







A Differentially Private Hybrid Approach to Traffic Monitoring

Rogério V. M. Rocha¹(✉) , Pedro P. Libório¹ , Harsh Kupwade Patil² ,
and Diego F. Aranha^{1,3} 

¹ Institute of Computing, University of Campinas, Campinas, Brazil
rogerio.rocha@ic.unicamp.br, liborio@lrc.ic.unicamp.br

² LG Electronics, Santa Clara, USA
harsh.patil@lge.com

³ Department of Computer Science, Aarhus University, Aarhus, Denmark
dfaranha@cs.au.dk

Abstract. In recent years, privacy research has been gaining ground in vehicular communication technologies. Collecting data from connected vehicles presents a range of opportunities for industry and government to perform data analytics. Although many researchers have explored some privacy solutions for vehicular communications, the conditions to deploy them are still maturing, especially when it comes to privacy for sensitive data aggregation analysis. In this work, we propose a hybrid solution combining the original differential privacy framework with an instance-based additive noise technique. The results show that for typical instances we obtain a significant reduction in outliers. As far as we know, our paper is the first detailed experimental evaluation of differentially private techniques applied to traffic monitoring. The validation of the proposed solution was performed through extensive simulations in typical traffic scenarios using real data.

Keywords: Differential privacy · Smooth sensitivity · Hybrid approach · Intelligent Transportation Systems (ITS)

1 Introduction

Mobility is a major concern in any city, and deploying Intelligent Transportation Systems (ITS) can make cities more efficient by minimizing traffic problems [1]. The adoption of ITS is widely accepted in many countries today. Because of its high potential, ITS has become a multidisciplinary field of connective work and therefore many organizations around the world have developed solutions to provide ITS applications to meet growing demand [2].

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and LG Electronics via Unicamp Development Foundation (FUNCAMP) Agreement 5296.

© The Author(s) 2021

K. Sako and N. O. Tippenhauer (Eds.): ACNS 2021, LNCS 12727, pp. 233–256, 2021.

https://doi.org/10.1007/978-3-030-78375-4_10

Data collection in connected vehicles presents numerous opportunities through aggregated data analysis for companies, industries, and governments. Among these opportunities, one can highlight investigation of the driver behavior, which helps vehicle manufacturers and insurers to improve and develop new services. Another interesting application is the monitoring of traffic conditions which allows transport departments to manage mobility and improve services [5].

Regarding traffic management, it is increasingly important to understand the behavior of urban mobility. It includes presenting the travel profile of drivers for future mobility planning and testing in new scenarios. A Traffic Data Center (TDC) is a vital component in the mobility management. All collected data is processed and analyzed by a TDC in order to manage traffic in real-time or simply store it for additional operations [4]. A vehicle periodically sends *beacons* collected by sensors to its neighbors, including base stations, which are then sent directly to a TDC. The vehicle sensors collect data such as identification, timestamp, position, speed (direction), acceleration, among other about 7700 signals, some of which are treated as sensitive [4,5].

It is undeniable that analyzing this volume of data brings substantial social benefits, but also concerns about data breaches and leakage. Disclosure of this data poses a serious threat to the privacy of contributors, and creates a liability for industry and governments. In Europe, the General Data Protection Regulation (GDPR) imposes stricter rules on the storage and management of personally identifiable information, with non-compliance resulting in severe penalties [5].

To put it in context, it is worth mentioning that any type of monitoring can lead to a privacy breach through tracking. The main privacy concerns for drivers are disclosure, vehicle tracking and commercial use of personal data [5]. The speed, object of study in this paper, is a vector quantity which has a module (numerical value) and direction. In this way, the speed is considered as confidential data, as it is possible to deduce the driver's absolute value on a specific time and, more importantly, what is the driver's direction at that time and place.

In recent years, a strong mathematical definition of privacy in the context of statistical databases became increasingly accepted as a standard privacy notion. The original differential privacy framework was introduced by Dwork et al. in 2006 [3]. Since then, there was a lot of progress, including the sample and aggregate framework developed by Nissim et al. [9]. Based on this framework, our main research question is *how to preserve the privacy of drivers while providing accurate aggregated information to a TDC, such as the average speed?*

This paper addresses the problem of calculating the average speed in a road segment under a differentially private solution while maintaining the utility of aggregated data. Our main contributions are the following:

- We propose a hybrid approach exploring the characteristics of the original differential privacy [3] and the sample and aggregate frameworks [9].
- We present a formal proof showing that the proposed approach satisfies the differential privacy definition.
- We validate the hybrid approach through extensive empirical evaluation in some typical traffic scenarios, focusing on accuracy of the average speed.

1.1 Related Work

In recent years, researchers have explored numerous solutions to the problem of preserving privacy in the context of ITS. Pseudonym change strategies are the main local privacy-preserving solutions found in the literature, where contributors do not trust service providers. However, due to the precise space-time information contained in beacons, these strategies are still vulnerable to tracing, even supposedly anonymous [6]. In addition, due to safety applications, which require availability and accurate information, the design of alternative local privacy-preserving solutions is very restricted.

Regardless of local privacy-preserving solutions, our purpose is to focus on centralized solutions for data aggregation analysis, where the database is held by a trusted party. In this direction, the main contribution is due to Kargl et al. [4] in 2013, which investigated how differential privacy can be applied to ITS. Specifically, they propose an architecture that enables differential privacy when using beacons for some ITS applications and services. This architecture integrates a differentially private module through an extension of the PRECIOSA PeRA policy enforcement framework. To illustrate the functioning of the proposed module and how it addresses the accuracy and privacy requirements, Kargl et al. designed a simple algorithm for average speed calculation, based on the original framework of differential privacy.

A comprehensive survey on introduction of differential privacy in the automotive domain is presented by Nelson and Olovsson [5], where they claim that one of the main problems to introduce differential privacy in the automotive domain is maintaining high utility for the analyses. Another important work in this direction is due to Hassan et al. [7]. They survey differential privacy techniques and their application to cyber-physical systems, including ITS, as basis for the development of modern differential privacy techniques to address various problems and data privacy scenarios. Both works claim that the most prominent study relating differential privacy and vehicular domain is due to Kargl et al. [4].

Regarding data, most of collected signals by vehicle sensors are numeric and, specially in traffic monitoring, the aggregation functions sum, count and average capture many of calculations utilized in ITS applications [4]. These aggregation functions tend to have high distortion for small databases, mainly, for the sum and average due to variable global sensitivity that may not be diluted at a small database [5]. ITS applications typically have defined accuracy standards for reported values. For example, U.S. standardization determines that the distortion (error) presented at the reported average speed should be up to 20%, depending on the application [4]. It represents an upper bound on the noise introduced by a differentially private mechanism.

Given these surveys, our aim is to explore the peculiarities of the addressed problem and associate them to the characteristics of differentially private techniques, in order to obtain more accurate results while maintaining the same level of privacy. Although in most situations the instances are misbehaved, our hypothesis is that well-behaved instances are produced in some situations. This is due to the fact that the addressed problem is dynamic. The main difference

compared to [4] is that while they focus on a differentially private architecture applied to ITS, this article aims to deepen in this architecture by proposing a robust and effective differentially private algorithm to calculate the average speed in a realistic scenario that meets privacy and accuracy requirements.

The remainder of this paper is organized as follows. In Sect. 2, we present the theoretical foundations related to the differential privacy required to build our approach. Section 3 describes the proposed solution. After that, the experimental evaluation is presented in Sect. 4. Finally, we conclude and give direction to future work in Sect. 5.

2 Background

Differential privacy emerged from the problem of performing statistical studies on a population while maintaining the privacy of its individuals. The definition models the risk of disclosing data from any individual belonging to a database by performing statistical analyses on it.

Definition 1. DIFFERENTIAL PRIVACY [3]. *A randomized algorithm A taking inputs from the domain D^n gives (ϵ, δ) -differential-private analysis if for all data sets $D_1, D_2 \in D^n$ differing on at most one element, and all $U \subseteq \text{Range}(A)$, denoting the set of all possible outputs of A ,*

$$\left| \ln \left\{ \frac{\Pr[A(D_1) \in U] - \delta}{\Pr[A(D_2) \in U]} \right\} \right| \leq \epsilon \quad (1)$$

where the probability space is over the coin flips of the mechanism A and $\frac{p}{0}$ is defined as 1 for all $p \in \mathbb{R}$.

The parameters ϵ and δ , known respectively as *privacy loss parameter* and *relaxation parameter*, control the level of privacy and, consequently, the level of utility in the model. While ϵ determines the level of indistinguishability between the two databases, δ allows negligible leakage of information from individuals under analysis.

The protection of the individual's privacy in a database is done by adding carefully-crafted noise to the individual contribution or the aggregated data. In this way, it is sufficient to mask the maximum possible contribution (upper bound) in the database, which is the maximum difference between the analyses performed over two databases differing only in one element. This difference is known as global sensitivity, denoted by Δ_f .

One of the main models of computation is the centralized model (also known as output perturbation). In this model, there is a trusted party that has access to the raw individuals' data and uses it to release noisy aggregate analyses. The Laplace and exponential [8, 11] mechanisms are two of the main primitives in the differential privacy framework used to perturb the output analysis. The first, is the most widely used mechanism and it is based on sampling continuous random variables from Laplace distribution. In order to sample a random variable, one

should calibrate the Laplace distribution by centering the location parameter at either zero or the aggregated value and setting the scale parameter as the ratio between Δ_f and ϵ .

On the other hand, the exponential mechanism is used to handle both numerical and categorical analysis [8, 16]. This mechanism outputs an element $o \in O$ with probability $\propto e^{\left(\frac{\epsilon q(D,o)}{2\Delta_q}\right)}$, where O is a set of all possible outputs and Δ_q is the sensitivity of the quality function.

McSherry and Talwar [16] observed that the Laplace mechanism can be viewed as a special case of the exponential mechanism, by using the quality function as $q(D, o) = -|f(D) - o|$, which provides $\Delta_q = \Delta_f$. In this way, we can use the continuous exponential distribution and it is sufficient to assume $q(D, o) = -[f(D) - o]$, whereas the output o can be set as zero, which gives the true value of the analysis. Li et al. [8] proves that if a quality function is monotonic we can omit the constant two in the exponential mechanism.

Regarding the composability, the composition theorems are essential to design differentially private solutions. It allows to combine multiple mechanisms or perform multiple analyses over the same database by controlling the privacy and relaxation parameters, that is, the privacy budget. The sequential and parallel composition theorems are the main ones present in the literature.

In sequential composition, the parameters will be accumulated according to the number of performed analyses. On the other hand, in parallel composition, the resulting differentially private analysis will take into account only the maximum values of the parameters.

In the original differential privacy framework [3], the noise magnitude depends on the global sensitivity (Δ_f) but not on the instance D . For many functions, such as the median, this framework yields high noise compromising the utility of the analysis. The smooth sensitivity framework [9] allows to add significantly less noise than calibration with global sensitivity.

The smooth sensitivity is the smallest upper bound on the local sensitivity (LS), which is a local measure of sensitivity, and takes into account only the two instances involved in the analysis [9]. Nissim et al. proved that adding noise proportional to this upper bound is safe.

Definition 2. SMOOTH SENSITIVITY [9]. *For $\beta > 0$, the β -smooth sensitivity of f is:*

$$S_{f,\beta}^*(D_1) = \max_{k=0,\dots,n} e^{-k\beta} \left(\max_{D_2:d(D_1,D_2)=k} LS_f(D_2) \right). \tag{2}$$

The following definition states that if a probability distribution that does not change too much under translation and dilation it can be used to add noise proportional to $S_{f,\beta}^*$.

Definition 3. ADMISSIBLE NOISE DISTRIBUTION [9]. *A probability distribution $h \in \mathbb{R}$ is (α, β) -admissible for $\alpha(\epsilon, \delta)$ and $\beta(\epsilon, \delta)$ if it satisfies the following inequalities:*

$$\left| \ln \left[\frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U + \Delta)} \right] \right| \leq \epsilon/2 \tag{3}$$

$$\left| \ln \left[\frac{\Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{\Pr_{X \sim h}(X \in U \cdot e^\lambda)} \right] \right| \leq \epsilon/2 \tag{4}$$

for all $\|\Delta\| \leq \alpha$, $|\lambda| \leq \beta$ and all subsets $U \subseteq \mathbb{R}$.

The following lemma arises from Definitions 2 and 3.

Lemma 1. [9]. *The Laplace distribution on \mathbb{R} with scale parameter b is (α, β) -admissible with $\alpha = b\frac{\epsilon}{2}$ and $\beta = \frac{\epsilon}{2ln(1/\delta)}$.*

Proof. The proof can be found in the Appendix A. □

Claim. [9]. In order to get an (ϵ, δ) -differentially-private algorithm, one can add noise proportional to $\frac{S_{f,\beta}^*(D)}{\alpha}$.

Let a database $D = \{d_1, \dots, d_n\}$ in non-decreasing order and $f_{med} = median(D)$ where $d_i \in \mathbb{R}$, with $d_i = 0$ for $i \leq 0$ and $d_i = \Delta_f$ for $i > n$. Nissim et al. [9] proved that the β -smooth sensitivity of *Median* function is

$$S_{f,\beta}^*(D) = \max_{k=0,\dots,n} \left[e^{-k\beta} \max_{t=0,\dots,k+1} (d_{m+t} - d_{m+t-k-1}) \right], \tag{5}$$

where m is the rank of median element and $m = \frac{n+1}{2}$ for odd n . It can be computed in time $O(n^2)$.

The intuition behind the sample and aggregate framework [9] is to replace an aggregate function f by f^* , a smoothed and efficient version of it. This framework evaluates f over random partitions of the original database and releases f^* over the results by calibrating the noise magnitude with smooth sensitivity.

In this work, we deal with an unbounded stream of events as a database. An event may be an interaction between a particular person and an arbitrary term [10]. In this way, we focus on event-level privacy where the protection is centered on a single reported beacon. As the data set is dynamic, the attribute will change for each interaction making an event unique where its ID (identification) is the combination of timestamp and user ID.

3 Hybrid Approach

In this section, we describe the proposed approach to calculate the average speed on a road segment satisfying the definition of differential privacy. This approach combines the original differential privacy framework (ODP) [3] to the sample and aggregate framework (SAA) [9]. The adoption of the latter was inspired by the hypothesis that most speed values are close to the average when measured in a short time interval and road segment yielding some well-behaved instances. The hybrid approach is justified by the dynamism of the application, which yields misbehaved instances leading to very high sensitivity in the SAA framework.

The noise magnitude from the original and smooth sensitivity techniques are not related. While the differences among the instance and its neighbors are taken

into account to get the noise magnitude in the smooth sensitivity, the original technique considers only the global sensitivity without examining the instance itself. The core of our contribution is to propose a formulation relating these techniques in order to obtain the lowest noise magnitude, which results in more accurate analyses.

From now on, we will refer to the collected set of beacons as a prefix, a finite length chain from an unbounded stream of beacons. In our approach, we calculate the noisy prefix size by using the exponential mechanism, since we are not interested in negative values. To calculate the average speed, we use the Laplace mechanism in both ODP and SAA frameworks.

A trivial procedure to calculate the differentially private average function using the ODP framework is to add a random variable, sampled from the Laplace distribution, to the true sum function, then divide it by the set size N to obtain the average. In this case, the scale parameter is set as $\frac{\Delta_f}{\epsilon}$. The following algorithmic construction illustrates this procedure.

Algorithm 1: Trivial-ODP ($prefix, N, \Delta_f, \epsilon$)

```

1 # Calculate the scale of Laplace distribution
2  $b \leftarrow \frac{\Delta_f}{\epsilon}$ 
3 # Calculate sum from prefix
4  $sum \leftarrow 0$ 
5 for  $e \in prefix$  do
6   |  $sum \leftarrow sum + e_{speed}$ 
7 end
8 # Sample random variable from Laplace distribution
9  $Y_s \leftarrow Laplace(b)$ 
10 # Calculate noisy sum
11  $sum_{noisy} \leftarrow sum + Y_s$ 
12 # Calculate the noisy average speed
13  $avg_{noisy} \leftarrow \frac{sum_{noisy}}{N}$ 
14
15 return  $avg_{noisy}, b$ 

```

On the other hand, using the SAA framework, we can divide the prefix into random partitions and evaluate the average function over each partition. After this process, we must sort the resulting data set where we will select the central element (median) as the average speed. The main idea is to reduce the impact of anomalies present in the prefix when calculating the aggregation. It allow us to introduce less but significant noise to protect the maximum element in well-behaved instances. This procedure is presented in more details in Algorithm 2.

The Hybrid approach is based in the following lemma and theorem.

Lemma 2. *Let a prefix $P = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ be a set of points over \mathbb{R} , such that $x_i \in [0, \Delta_f]$ for all i . Sampling a random variable from the Laplace*

Algorithm 2: SAA (*prefix*, N , M , Δ_f , ϵ , δ)

```

1 # Partition prefix into  $M$  random samples of size  $N/M$ 
2 count  $\leftarrow 0$ 
3 average_speeds  $\leftarrow \emptyset$ 
4 while count  $< M$  do
5     # Extract the partition using a uniformly random sample
6     partition  $\leftarrow \text{RandomSample}(\text{prefix}, N/M)$ 
7     # Calculate average speed from partition adding to a list
8     avg  $\leftarrow \frac{\text{Sum}(\text{partition})}{N/M}$ 
9     average_speeds  $\leftarrow \text{Append}(\text{avg})$ 
10    count  $\leftarrow \text{count} + 1$ 
11 end
12 # Sort average speeds set in non-decreasing order
13 sorted_average_speeds  $\leftarrow \text{Sort}(\text{average\_speeds})$ 
14 # Calculate the scale of Laplace distribution
15  $b \leftarrow \frac{\Delta_f}{\epsilon}$ 
16 # Calculate alpha and beta parameters
17  $\alpha \leftarrow b \frac{\epsilon}{2}; \beta \leftarrow \frac{\epsilon}{2 \ln(1/\delta)}$ 
18 # Calculate smooth sensitivity of median function by Eq. (5)
19 smooth_sensitivity_median  $\leftarrow S_{\text{median}, \beta}^*(\text{sorted\_average\_speeds}, M, \Delta_f)$ 
20 # Get random variable from Laplace distribution
21  $Y_m \leftarrow \text{Laplace}\left(\frac{\text{smooth\_sensitivity\_median}}{\alpha}\right)$ 
22 # Calculate noisy average speed
23 avg_noisy  $\leftarrow \text{Median}(\text{sorted\_average\_speeds}) + Y_m$ 
24
25 return avg_noisy,  $\frac{\text{smooth\_sensitivity\_median}}{\alpha}$ 

```

distribution with scale parameter set as $\frac{\Delta_f/N}{\epsilon}$ and add it to the true average function is equivalent to Algorithm 1, both performed over P .

Proof. Consider the cumulative distribution function of the Laplace distribution with mean ($\mu = 0$) [17]. Suppose S is the sum of P and $r_s = \lambda \cdot S$ represents a proportion of S . The probability of sampling any value greater than r_s is given by

$$p_s(X > r_s) = \frac{1}{2} e^{-\frac{r_s}{b_s}} \quad (6)$$

where $b_s = \frac{\Delta_f}{\epsilon}$.

Now, suppose A is the average of P and $r_a = \lambda \cdot A$ represents a proportion of A . The probability of sampling any value greater than r_a is given by

$$p_a(X > r_a) = \frac{1}{2} e^{-\frac{r_a}{b_a}} \quad (7)$$

In order to conclude the proof, we need to determine b_a . So, it is a fact that $S = A \cdot N$. Thus, we have $r_s = \lambda \cdot A \cdot N$, which results in $r_s = r_a \cdot N$.

By substituting it in Eq. (6) and equating to Eq. (7), i.e., $p_s = p_a$, we obtain $b_a = \frac{\Delta_f/N}{\epsilon}$. \square

Based on Lemma 2, the following algorithmic construction is an alternative to Algorithm 1.

Algorithm 3: ODP (*prefix*, N , Δ_f , ϵ)

```

1 # Calculate the scale of Laplace distribution
2  $b \leftarrow \frac{\Delta_f/N}{\epsilon}$ 
3 # Calculate sum from prefix
4  $sum \leftarrow 0$ 
5 for  $e \in prefix$  do
6   |  $sum \leftarrow sum + e_{speed}$ 
7 end
8 # Calculate true average
9  $avg \leftarrow \frac{sum}{N}$ 
10 # Sample random variable from Laplace distribution
11  $Y_s \leftarrow Laplace(b)$ 
12 # Calculate the noisy average speed
13  $avg_{noisy} \leftarrow avg + Y_s$ 
14
15 return  $avg_{noisy}, b$ 

```

Theorem 1. *Let a prefix $P = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ be a set of points over \mathbb{R} , such that $x_i \in [0, \Delta_f]$ for all i . Then, Algorithm 2 provides more accurate results than Algorithm 3, if $S_{f_{median},\beta}^*(D) < \alpha \cdot \frac{\Delta_f/N}{\epsilon}$, both performed over P .*

Proof. Let b_{SAA} and b_{ODP} be the scale parameter of the Laplace distribution in Algorithms 2 and 3, respectively. Then, we obtain

$$b_{SAA} = \frac{S_{f_{median},\beta}^*(D)}{\alpha} \quad (8)$$

$$b_{ODP} = \frac{\Delta_f/N}{\epsilon} \quad (9)$$

Rearranging Eq. (8) and setting b_{ODP} as an upper bound on b_{SAA} , we get $S_{f_{median},\beta}^*(D) < \alpha \cdot b_{ODP}$, which results in

$$S_{f_{median},\beta}^*(D) < \alpha \cdot \frac{\Delta_f/N}{\epsilon}. \quad (10)$$

In order to prove this theorem, assume for the sake of contradiction that Algorithm 3 provides more accurate results than Algorithm 2, both performed over P . Then, b_{ODP} is less than b_{SAA} . By Eq. (10), it is a contradiction.

Therefore, if Eq. (10) holds, then Algorithm 2 provides more accurate results than Algorithm 3. \square

From Theorem 1 and Lemma 2, the noise magnitude of the Hybrid approach is formulated as follows:

$$b_{Hybrid} = \begin{cases} b_{SAA}, & \text{if } S_{f_{median},\beta}^*(D) < \alpha \cdot \frac{\Delta_f/N}{\epsilon} \\ b_{ODP}, & \text{otherwise.} \end{cases} \quad (11)$$

The algorithmic construction of the Hybrid approach is presented in Algorithm 4. This algorithm calculates the average speed in a differentially private way using all beacons reported in a short time interval in a specific road segment. It gets as input a privacy budget ϵ related to each received event in the base station, the prefix size N to calculate the average speed, the number of partitions for SAA framework, the global sensitivity of the average function (speed limit in the road segment), the privacy loss parameters for count and average functions, and the relaxation parameter for average function (non-zero).

The algorithm starts by checking the privacy budget of the privacy loss and relaxation parameters. After that, it initializes an empty list called *beacons* used to store all beacons received through the base station. Next, the base station starts collecting data (beacons/events) adding each of them to the list. The collection control is made by a differentially private *Count* function which uses the exponential mechanism, Algorithm 5. The event collection is performed by the *Receive Beacon* function. Each beacon includes the vehicle speed (*m/s*) between 0 and Δ_f . It is worth mentioning that, in a realistic scenario, some values can be above the speed limit Δ_f but these values are intentionally not protected in proportion to their magnitude, since in our scenario they are reckless drivers. After collecting enough data to compose the prefix, the algorithm selects the most recent beacons to calculate the average speed. The next step is to calculate the noisy average speed through the two frameworks, *ODP* and *SAA*. Then, we choose the average noisy speed calculated with the lowest noise magnitude. Finally, the privacy loss and relaxation parameters are deducted from the privacy budget for each event in the prefix.

3.1 Security Analysis

A Threat Model. Differential privacy was designed considering a very strong adversary, with an infinite computational power, who has the knowledge of the entire data set, except a single element. It is considered that the adversary cannot glean any additional information about this element beyond what it known before interacting with the privacy mechanism. This assumption is not unrealistic since differential privacy is supposed to provide privacy given adversaries with arbitrary background knowledge. Then, the adversary tries to obtain additional information about this element using the knowledge of the entire data set except it and the auxiliary information about it before the data set analysis.

In our scenario, for simplicity, consider that there are two service providers (carriers *A* and *B*) that provide aggregate information to customers (drivers), such as average speed on a road segment. Also, we consider that all drivers on a road segment are customers of both carriers except by a single customer *e* who is

a customer of only one of them, B , for example. As we are dealing with a strong adversary, it is supposed that they have knowledge about all others customers except by e . That is, the speed of all drivers which are customers of carrier A . Then, from the entire data set (the selected prefix by carrier A) which has length N , the adversary can obtain the sum of all speeds and calculate the difference between this sum and the result of the product between the average speed from B , which includes the driver e 's speed, multiplied by $N + 1$. This procedure gives the correct contribution of e .

Algorithm 4: Hybrid (ϵ , δ , N , M , Δ_f , ϵ_c , ϵ_a , δ_a)

```

1 if  $\epsilon_c + \epsilon_a \leq \epsilon$  and  $\delta_a \leq \delta$  then
2   # Initialize beacon list
3    $beacons \leftarrow \emptyset$ 
4   # Receive first event and add it to the beacon list
5    $e \leftarrow ReceiveBeacon()$ 
6    $beacons \leftarrow Append(e)$ 
7   # Receive the remaining events and add them to the beacon list
8   while  $Count(beacons, \epsilon_c) < N$  do
9      $e \leftarrow ReceiveBeacon()$ 
10     $beacons \leftarrow Append(e)$ 
11  end
12  # Select the  $N$  more recent events
13   $prefix \leftarrow SelectLatestBeacons(beacons, N)$ 
14  # Calculate the noisy average speed through ODP and SAA
15   $avg_{ODP}, b_{ODP} \leftarrow ODP(prefix, N, \Delta_f, \epsilon_a)$ 
16   $avg_{SAA}, b_{SAA} \leftarrow SAA(prefix, N, M, \Delta_f, \epsilon_a, \delta_a)$ 
17  # Choosing the lowest noise magnitude
18  if  $b_{SAA} < b_{ODP}$  then
19     $avg_{noisy} \leftarrow avg_{SAA}$ 
20  end
21  else
22     $avg_{noisy} \leftarrow avg_{ODP}$ 
23  end
24  # Deduce count and average privacy loss parameter from each event
   privacy budget in prefix
25  for  $e \in prefix$  do
26     $\epsilon \leftarrow \epsilon - \epsilon_c - \epsilon_a$ 
27     $\delta \leftarrow \delta - \delta_a$ 
28  end
29
30  return  $avg_{noisy}$ 
31 end
32 else
33   write "Privacy budget exceeded!"
34 end

```

Algorithm 5: Count (*beacons*, ϵ)

```

1 # Calculate the scale of Exponential distribution
2  $\lambda \leftarrow \epsilon$ 
3 # Calculate count from beacon list
4 count  $\leftarrow 0$ 
5 for  $e \in \textit{prefix}$  do
6   | count  $\leftarrow \textit{count} + 1$ 
7 end
8 # Get random variable from exponential distribution
9  $Y_c \leftarrow \textit{Exponential}(\lambda)$ 
10 # Calculate noisy count
11 countnoisy  $\leftarrow \textit{count} - Y_c$ 
12
13 return countnoisy

```

Privacy Analysis. The security of the Hybrid approach is supported by the following lemmas and theorem. In Lemma 3, we prove that the randomized *Count* function, presented in Algorithm 5, is differentially private. After that, Lemma 4 shows that Algorithm 3 satisfies differential privacy. Next, we prove through Lemma 5 by parallel composition that Algorithm 2 is differentially private. Finally, in Theorem 2, we prove that the Hybrid approach presented in Algorithm 4 satisfies differential privacy by sequential composition.

Lemma 3. *From the beacon list, let $B = \{x_1, x_2, \dots, x_{n-1}, x_n\}$ be a set of points over \mathbb{R} such that $x_i \in [0, \Delta_f]$ for all i and $|B|$ be the length of the beacon list. Then, Algorithm 5 satisfies $(\epsilon, 0)$ -differential privacy.*

Proof. Assume that, without loss of generality, A represents Algorithm 5. Let B_1 and B_2 be two neighboring beacon lists differing at most one event, i.e., $||B_1| - |B_2|| = 1$. From Eq. (1) in the differential privacy definition, we must evaluate two cases: when the ratio is greater than 1 and less or equal to 1. Since the quality of the *Count* function is monotonic [11], we get:

– When $\frac{\Pr[A(B_1) \in U]}{\Pr[A(B_2) \in U]} \geq 1$, we have

$$\frac{\Pr[A(B_1) \in U]}{\Pr[A(B_2) \in U]} = \frac{\epsilon \int_a^b e^{-\epsilon x} dx}{\epsilon \int_a^b e^{-\epsilon(x+1)} dx} = \frac{e^{-(\epsilon a)} - e^{-(\epsilon b)}}{\frac{\epsilon}{e^{-\epsilon} [e^{-(\epsilon a)} - e^{-(\epsilon b)}]}} \leq e^\epsilon. \quad (12)$$

– When $\frac{\Pr[A(B_1) \in U]}{\Pr[A(B_2) \in U]} < 1$, we have by symmetry that the ratio is $\geq e^{-\epsilon}$.

□

Lemma 4. *Let P be a prefix from a beacon list $B = \{x_1, \dots, x_n\}$ such that $N = |P| \leq |B|$ and $x_i \in [0, \Delta_f]$ for all i . Then, Algorithm 3 satisfies $(\epsilon, 0)$ -differential privacy.*

Proof. Assume now, without loss of generality, that A represents Algorithm 3. Let P_1 and P_2 be two neighboring prefixes differing at most one event, $|A(P_1) - A(P_2)| = \Delta_f/N$. From the definition of differential privacy, we obtain

- When $\frac{Pr[A(P_1) \in U]}{Pr[A(P_2) \in U]} \geq 1$, we have

$$\frac{Pr[A(P_1) \in U]}{Pr[A(P_2) \in U]} = \frac{\frac{\epsilon_s N}{2\Delta_f} \int_U e^{-\frac{\epsilon_s N|x|}{\Delta_f}} dx}{\frac{\epsilon_s N}{2\Delta_f} \int_U e^{-\frac{\epsilon_s N|x+\Delta_f/N|}{\Delta_f}} dx} = \frac{\int_a^b e^{-\frac{\epsilon_s N|x|}{\Delta_f}} dx}{\int_a^b e^{-\frac{\epsilon_s N|x+\Delta_f/N|}{\Delta_f}} dx}. \quad (13)$$

We will solve this ratio in two parts. First, considering numerator of Eq. (16), evaluating the cases when $x \geq 0$ and $x < 0$, we obtain respectively

$$\int_a^b e^{\mp \frac{\epsilon_s N x}{\Delta_f}} dx = \pm \frac{\Delta_f [e^{\mp(\epsilon_s a N)/\Delta_f} - e^{\mp(\epsilon_s b N)/\Delta_f}]}{\epsilon_s N}. \quad (14)$$

Now, considering denominator of Eq. (16), evaluating the cases when $x \geq -\Delta_f/N$ and $x < -\Delta_f/N$, we obtain respectively

$$\int_a^b e^{\mp \frac{\epsilon_s N(x+\Delta_f/N)}{\Delta_f}} dx = \pm \frac{e^{-\epsilon_s N} \Delta_f [e^{\mp(\epsilon_s a N)/\Delta_f} - e^{\mp(\epsilon_s b N)/\Delta_f}]}{\epsilon_s N}. \quad (15)$$

By replacing Eq. (14) and Eq. (15) in Eq. (16), we obtain

$$\frac{\pm \frac{\Delta_f [e^{\mp(\epsilon_s a N)/\Delta_f} - e^{\mp(\epsilon_s b N)/\Delta_f}]}{\epsilon_s N}}{\pm \frac{e^{-\epsilon_s N} \Delta_f [e^{\mp(\epsilon_s a N)/\Delta_f} - e^{\mp(\epsilon_s b N)/\Delta_f}]}{\epsilon_s N}} \leq e^{\epsilon_s}. \quad (16)$$

- When $\frac{Pr[A(P_1) \in U]}{Pr[A(P_2) \in U]} < 1$, we have by symmetry that the ratio is $\geq e^{-\epsilon_s}$.

□

Lemma 5. *Let P be a prefix from a beacon list $B = \{x_1, \dots, x_n\}$ such that $N = |P| \leq |B|$ and $x_i \in [0, \Delta_f]$ for all i . Then, Algorithm 2 is ϵ -differentially private with probability $1 - \delta$.*

Proof. Our construction is based on uniformly distributed samples from the prefix P . These random samples are extracted without replacement, producing M partitions of size N/M . The M partitions form a set from which we can calculate the average speed. In order to do it, we first need to sort this set of partitions in a non-decreasing order and then calculate the smooth sensitivity of *Median* function from Eq. (5). Thus, Algorithm 2 follows the sample and aggregate framework.

The proof of this lemma follows directly by combination of Definition 3, Lemma 1 and parallel composition theorem [8]. □

Theorem 2. *Let P be a prefix from a beacon list $B = \{x_1, \dots, x_n\}$ such that $N = |P| \leq |B|$ and $x_i \in [0, \Delta_f]$ for all i . Then, Algorithm 4 satisfies (ϵ, δ) -differential privacy.*

Proof. From Lemma 3, Lemma 4 and Lemma 5 we have that Algorithm 5, 3 and 2 are differentially private. By the sequential composition theorem [8], the combination of Algorithm 5, 3 or 2 occurs when $\epsilon_c + \epsilon_a \leq \epsilon$ and $\delta_a \leq \delta$ in Algorithm 4. Therefore, Algorithm 4 satisfies (ϵ, δ) -differential privacy. \square

4 Empirical Evaluation

In this section, we present and discuss the results obtained from the evaluation of the Hybrid approach for differentially private computation of average speed. Since the evaluation focuses on the accuracy of the proposed solution, the two fundamental parameters were fixed and calibrated as suggested in the literature [11]. In this evaluation, we set the privacy loss parameter as $\ln(2) - 0.15$ for *average* function and 0.15 for *count* function. Since we have defined the prefix size in this evaluation as 55, it is sufficient to calibrate the relaxation parameter with 0.01, which allows negligible leakage information in the size of the prefix, $o(1/N)$. For the SAA approach, we partition the prefix into 11 random partitions with 5 elements each.

In order to evaluate the approach, we adopted the open source traffic mobility (SUMO) [13] and the discrete event-based (OMNeT++) [15] simulators. In addition, as a interface of the two simulators, we use the open source framework for running vehicular network simulations (Veins) [14]. The evaluation was performed on a realistic mobility scenario provided by Codeca et al. in [12], using the SUMO simulator. The realistic mobility scenario is able to meet all basic requirements in size, realism and duration of a real-world city (Luxembourg) with a typical topology in mid-size European cities. From now on, we will refer to the realistic mobility scenario as the Luxembourg scenario. This scenario is available to industrial and scientific communities working on vehicular traffic congestion, intelligent transportation systems and mobility patterns.

As a utility metric, we adopt the absolute deviation and create filters on the reported original average speed. The values to calibrate the filters are in line with US standardization (in Subsect. 1.1). The scenario of evaluation and the numerical and graphical results are presented in following sections.

4.1 Luxembourg Scenario

As mentioned before, the realistic mobility scenario is based on the city of Luxembourg and contains residential or arterial roads and highways, see Fig. 1a. The Luxembourg scenario has an area of 156 km², 930 km of roads, 4, 500 intersections, 200 traffic lights, and 290,000 cars. This scenario works on two types of mobility traces, which have duration of about 88,000s (24 h), and peaks of traffic in about 8AM, 13PM and 6PM, as it can be seen in Fig. 1b.

DUA-T (Dynamic User Assigned Traces) is one of the mobility traces, which provides the optimal path for each origin-destination pair in terms of time and length. It is not very realistic because it does not take other vehicles and congestion into account. DUE-T (Dynamic User Equilibrium Traces) is the other

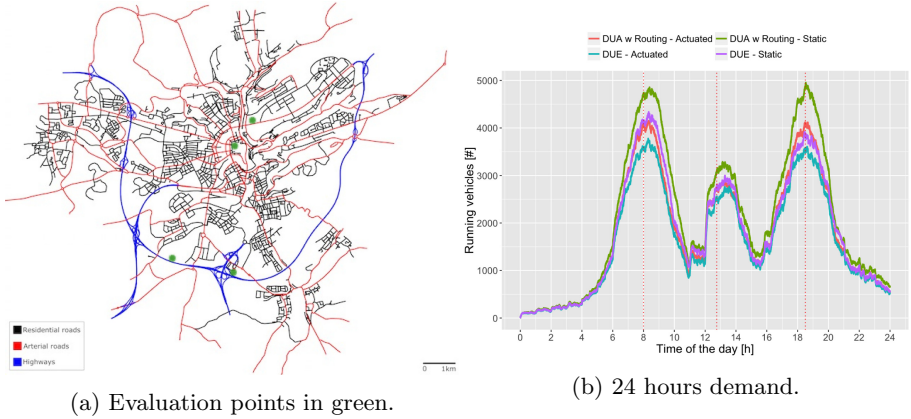


Fig. 1. Luxembourg scenario and traffic demand. **Source:** (a) <https://www.vehicularlab.uni.lu/lust-scenario/> (b) <https://github.com/lcodeca/LuSTScenario/tree/master> (Color figure online)

mobility trace that provides the approximated equilibrium for the scenario’s traffic demand [12]. The latter can be combined with static or actuated traffic light systems. The static case isolates the impact of routing, while the actuated case would imply two independent optimization problems, the traffic light timing and the vehicular rerouting [12]. The combination of DUE-T and actuated traffic lights seems more realistic. However, we opt by combination of DUA-T and static traffic lights because this setting cause more traffic congestion and it fits well with our problem.

We evaluate the Hybrid approach in four strategic points of the Luxembourg scenario (green points in Fig. 1) in a rush period, between about 6AM and 10AM, as it can be seen in Fig. 1b. The Road Side Units (RSU’s) or base stations were positioned using the Geodetic (Longitude/Latitude) coordinate system. The first and third points are located on a highway with low vehicle density. The first point has an RSU with a range of 250 m monitoring traffic on the road with no congestion. At the third point, we have a substantial traffic jam caused by a maintenance on the road, which has an RSU with a range of 150 m. The second and fourth points are located at the center of the city containing high vehicle density and traffic lights. RSU’s are monitoring arterial roads with a respective range of 75 and 42.5 m, all congested in different levels. The second point has a regular traffic flow, with very little jam caused by traffic lights and the last point has a lot of congestion because it is the main avenue in the city center where several streets lead to it. The next subsection summarizes our numerical results.

4.2 Experimental Results

The filters were created with deviation tolerances (tol) of 5, 10 and 20% over the reported original average speed ($avg_{original}$). The reported noisy average speed

(avg_{noisy}) is expected to remain within the respective range and any measurement reported outside this range is considered an outlier. Thus, the reported average noisy speed can be represented as

$$avg_{noisy} = avg_{original} \cdot (1 \pm tol) \tag{17}$$

where tol is divided by 100.

As numerical result, we calculate the number of outliers obtained in the simulation time window for each approach: ODP, SAA (all deviation tolerances) and Hybrid (deviation tolerance of 10%). In addition, we calculate the number of misbehaved (bad) instances, those that produce SAA scale parameters larger than the expected SAA scale parameter, and also the number of SAA scale parameters that are lower than the ODP scale parameter. The expected SAA scale parameter is calculated based on $Pr(-avg_{original} \cdot tol \leq X \leq avg_{original} \cdot tol) = 0.95$, where the random variable X is the noise to be added to the original average speed. Furthermore, in order to enrich our discussion, we present the following graphic results: the behavior of the real average speed, the quality of the instances by presenting the scale parameter for each instance and the relative deviation between the results of the hybrid approach and the original average speed. Table 1 summarizes the numerical results.

Table 1. Results of the Luxembourg scenario evaluation. The coordinates and speed limits (m/s) of the points correspond respectively to (49.579464, 6.100917), limit of 36.11; (49.617679, 6.132573), 13.89; (49.575654, 6.131255), 36.11; and (49.611492, 6.126152), 13.89.

Point	Number of events	Bad instances (%)	Lower b_{SAA} (%)	Outliers (%)								
				ODP			SAA			Hybrid		
				5%	10%	20%	5%	10%	20%	5%	10%	20%
1st	3,648	65.25	41.31	25.54	9.46	4.52	25.94	9.51	4.54	9.33	1.05	0.00
2nd	2,046	77.74	25.34	32.08	15.59	7.37	33.52	15.98	7.50	13.36	3.42	0.68
3rd	4,210	99.34	34.17	77.17	57.98	36.89	80.23	65.96	47.68	45.77	30.19	15.29
4th	5,068	99.70	62.23	96.01	87.24	84.78	90.53	77.13	75.13	87.89	72.37	70.35

We initiate our discussion by pointing out that the number of outliers is decreasing when it varies among the deviation tolerances from 5% to 20% in all points of evaluation. It is an expected behavior since we expand the tolerance range. Although we are getting an improvement in the number of outliers in all cases when moving among deviation tolerances, the rate of variation when switching among the evaluation points are decreasing. For example, this rate varies from about 8.88 in the 1st point (cell Hybrid 5% divided by cell Hybrid 10% in Table 1) until about 1.21 in the 4th point of evaluation (cell Hybrid 5% divided by cell Hybrid 10% in Table 1). This shows that the greater the congestion, the lower the rate of variation among the deviation tolerances.

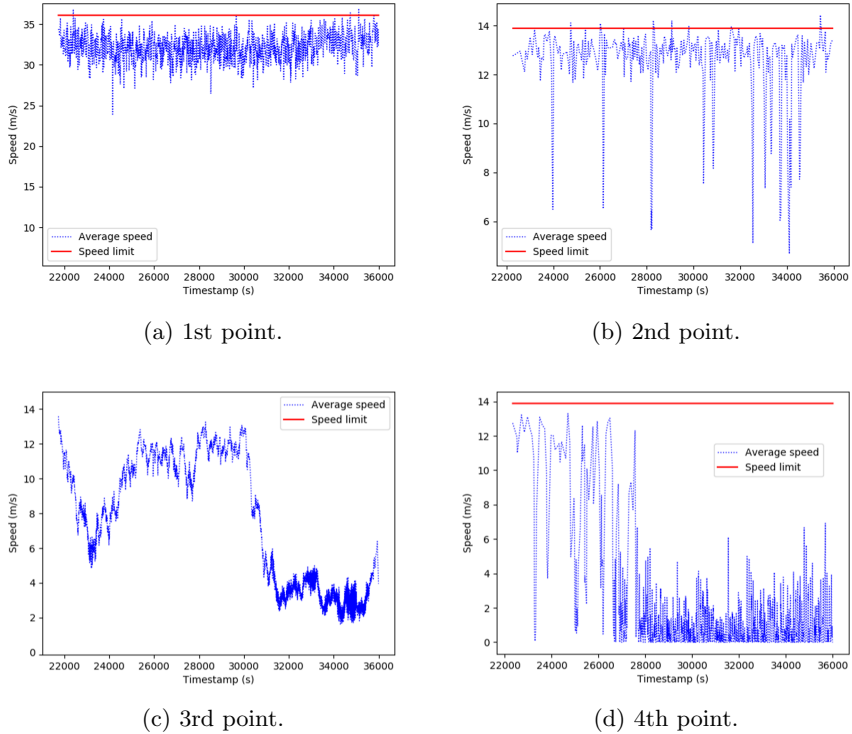


Fig. 2. Average and limit speed behavior during the simulation time window.

At the first and second point of evaluation, which are respectively located in a highway and an arterial road, we can see that ODP and SAA provide virtually the same result (number of outliers) for all deviation tolerance. For instance, it is about 9.5% and 16% in the 1st point and 2nd point, respectively, for a deviation tolerance of 10%. The good result at these points is due to the ideal flow both on the highway and on the arterial road, so that the ODP has the same behavior as the SAA. Observe in Fig. 2a and 2b that the behavior of the average speed is very close to the speed road limit. The results of the 2nd point are worse than the 1st due to the traffic lights present in the second point yielding a small traffic jam. Note that in the Fig. 2b there are measurements far below the speed limit.

Still in the 1st and 2nd evaluation points, the results related to the Hybrid approach show that we obtain a significant reduction in outliers. At the 1st point, in Table 1, the number of outliers is reduced from about 9.5% (ODP and SAA) to about 1% (Hybrid) for a deviation tolerance of 10%, a reduction rate of more than nine times. When we move to the deviation tolerance of 20% the number of outliers is reset to zero (Hybrid) from about 4.5% (ODP and SAA). Figure 3a shows the behavior of the relative deviation for all measurements and from it we can see that all deviation are below 20%. Note, in Fig. 4a, that even with more than 65% of badly behaved instances, we get about 41% of the SAA scale

parameters (yellow dots) below the ODP scale parameter (solid red line), these smaller obtained scale parameters is sufficient to obtain a significant reduction in the outliers. Observe further that in Fig. 4a, the expected SAA scale parameter (dashed green line) is slightly below the ODP scale parameter (solid red line) and most of the 41% of the SAA scale parameters (yellow dots) below the ODP scale parameter (solid red line) are also below the expected SAA scale parameter (dashed green line).

In Table 1, at the 2nd point, the reduction rate related to the Hybrid approach compared to ODP and SAA for the deviation tolerance of 10% is a bit lower than at the 1st point, around 4.5 (15.59 ODP or 15.98 SAA divided by 3.42 Hybrid), half of the 1st point but still a great result, especially when we consider the results for the deviation tolerance of 20% which reaches a reduction rate of more than 11 times. See in Fig. 3b that most deviation are below 15% which shows a good performance of the Hybrid approach. Although 78% of instances are misbehaved, more than 25% of all instances have SAA scale parameters (yellow dots) smaller than the ODP scale parameter (solid red line), see Fig. 4b, these smaller scale parameters are crucial to get this improvement.

The reason the Hybrid approach provides great results is because most outliers do not overlap between the ODP and SAA approaches.

Now, considering the 3rd evaluation point located in a highway, we can see that the results of all approaches suffered a huge negative impact caused by a substantial traffic jam. The number of outliers reached about 80% (ODP and SAA) and almost half of it with the Hybrid approach for deviation tolerance of 5%. When moving to deviation tolerance of 20% the result of Hybrid approach is less than a half of the ODP and SAA results. Figure 2c shows the behavior of the average speed in this point. Observe that all the measurements are too far from the speed limit (36.11). There are two declines in the average speed behavior, one at the beginning of the simulation reaching about 6 m/s and another after 30000s that reaches about 2 m/s. This is due to the high traffic demand at around 8AM where vehicles will abruptly reduce their speed when they are very close to road maintenance in order to avoid collisions, contributing to congestion.

Still in the 3rd evaluation point, the SAA result has a considerable worsening in relation to the ODP result, about 8% points in the deviation tolerance of 10% reaching until about 11% in the deviation tolerance of 20%. This is explained by the traffic jam yielding misbehaved instances which directly impacts the good performance of the SAA approach. From Table 1 we obtain about 99% of badly behaved instances, this lead to very little measurements below the expected scale parameter (dashed green line), see Fig. 4c. Even so, we get about 34% of the SAA scale parameters (yellow dots) below the ODP scale parameter (solid red line), sufficient to obtain a reduction rate of about 2 times in the number of outliers for the deviation tolerance of 5% and 10% with Hybrid approach compared to ODP and SAA, and reaching more than 3 times for the deviation tolerance of 20%.

Finally, in the 4th evaluation point, the growth in the number of outliers is even more evident when compared to the 3rd evaluation point, reaching about

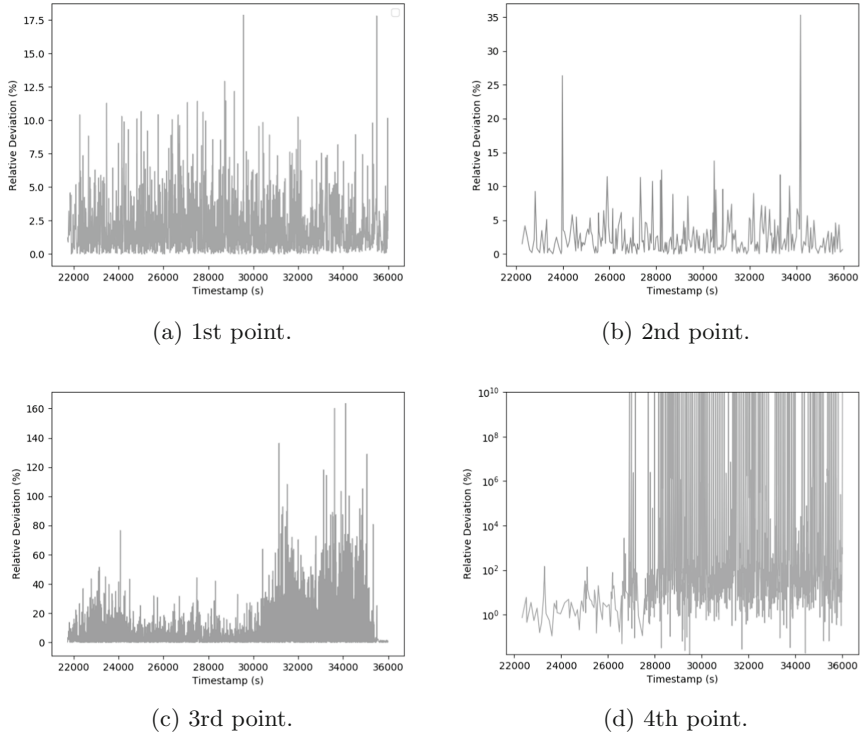


Fig. 3. Relative deviation between the hybrid approach and the original average speed for each instance during the simulation time window.

2.3 times more in the ODP approach and about 4.6 times in Hybrid approach for the deviation tolerance of 20%. This worsening occurs because most of SAA scale parameters when applied over an average speed very close to zero leads to an outlier. See, in Fig. 2d, that most measurements are close to zero. We can also see in Fig. 3d that the relative deviation is very high in most measurements in the simulation time window.

The SAA result improves considerably compared to the ODP result at the 4th valuation point, about 10% below in the deviation tolerance of 10%. Although almost all (99.7%) instances are misbehaved, close to 63% of the SAA scale parameters (yellow dots) are below the ODP scale parameter (solid red line) as it can be seen in Fig. 4d, which explains this improvement. However, it was not sufficient to help the Hybrid approach provide good results (significant reduction in outliers), this is due to the huge number of average speed very close to zero. We can conclude that all approaches are very sensitive to an average speed close to zero.

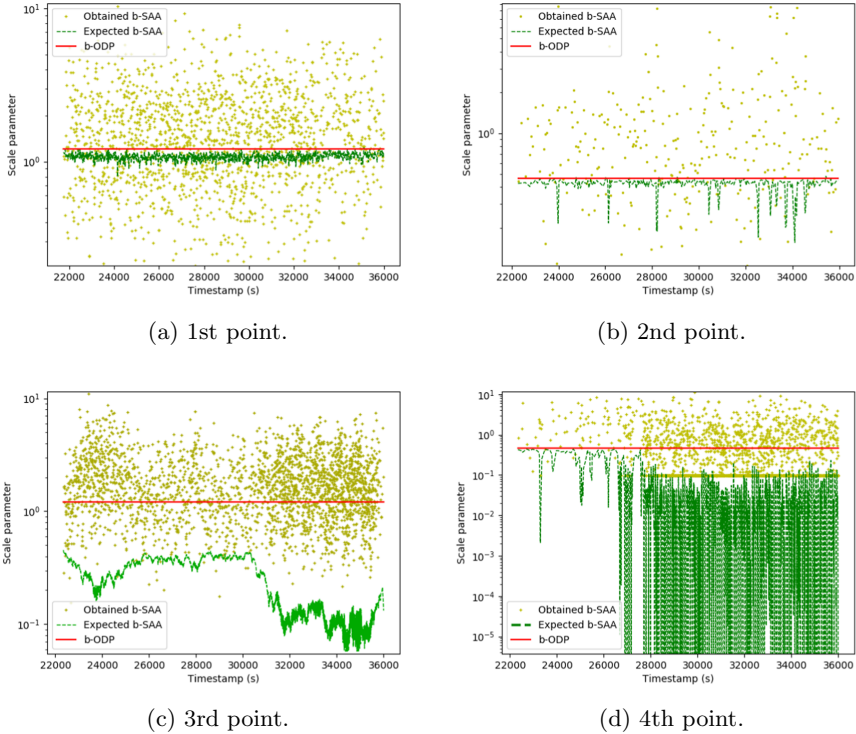


Fig. 4. Scale parameter for each instance during the simulation time window. (Color figure online)

5 Conclusion

We proposed in this paper a hybrid privacy-preserving data aggregation solution for traffic monitoring focusing on event-level privacy. This solution was designed to calculate the average speed on a road segment combining the original differential privacy to the sample and aggregation frameworks.

Experimental results have shown that the Hybrid approach is superior to the singular use of ODP and SAA approaches in situations that present none or at most some congestion, following the hypothesis that vehicles will travel in the same speed in a short period of time and space. The results of the first and second points of evaluation confirm this statement. However, at points where there is a lot of traffic jam, the performance of the Hybrid approach is negatively affected by the misbehaved produced instances. This shows how dependent the Hybrid approach is on the SAA approach.

As future work, we intend to propose a concurrent solution to this proposal by looking for improvements on the smooth sensitivity framework or alternatives to this one, or by using other techniques to get the median of a set with little noise, such as combining the sample and aggregate framework with exponential mech-

anism. Furthermore, we plan to evaluate the performance and security results of proposed approaches against a solution in a local model of computation.

A Proof Lemma 1

The Laplace distribution on \mathbb{R} with scale parameter b , is (α, β) -admissible with $\alpha = b\frac{\epsilon}{2}$ and $\beta = \frac{\epsilon}{2\ln(1/\delta)}$.

Proof. From Definition 3, we can obtain parameters α and β . Since the Laplace distribution is not a heavy tail distribution, then $\delta > 0$.

– Considering Eq. 3, we have

- When $\frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U + \Delta)} \geq 1$, we have

$$\begin{aligned} \frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U + \Delta)} &= \frac{\int_U \frac{1}{2b} e^{-\frac{|x|}{b}} dx - \frac{\delta}{2}}{\int_{U+\Delta} \frac{1}{2b} e^{-\frac{|x|}{b}} dx} \\ &= \frac{\frac{1}{2b} \int_c^d e^{-\frac{|x|}{b}} dx - \frac{\delta}{2}}{\frac{1}{2b} \int_c^d e^{-\frac{|x+\Delta|}{b}} dx} = \frac{\int_c^d e^{-\frac{|x|}{b}} dx - \frac{\delta}{2}}{\int_c^d e^{-\frac{|x+\Delta|}{b}} dx} \end{aligned} \quad (18)$$

Considering numerator of Eq. (18), we have to evaluate interval $[c, d]$ in two cases,

- * when $x \geq 0$:

$$\int_c^d e^{-\frac{x}{b}} dx = b(e^{-c/b} - e^{-d/b}), \quad (19)$$

- * and when $x < 0$:

$$\int_c^d e^{\frac{x}{b}} dx = -b(e^{c/b} - e^{d/b}). \quad (20)$$

Now, considering denominator of Eq. (18), we have

- * when $x \geq -\Delta$:

$$\int_c^d e^{-\frac{x+\Delta}{b}} dx = e^{-\Delta/b} b(e^{-c/b} - e^{-d/b}), \quad (21)$$

- * and when $x < -\Delta$:

$$\int_c^d e^{\frac{x-\Delta}{b}} dx = -e^{-\Delta/b} b(e^{c/b} - e^{d/b}). \quad (22)$$

By substituting Eq. (19) and Eq. (21) in Eq. (18) we obtain

$$\begin{aligned} &\frac{b(e^{-c/b} - e^{-d/b}) - \frac{\delta}{2}}{e^{-\Delta/b} b(e^{-c/b} - e^{-d/b})} \\ &= e^{\Delta/b} \frac{b(e^{-c/b} - e^{-d/b}) - \frac{\delta}{2}}{b(e^{-c/b} - e^{-d/b})} \leq e^{\epsilon/2} \\ &\Leftrightarrow e^{\Delta/b} \leq e^{\epsilon/2} \frac{b(e^{-c/b} - e^{-d/b})}{b(e^{-c/b} - e^{-d/b}) - \frac{\delta}{2}}. \end{aligned} \quad (23)$$

When δ tends to zero in Eq. (23), the ratio tends to 1. Thus, assuming a negligible δ , we get

$$\Delta \leq b(\epsilon/2) + \ln \left[\frac{b(e^{-c/b} - e^{-d/b})}{b(e^{-c/b} - e^{-d/b}) - \frac{\delta}{2}} \right] \approx b(\epsilon/2). \tag{24}$$

Similarly, by replacing Eq. (20) and Eq. (22) in Eq. (18) we get the same result, $\Delta \leq b(\epsilon/2)$.

- When $\frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U + \Delta)} < 1$, we have by symmetry that

$$\begin{aligned} \frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U + \Delta)} &\geq e^{-\epsilon/2} \\ &\approx e^{-\Delta/b} \geq e^{-\epsilon/2} \\ &\approx \Delta \leq b(\epsilon/2). \end{aligned} \tag{25}$$

Therefore, it is sufficient to admit $\alpha = b(\epsilon/2)$, so that the translation property is satisfied with probability $1 - \frac{\delta}{2}$.

– Considering Eq. (4), we have

- When $\frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U \cdot e^\lambda)} \geq 1$, we have

$$\begin{aligned} \frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U \cdot e^\lambda)} &= \frac{\int_U \frac{1}{2b} e^{-\frac{|x|}{b}} dx - \frac{\delta}{2}}{\int_{U \cdot e^\lambda} \frac{1}{2b} e^{-\frac{|x|}{b}} dx} \\ &= \frac{\int_c^d e^{-\frac{|x|}{b}} dx - \frac{\delta}{2}}{\int_c^d e^{-\frac{|e^\lambda x|}{b}} dx} \end{aligned} \tag{26}$$

Numerator of Eq. (26) is given by Eq. (19) and (20). On the other hand, denominator of Eq. (26) is given by evaluating interval $[c, d]$ in two cases,

* when $x \geq 0$:

$$\int_c^d e^{-\frac{e^\lambda x}{b}} dx = e^{-\lambda} b [e^{-(e^\lambda c)/b} - e^{-(e^\lambda d)/b}], \tag{27}$$

* and when $x < 0$:

$$\int_c^d e^{\frac{e^\lambda x}{b}} dx = -e^{-\lambda} b [e^{(e^\lambda c)/b} - e^{(e^\lambda d)/b}]. \tag{28}$$

By replacing Eq. (19) and Eq. (27) in Eq. (26) we obtain

$$\begin{aligned} \frac{b(e^{-c/b} - e^{-d/b}) - \frac{\delta}{2}}{e^{-\lambda} b [e^{-(e^\lambda c)/b} - e^{-(e^\lambda d)/b}]} &\leq e^{\epsilon/2} \\ e^\lambda &\leq e^{\epsilon/2} \frac{b [e^{-(e^\lambda c)/b} - e^{-(e^\lambda d)/b}]}{b(e^{-c/b} - e^{-d/b}) - \frac{\delta}{2}}. \end{aligned} \tag{29}$$

From an analysis of Eq. (29), we can conclude that, regardless of values of b, c and d , where $d > c$, the ratio tends to zero when we get high values of λ . This is because the value of δ is negligible. When we get λ tending to zero, the ratio tends to 1. Thus, an acceptable upper bound for λ , so that Eq. (29) is satisfied with high probability, is $\epsilon/(2\ln(1/\delta))$. This value tends to zero when we get a very small value for δ .

Similarly, by replacing Eq. (20) and Eq. (28) in Eq. (26) we obtain the same result, $\lambda \leq \epsilon/(2\ln(1/\delta))$.

- When $\frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U \cdot e^\lambda)} < 1$, we have by symmetry that

$$\frac{Pr_{X \sim h}(X \in U) - \frac{\delta}{2}}{Pr_{X \sim h}(X \in U \cdot e^\lambda)} \geq e^{-\epsilon/2}, \quad (30)$$

which results in $-\lambda \geq -\epsilon/(2\ln(1/\delta))$.

Therefore, to satisfy the dilation property with probability $1 - \frac{\delta}{2}$, it is enough to assume $\beta = \epsilon/(2\ln(1/\delta))$.

□

References

1. Xiong, Z., Sheng, H., Rong, W., Cooper, D.: Intelligent transportation systems for smart cities: a progress review. *Sci. China Inf. Sci.* **55**(12), 2908–2914 (2012)
2. Research and Consultation Summary Report. <https://www.transport.gov.scot/medi-a/41636/its-strategy-research-and-consultation-summary-report-july-2017.pdf>. Accessed 21 Oct 2019
3. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: TCC 2006, pp. 265–284 (2006)
4. Kargl, F., Friedman, A., Boreli, R.: Differential privacy in intelligent transportation systems. In: WiSec 2013, pp. 107–112 (2013)
5. Nelson, B., Olovsson, T.: Introducing differential privacy to the automotive domain: opportunities and challenges. In: IEEE 86th VTC-Fall, pp. 1–7 (2017)
6. Jemaa, I., Kaiser, A., Lonc, B.: Study of the impact of pseudonym change mechanisms on vehicular safety. In: IEEE Vehicular Networking Conference (VNC), pp. 259–262 (2017)
7. Hassan, M., Rehmani, M., Chen, J.: Differential privacy techniques for cyber physical systems: a survey. [arXiv:1812.02282](https://arxiv.org/abs/1812.02282) (2018)
8. Li, N., Lyu, M., Su, D.: Differential Privacy: From Theory to Practice. 1st edn. Morgan & Claypool Publishers (2016)
9. Nissim, K., Raskhodnikova, S., Smith, A.: Smooth sensitivity and sampling in private data analysis. In: 39th ACM STC, pp. 75–84 (2007)
10. Dwork, C., Naor, M., Pitassi, T., Rothblum, G.: Differential privacy under continual observation. In: Association for Computing Machinery Symposium on Theory of Computing, pp. 715–724 (2010)
11. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. *Found. Trends Theoret. Comput. Sci.* **9**, 211–407 (2014)
12. Codeca, L., Frank, R., Faye, S., Engel, T.: Luxembourg SUMO traffic (LuST) scenario: traffic demand evaluation. *IEEE Intell. Transp. Syst. Mag.* **9**, 52–63 (2017)

13. Krajzewicz, D., Behrisch, M., Bieker, L., Erdmann, J.: Recent development and applications of SUMO - Simulation of Urban MObility. *Int. J. Adv. Syst. Measur.* **5**(3 & 4), 128–138 (2012)
14. Sommer, C., German, R., Dressler, F.: Bidirectionally coupled network and road traffic simulation for improved IVC analysis. *IEEE Trans. Mob. Comput.* **10**(1), 3–15 (2011)
15. OMNeT++. <https://omnetpp.org/>. Accessed 22 Aug 2019
16. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: *FOCS 2007*, pp. 94–103 (2007)
17. Jaynes, E.: *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge (2003)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

