

[16] Xiao, X., Bender, G., Hay, M., and Gehrke, J. iReduct: differential privacy with reduced relative errors. In *SIGMOD Conference* (2011), pp. 229–240.