React to the Worst: Lightweight and proactive protection of location privacy

Emilio Molina, Mirko Fiacchini, Sophie Cerf, Bogdan Robu

React to the Worst: Lightweight and proactive protection of location privacy

Emilio Molina, Mirko Fiacchini, Sophie Cerf, and Bogdan Robu

***Abstract*— This work presents a novel optimal control method for privacy protection of mobility data. Protection is based on data obfuscation, consisting in sending to the geolocated service a finely tuned fake location. The objective is twofold, keeping privacy values at an acceptable level and guaranteeing a reasonable utility loss, with a lightweight algorithm able to run on mobile devices. The proposed method consists of an offline modeling stage, based on privacy worst-case anticipation, and a fast algorithm executed online. In the offline stage, the algorithm computes the average amount of allowed utility loss necessary to maintain the privacy value of the following h steps above a given lower bound. For this purpose, the worst possible scenario over the future steps is computed and compared with the privacy function of the solution obtained by an MPC method. The online stage uses the information computed offline to solve an optimization problem whose decision variable is the location to transmit and whose objective is to maintain the privacy value above a minimal level, by avoiding large utility losses. The method is validated on an instance of a database of real records and compared with a state-of-the-art competitor.**

***Index Terms*— Control applications, optimal control, optimization, predictive control for linear systems.**

I. INTRODUCTION

MASSIVE flows of data are constantly generated by connected devices. Among those, location data are notably sensitive, as they can reveal the identity of anonymous users [1], their homes and workplaces, favorite venues, and even social relationships, sexual orientation, or religion [2]. Leveraging of the shared information is needed to preserve privacy while ensuring the utility of geolocated services (navigation, recommendations, etc.) [3]. Mechanisms realizing privacy protection can be: (i) based on theoretical privacy definitions [4], [5], hardly usable in practice due to limited utility performance [6]; (ii) an optimal defense against an attack [7], therefore with limited robustness to different attacks and user profiles [8]. Most approaches are computed offline and require the knowledge of the entire mobility dataset, hence trust in a third-party [9]. The limitation of the literature relies on the practicality of protection mechanisms, to offer scalable and lightweight protection for individual users.

This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003) and by the French LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

Emilio Molina, Mirko Fiacchini and Bogdan Robu are with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France. {name.surname}@gipsa-lab.fr

Sophie Cerf is with the Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, Lille, F-59000, France. sophie.cerf@inria.fr

This work tackles the privacy protection challenge with a dynamic system's perspective. Privacy control with a reactive PI controller [10] allows for protection with negligible computing overhead; however, with limited precision and slow reaction. Model Predictive Control significantly improves privacy performances [11], but requires the knowledge of future locations and has a high computing cost. The objective of this work is to achieve a lightweight protection able to prevent the privacy levels to be lower than a reference value. More precisely, the objective is to keep acceptable values of privacy, higher than a given bound, by using worst-case predicted information to anticipate a violation of this constraint, without solving online any nonconvex optimization problem. To achieve this objective, the transmitted locations are used as control variables.

This work presents a scheme divided in two phases. The first phase, performed offline, learns the obfuscation needed on the actual location, at each step, to prevent the violation of a given bound on the privacy values, for keeping it at acceptable levels. This phase uses historical data and the MPC method proposed in [11]. The second phase is performed online: based on the information obtained in the first phase, an optimal obfuscated location is computed using the analytical solution of the optimization problem. Calculations in this phase are fast and lightweight, allowing its implementation on mobile devices. Compared to previous work [11], this approach does not need all-knowing prediction of future locations, as it is based on worst-case anticipation. Additionally, it is lightweight, hence feasible in practice, while the previous MPC solution requires computationally demanding solvers. Moreover, the proposed scheme permits to preserve a minimal privacy level with a high probability, property not present in [11].

Notation: : We denote by $l_k = (x(k), y(k)) \in \mathbb{R}^2$ the actual location of a user at time τ_k, and $\bar{l}_k = (x(k), y(k)) \in \mathbb{R}^2$ the obfuscated location transmitted at time τ_k to a third-party service. For a fixed finite duration $T > 0$ and a time $0 \leq T \leq \tau_k$ we denote with N the number of locations transmitted in $[\tau_k - T, \tau_k]$, with $\{\tau_{k-N+1}, \tau_{k-N+2}, \ldots, \tau_{k-1}, \tau_k\} \subset [\tau_k - T, \tau_k]$ the respective transmission times.

II. PROBLEM STATEMENT

With the aim of preventing any external agent to infer sensitive locations to protect user privacy, we consider the problem of obfuscation of user mobility data by transmitting modified locations $\bar{l}_k$. The main objective is to maintain the privacy level $p(t)$ above a threshold $\underline{p}$ with a reduced amount

of utility loss $q(t)$ (quality of service). Moreover, to obtain an anticipative effect, the optimization objective considers the privacy and the utility loss over a horizon of h future locations.

Some definitions concerning the privacy preservation problem are introduced. At time t, the privacy, as function in [9], is:

$$p(k) = \frac{1}{N} \sum_{j=k+1-N}^{k} \|\bar{l}_j - c(k)\|^2, \tag{1}$$

where $c(k) \in \mathbb{R}^2$ is the centroid of N locations $\bar{l}_1, \ldots, \bar{l}_N$ transmitted in $[t-T, t]$, computed as:

$$c(k) = (x_c(k), y_c(k)) = \frac{1}{N} \sum_{j=k+1-N}^{k} \bar{l}_j. \tag{2}$$

where $l_N = l(t)$. The privacy function $p(k)$ measures the spatial spread of the data transmitted within the horizon N. Low values of $p(k)$ represent significant stops of a user and then points of interest to be obfuscated. Note that this function is differentiable and then well adapted to use in mathematical optimization.[1]

The utility loss function at time τ_k is defined as the distance between the actual location l_k and the transmitted obfuscated location $\bar{l}_k$, that is :

$$q(k) = \|l_k - \bar{l}_k\|. \tag{3}$$

The bigger distance, the higher the service degradation. Since the aim of this work is to guarantee a certain minimal level of privacy with the minimal possible utility loss within a future horizon, the problem is posed in terms of a dynamical system whose state is the vector of locations transmitted within the past interval, as in [11], recalled hereafter. Consider a time interval, discretized in M points $\{\tau_k\}_{k=1}^{M}$, over which the future privacy evolution is evaluated.

The actual location at time τ_k is $(x(k), y(k))$ and, since at any time transmission might or might not have occurred, a binary variable $n(k)$ is defined, taking value 1 if the location at time τ_k is transmitted and 0 otherwise. The state $z(k) = (\mathbf{x}(k), \mathbf{y}(k), \mathbf{n}(k)) \in \mathbb{R}^N \times \mathbb{R}^N \times \{0,1\}^N$ acts as a buffer, storing the N transmission values in the time window $[\tau_k - T, \tau_k]$. Vectors $\mathbf{x}$ and $\mathbf{y}$ are the location states, and the vector $\mathbf{n}$ is the state of transmission occurrences. The transition system is defined as:

$$z(k+1) = \mathcal{A} \cdot z(k) + \mathcal{B} \cdot u(k) \tag{4}$$

where

$$\mathcal{A} = \begin{pmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & A \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{pmatrix}$$

with $u(k) = (x(k+1), y(k+1), n(k+1))$ and

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{N \times N}, \quad b = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^N$$

Note that as solution of this system we obtain:

$$\begin{aligned} \mathbf{x}_i(k) &= x(k+i-N), \\ \mathbf{y}_i(k) &= y(k+i-N), \\ \mathbf{n}_i(k) &= n(k+i-N) \end{aligned}$$

where i correspond to the ith coordinate of vectors $\mathbf{x}, \mathbf{y}$ and $\mathbf{n}$.

Using this notation, the centroid can be expressed as

$$x_c(z) = \frac{\sum_{i=1}^{N} \mathbf{x}_i \cdot \mathbf{n}_i}{\sum_{i=1}^{N} \mathbf{n}_i}, \qquad y_c(z) = \frac{\sum_{i=1}^{N} \mathbf{y}_i \cdot \mathbf{n}_i}{\sum_{i=1}^{N} \mathbf{n}_i} \tag{5}$$

and then the privacy function

$$p(k) = \frac{\sum_{i=1}^{N} ((\mathbf{x}_i(k) - x_c(k))^2 + (\mathbf{y}_i(k) - y_c(k))^2) \cdot \mathbf{n}_i(k)}{\sum_{i=1}^{N} \mathbf{n}_i(k)}. \tag{6}$$

Thus, we solve the following non-convex optimization problem:

$$\begin{aligned} \min_{D,\ (\delta x_i, \delta y_i)_{i=0}^{h}} \quad & D^2 \\ & p(\tilde{z}(k+i)) \geq \underline{p} && i \in \{0, .., h\}, \\ & \tilde{z}(k+i) = \mathcal{A}\tilde{z}(k+i-1) + \mathcal{B}\bar{u}(k+i-1), && i \in \{1, .., h\}, \\ & \bar{u}(k+i-1) = \begin{pmatrix} x(k+i) + \delta x_i \\ y(k+i) + \delta y_i \\ n(k+i) \end{pmatrix}, && i \in \{1, .., h\}, \\ & \delta x_i^2 + \delta y_i^2 = q^2(k) \leq D^2, && i \in \{0, .., h\}, \\ & \tilde{z}(k) = \bar{z}_h(k), \end{aligned} \tag{7}$$

where x and y denote the actual locations and $\bar{z}$ the transmitted, obfuscated position. The variable D corresponds to the utility loss upper bound which is minimized. Variables δx_i and δy_i represent the spatial perturbation to be applied to the actual positions. Finally, $\bar{z}$ and $\bar{u}$ are auxiliaries variables of the optimization problem, related to z and u in (4). See [11] for more details.

III. Preliminaries

The aim of this work is to propose a lightweight method to maintain the privacy values above a given bound in the following h steps by using a minimal utility loss, without the need of solving online any optimization problem.

We introduce two optimization problems that will be the key ingredient of our method:

1) since the future values of the actual location are not available at time k, the prediction should be performed

[1] An alternative expression of the privacy is $p(k) = \frac{1}{N} \sum_{j=k+1-N}^{k} \|\bar{l}_j - c(k)\|$, not differentiable at the origin, though.

by considering the worst-case scenario, i.e. the trajectory leading to the minimal values of the privacy function. The worst case location is given by the solution of

$$\min_{\bar{l}_N \in \mathbb{R}^2} \frac{1}{N} \sum_{k=1}^{N} \left\| \bar{l}_k - \frac{\sum_{j=1}^{N} \bar{l}_j}{N} \right\|^2 ; \tag{8}$$

2) since no optimization problem has to be solved online, an explicit solution is required for the problem of computing the obfuscated location that maximizes the privacy:

$$\begin{aligned} \max_{\bar{l}_N \in \mathbb{R}^2} \quad & \frac{1}{N} \sum_{k=1}^{N} \left\| \bar{l}_k - \frac{\sum_{j=1}^{N} \bar{l}_j}{N} \right\|^2 \\ \text{s.t.} \quad & \|\bar{l}_N - l_N\|^2 \leq D^2 \end{aligned} \tag{9}$$

for a given the value of D.

Both issues, whose solutions are addressed hereafter, are the basis of the proposed algorithm presented in the subsequent section.

A. Privacy Worst-Case Scenario

In this section, an explicit expression for the solution of (8) is provided.

Proposition 1: The only solution of problem (8) is:

$$\bar{l}_N^* = \frac{1}{N-1} \sum_{k=1}^{N-1} \bar{l}_k. \tag{10}$$

Proof: Note that the objective function is strictly convex and smooth, so, it has a unique minimum, and we can obtain it equaling its gradient to 0.

Writing $\bar{l}_k = (\bar{x}_k, \bar{y}_k)$, then the function to minimize is

$$\frac{1}{N} \sum_{k=1}^{N} \left(\bar{x}_k - \frac{\sum_{j=1}^{N} \bar{x}_j}{N} \right)^2 + \left(\bar{y}_k - \frac{\sum_{j=1}^{N} \bar{x}_j}{N} \right)^2$$

whose derivative with respect to x_N is:

$$\begin{aligned} \frac{\partial p_2}{\partial x_N} =& -\frac{2}{N^2} \sum_{k=1}^{N-1} \left(\bar{x}_k - \frac{\sum_{j=1}^{N} \bar{x}_j}{N} \right) \\ &+ \frac{2(N-1)}{N^2} \left(\bar{x}_N - \frac{\sum_{j=1}^{N} \bar{x}_j}{N} \right) \\ =& -\frac{2}{N^2} \sum_{k=1}^{N-1} \bar{x}_k + \frac{2(N-1)}{N^2} \bar{x}_N. \end{aligned} \tag{11}$$

The function (11) is null when $\bar{x}_N = \frac{1}{N-1} \sum_{k=1}^{N-1} \bar{x}_k$. We prove analogously $\bar{y}_N = \frac{1}{N-1} \sum_{k=1}^{N-1} \bar{y}_k$. ∎

From Proposition 1, it can be inferred that the worst future location in terms of privacy is the centroid of the previous $N-1$ transmitted ones since it makes the privacy level decrease the most, given by

$$\bar{l}_N^* = \frac{\sum_{i=1}^{N-1} (\mathbf{x}_i, \mathbf{y}_i) \cdot \mathbf{n}_i}{\sum_{i=1}^{N-1} \mathbf{n}_i}, \tag{12}$$

in terms of the transition system state.

B. Optimal Obfuscated location

Concerning item 2 above, since problem (9) is non-convex, it may admit many local solutions. Moreover, since we are maximizing a strictly convex function over a compact set, then the maximum value is reached at the boundary of the set, implying that its solutions satisfy $\|\bar{l}_N - l_N\|^2 = D^2$. The following proposition characterizes the solutions to the problem (9).

Proposition 2: Given $\bar{l}_N^*$ as in (10) and denoting by S the solution set of (9), the following claim holds:

i) If $l_N = \bar{l}_N^*$, i.e, if the actual location at time t is equal to the centroid of the previous $N-1$ transmitted locations, then the solution set of (9) is

$$S = \{\bar{l}_N \in \mathbb{R}^2 : \|\bar{l}_N - l_N\|^2 = D^2\},$$

and the optimal privacy value is:

$$\frac{1}{N} \sum_{k=1}^{N-1} \|\bar{l}_k - l_N\|^2 + D^2 \frac{N-1}{N^2} \tag{13}$$

ii) Otherwise, S has just two elements:

$$S = \left\{ l_N + D \frac{\bar{l}_N^* - l_N}{\|\bar{l}_N^* - l_N\|}, l_N - D \frac{\bar{l}_N^* - l_N}{\|\bar{l}_N^* - l_N\|} \right\} \tag{14}$$

Proof: Without loss of generality we can assume $l_N = 0$, then the condition in i) corresponds to $\bar{l}_N^* = 0$. We will prove that for $\|\bar{l}_N\|^2 = D^2$ the objective function in (9) is constant.

$$\begin{aligned} \frac{1}{N} \sum_{k=1}^{N} \left\| \bar{l}_k - \frac{\sum_{j=1}^{N} \bar{l}_j}{N} \right\|^2 &= \frac{1}{N} \sum_{k=1}^{N} \left\| \bar{l}_k - \frac{\bar{l}_N}{N} \right\|^2 \\ &= \frac{1}{N} \sum_{k=1}^{N} \left[\|\bar{l}_k\|^2 - 2 \left\langle \bar{l}_k, \frac{\bar{l}_N}{N} \right\rangle + \left\| \frac{\bar{l}_N}{N} \right\|^2 \right] \\ &= \frac{1}{N} \sum_{k=1}^{N-1} \|\bar{l}_k\|^2 + \frac{\|\bar{l}_N\|^2}{N} + \frac{\|\bar{l}_N\|^2}{N^2} - 2 \left\langle \sum_{k=1}^{N} \bar{l}_k, \frac{\bar{l}_N}{N^2} \right\rangle \\ &= \frac{1}{N} \sum_{k=1}^{N-1} \|\bar{l}_k\|^2 + \frac{\|\bar{l}_N\|^2}{N} + \frac{\|\bar{l}_N\|^2}{N^2} - 2 \left\langle \bar{l}_N, \frac{\bar{l}_N}{N^2} \right\rangle \\ &= \frac{1}{N} \sum_{k=1}^{N-1} \|\bar{l}_k\|^2 + \frac{D^2}{N} - \frac{D^2}{N^2} \end{aligned}$$

We obtain then expression (13). Moreover, as the solution of problem (9) is reached when $\|\bar{l}_N\|^2 = D^2$, therefore every point in this circumference is an optimal solution.

Consider now the case $\bar{l}_N^* \neq 0$, and recall the notation $\bar{l}_N = (\bar{x}_N, \bar{y}_N)$. Thanks to Karush-Kuhn-Tucker theorem, there exists $\mu \geq 0$ such that

$$-\frac{\partial p_2}{\partial x_N} + 2\mu \bar{x}_N = 0, \tag{15}$$

$$-\frac{\partial p_2}{\partial y_N} + 2\mu \bar{y}_N = 0 \tag{16}$$

Using the equation (15) and (11), it follows

$$\frac{2}{N^2}\sum_{k=1}^{N-1}\bar{x}_k - \frac{2(N-1)}{N^2}\bar{x}_N + 2\mu\bar{x}_N = 0$$

and then

$$\bar{x}_N = \frac{-1}{N^2\mu-(N-1)}\sum_{k=1}^{N-1}\bar{x}_k. \tag{17}$$

Similarly, we obtain

$$\bar{y}_N = \frac{-1}{N^2\mu-(N-1)}\sum_{k=1}^{N-1}\bar{y}_k \tag{18}$$

And from $\|\bar{l}_N\|^2 = x_N^2 + y_N^2 = D^2$ we get:

$$\frac{1}{N^2\mu-(N-1)} = \pm\frac{D}{\sqrt{\left(\sum_{k=1}^{N-1}\bar{x}_k\right)^2+\left(\sum_{k=1}^{N-1}\bar{y}_k\right)^2}}$$

Replacing in equations (17) and (18), we obtain

$$\pm D\frac{\sum_{k=1}^{N-1}\bar{l}_k}{\left\|\sum_{k=1}^{N-1}\bar{l}_k\right\|} = \pm D\frac{\bar{l}_N^*}{\|\bar{l}_N^*\|}.$$

To recover (13) and (14), when $l_N \neq 0$, just consider the change of coordinates $\tilde{l}_k = \bar{l}_k - l_N$ ■

The method proposed in the following section exploits Propositions 1 and 2, that, based on explicit solutions, considerably improve the online execution times, with no need of optimization solvers.

IV. Fast MPC-based obfuscation

The proposed method is composed of two stages, first an offline data-based structure computation and then its online implementation. The offline stage consists of learning the value of the utility loss bound, D in (9), necessary to have a privacy level higher than a reference value $\underline{p}$ in the next h steps, being $\underline{p} \geq 0$ and $h \in \mathbb{N}$ two parameters of the method. In addition, the value N used to compute the privacy values will be fixed for both offline and online instances.

In the online implementation, the information learned in the first stage is used to solve problem (9). In this stage the objective is the same, i.e. to prevent the drop of the privacy value below $\underline{p}$ but using faster algorithms. In the prediction, the worst possible evolution, given by the solution of problem (8), is considered, to react to the most adverse scenario for the privacy. In the following, both stages are illustrated.

A. Privacy Gain Computation Stage

Given a privacy bound $\underline{p}$ and the horizon h, and an interval $\{\tau_k\}_{k=1}^M$ we train the model as follows. Based on the transition system in (4), for an index $k \in \{1,\dots,M-h\}$, and the actual location $(x(k),y(k))$, the points $(x^1(k),y^1(k))$, $(x^2(k),y^2(k))$, …, $(x^h(k),y^h(k))$ correspond to future locations leading to the minimal value of privacy. These locations are iteratively computed from 1 to h using (12) and the previous $N-1$ actual locations. We denote by $p_{wc}(k)$ the privacy computed using (6) and the h predicted locations along with $(x(k),y(k))$ and its $N-h-1$ previous actual locations.

Consider a set of d upper bounds of the utility loss $\{D_1,\dots,D_d\} \subseteq [0, D_{max}]$ where D_{max} is the maximum value allowed for the utility loss. The procedure in the offline stage is as follows.

Given D_j with $j \in \{1,\dots,d\}$, the proposed method iterates over $k \in \{1,\dots,M-h\}$. When $n(k) = 1$ we compute $(x^1(k),y^1(k))$ to $(x^h(k),y^h(k))$ and then $p_{wc}(k)$. Then, we solve an MPC instance using as predicted locations $(x^1(k),y^1(k))$ to $(x^h(k),y^h(k))$ and D_j as the upper bound for the utility loss (the particular MPC method used corresponds to the one presented in section 3.2 in [11]). We obtain $(\bar{x}(k),\bar{y}(k))$, $(\bar{x}^1(k),\bar{y}^1(k))$,…, $(\bar{x}^h(k),\bar{y}^h(k))$ obfuscated locations. We use those points along with $N-h-1$ previous actual locations to compute the resulting privacy at point $(\bar{x}^h(k),\bar{y}^h(k))$ that we call $p_{MPC}(k)$. We finish the iteration saving the value $\Delta p^{D_j}(k) = p_{MPC}(k) - p_{wc}(k) \geq 0$ and updating the transition system using the actual location .

The value $\Delta p^{D_j}(k)$ is the gain in the privacy value at $k+h$ obtained if the MPC-based optimized obfuscation is used with the worst case as predicted real trajectory. At the end of the iteration process, we can then derive statistical information as the maximal or average gain. In this work the average is considered, but other statistical information could be used in future work. Using the average gain computed for every D_j, we build a piece-wise linear function $f_h : [0, D_{max}] \to [0,\infty)$ using linear interpolation. The value $f_h(D)$ represents the compared difference after h steps between the privacy obtained in the worst case and the privacy got using an MPC method, with a bound in the utility loss equal to D. We assume $f_h(0) = 0$. This function is the structure used to implement the method in the online stage. The pseudocode of this stage is in Algorithm 1.

B. Transmission Stage

Suppose we apply the online stage from a time t_0 until a time t_F. Let $t_k > t_0$ be a time at which a new location has to be transmitted. Recall that $(x(k),y(k))$ and $(\bar{x}(k),\bar{y}(k))$ are the actual and the transmitted locations, respectively. To determine the obfuscated location to be transmitted, the method consists of these steps:

1) Compute the worst case locations in the following h steps, i.e., $(x^1(k),y^1(k))$ to $(x^h(k),y^h(k))$. Along with the transmitted locations $(\bar{x}(k-(N+1-h)),\bar{y}(k-(N+1-h)),\dots,(\bar{x}(k-1),\bar{y}(k-1))$ compute $p_{wc}(k)$ and $\Delta p = p_{wc}(k) - \underline{p}$.
2) If $\Delta p \geq 0$, then in the following h steps, the privacy values will be over the lower bound, even in the worst scenario, and thus the location obfuscation is not necessary. The actual location is transmitted, i.e.

$$(\bar{x}(k),\bar{y}(k)) = (x(k),y(k)).$$

3) If $\Delta p < 0$, the function f_h is used to compute $D(k) = f_h^{-1}(-\Delta p)$ when $-\Delta p \leq f_h(D_{max})$ and $D(k) = D_{max}$ in the other case. This value represents the utility loss

Algorithm 1 Offline training

Input: actual locations (x, y, n), h, t_F, $\{D_1, \dots, D_d\}$
Output: function f_h
for $j := 1 \dots d$ **do**
  $k \leftarrow 0$, $z(0) \leftarrow 0$
  while $k \leq t_F$ **do**
    $z(k+1) \leftarrow \mathcal{A} \cdot z(k) + \mathcal{B} \cdot u(k)$
    where $u(k) = (x(k+1), y(k+1), n(k+1))$
    $z^1_{aux}(0) \leftarrow z(k+1)$
    for $i := 1, \dots, h$ **do**
      $(x^i(k), y^i(k)) \leftarrow \bar{l}^*_N(z^1_{aux}(i-1))$ using (12)
      $z^1_{aux}(i) \leftarrow \mathcal{A} \cdot z^1_{aux}(i-1) + \mathcal{B} \cdot v(i)$
      where $v(i) = (x^i(k), y^i(k), 1)$
    end for
    $p_{wc}(k) \leftarrow p(z^1_{aux}(h))$ using (6)
    $z^2_{aux}(0) \leftarrow z(k+1)$
    for $i := 1, \dots, h$ **do**
      $(\bar{x}^i(k), \bar{y}^i(k)) \leftarrow$ solution of MPC$-h$ instance using $\{(x^i(k), y^i(k))\}_{i=1}^h$ and D_j
      $z^2_{aux}(i) \leftarrow \mathcal{A} \cdot z^2_{aux}(i-1) + \mathcal{B} \cdot w(i)$
      where $w(i) = (\bar{x}^i(k), \bar{y}^i(k), 1)$
    end for
    $p_{MPC}(k) \leftarrow p(z^2_{aux}(h))$ using (6)
    $\Delta p^{D_j}(k) \leftarrow p_{MPC}(k) - p_{wc}(k)$
  end while
  $\Delta p^{D_j}_m \leftarrow mean\left(\{\Delta p^{D_j}(k) | n(k) = 1\}\right)$
end for
$f_h \leftarrow$ linear interpolation using $\left\{\Delta p^{D_j}_m\right\}_{j=1}^d$

Algorithm 2 Online implementation

Input: actual locations (x, y, n), h, f_h, $\underline{p}$, t_F
Output: obfuscated locations $(\bar{x}, \bar{y})$
Start: $k \leftarrow 0$, $z(0) \leftarrow 0$
while $k \leq t_F$ **do**
  $z(k+1) \leftarrow \mathcal{A} \cdot z(k) + \mathcal{B} \cdot u(k)$
  where $u(k) = (x(k+1), y(k+1), n(k+1))$
  $z_{aux}(0) \leftarrow z(k+1)$
  for $i := 1, \dots, h$ **do**
    $(x^i(k), y^i(k)) \leftarrow \bar{l}^*_N(z_{aux}(i-1))$ using (12)
    $z_{aux}(i) \leftarrow \mathcal{A} \cdot z_{aux}(i-1) + \mathcal{B} \cdot v(i)$
    where $v(i) = (x^i(k), y^i(k), 1)$
  end for
  $p_{wc}(k) \leftarrow p(z_{aux}(h))$ using (6)
  $\Delta p \leftarrow p_{wc}(k) - \underline{p}$
  if $\Delta p \leq 0$ **then**
    if $-\Delta p \leq f_h(D_{max})$ **then**
      $D(k) \leftarrow f_h^{-1}(-\Delta p)$
    else
      $D(k) \leftarrow D_{max}$
    end if
    $\bar{l}_{\pm} \leftarrow \bar{l}^{**}_N$ using (14)
    $(\bar{x}(k), \bar{y}(k)) \leftarrow \arg\max(l_+, l_-)$
  else
    $(\bar{x}(k), \bar{y}(k)) \leftarrow (x(k), y(k))$
  end if
end while

quantity needed in average to keep the privacy value higher than $\underline{p}$ but using no more than D_{max}.

4) The problem (9) is solved using $D(k)$ instead of D, by comparing the objective values of two solutions in (14), which consistently simplifies its online implementation and speed up its resolution. Finally, the transmitted location $(\bar{x}(k), \bar{y}(k))$ is the obtained solution.

This algorithm keeps the privacy over $\underline{p}$ most of the time, but since we are using average values of the privacy gain, there could have some particular cases when privacy goes under $\underline{p}$. This can usually happen when $\Delta p > f_h(D_{max})$ because in that case, we use a utility loss equal to D_{max} which just guarantees that we can gain in average $f_h(D_{max})$. When Δp^{D_j} does not take extreme values, this method should offer good performances by requiring very limited computational resources. The pseudocode is in Algorithm 2.

V. Numerical examples

We present the results of applying our method to an instance from the datasets Cabspotting [12]. In particular, we took Oilrag user, and to have a uniform transmitted times distribution, data are re-sampled to obtain intervals of $30s$. Oilrag is the location trace of a cab, its whole duration is 18.7 hours. We used for the offline stage the first $20000s$, that is, around 5.5 hours. In that period, the location was transmitted 436 times. Figure 1 shows the function f_h that we obtained in the offline stage for $h = 4$ and $D_{max} = 2000$. We also plot the lines corresponding to percentiles related to 5% and 95%, giving us information about the data distribution.

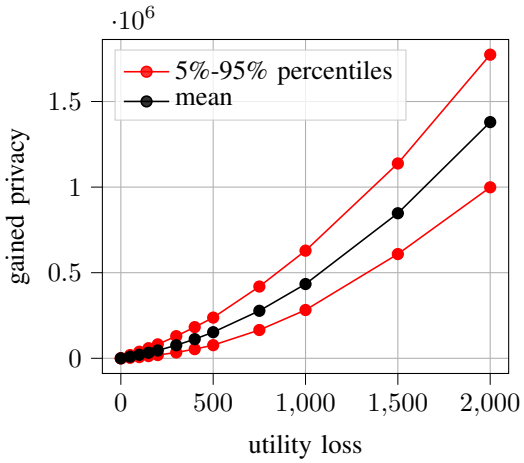

Fig. 1. Privacy gained comparing h steps in the future.

We assess online performance using the data from $20000s$ to $30000s$ (duration of 2.7 hours). In figure 2 we show three different scenarios, low ($\underline{p} = 5 \cdot 10^5$), regular ($\underline{p} = 10^6$) and high ($\underline{p} = 1.5 \cdot 10^6$) values for the lower bound of the privacy. We observe that the percentage of failure is low in every instance (lower than 3%). Fails increase with the lower bound value, due to the fact that the maximal utility loss, equal to 2000, might not always be high enough to raise the privacy over the higher lower bound, in a sort of saturation effect.

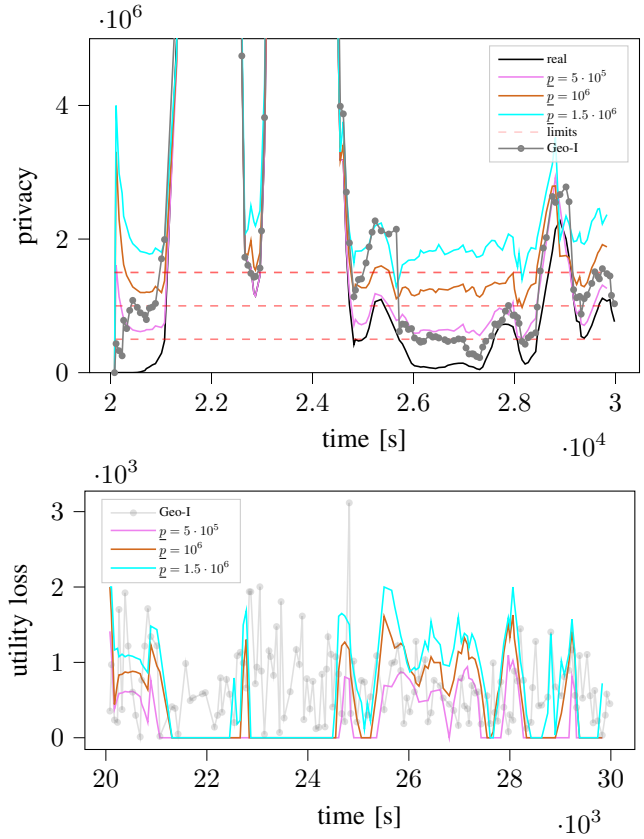

Fig. 2. Privacy and utility through time obtained in online implementation and Geo-I.

The performance of our method are compared with those of the state-of-the-art Geo-I mechanism [13]. This mechanism adds spatial noise using the expression:

$$\bar{l}(t) = l(t) - \frac{1}{\epsilon}\left[W_{-1}\left(\frac{\alpha(t)-1}{e}\right)+1\right]\begin{pmatrix}\cos\ \theta(t)\\ \sin\ \theta(t)\end{pmatrix} \quad (19)$$

where W_{-1} is the Lambert W function (the -1 branch), e is Euler's number, $\alpha(t)$ is drawn uniformly in $[0,\ 1)$ and $\theta(t)$ in $[0,\ 2\pi)$. Taking $\epsilon = 0.00275$ we obtain privacy values of the order of those given by the method proposed here. In table I and figure 2 we note that Geo-I has a higher utility loss than our method while obtaining similar privacy values. Moreover, its associated privacy has several oscillations, violating the lower bound in many times.

method	av. privacy	improvement	av. utility loss	% fails
$\underline{p} = 5 \cdot 10^5$	6137799	5.5%	253.6	0.6%
$\underline{p} = 10^6$	6465300	11.1%	478.7	1.4%
$\underline{p} = 1.5 \cdot 10^6$	6885697	18.3%	662.5	2.8%
Geo-I	6581841	13.1%	698.9	N.A.

TABLE I

PERFORMANCE OF THE THREE SCENARIOS AND $\boldsymbol{Geo-I}$. IMPROVEMENTS REGARDING REAL PRIVACY VALUE $\mathbf{5815745}$. THE LAST COLUMN SHOWS THE PERCENTAGE OF TIMES WHEN PRIVACY IS LOWER THAN $\underline{\boldsymbol{p}}$. AVERAGE VALUES OVER TIME.

VI. CONCLUSION

This paper presents a lightweight obfuscation approach, based on MPC, to ensure a minimal privacy level by reducing the required utility loss. Particular attention is given to computational feasibility on mobile devices. The proposed method is consists of two phases, one offline and one online, and is based on worst case anticipation of future privacy. The instance solved shows promising results, respecting almost always the privacy constraint, with utility loss lower than the state-of-the-art Geo-I for similar privacy value. Future works will consider testing and evaluating the method on more datasets, analyzing the computation overhead and runtimes, a sensitivity analysis of the solutions with respect to the design parameters h, $\underline{p}$ and D.

REFERENCES

[1] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, "Show Me How You Move and I Will Tell You Who You Are," Transactions on Data Privacy, vol. 4, pp. 103–126, Aug. 2011.

[2] V. Srivastava, V. Naik, and A. Gupta, "Privacy breach of social relation from location based mobile applications," in 2014 7th International Conference on Contemporary Computing, pp. 324–328, IEEE, 2014.

[3] H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, and A. Iyengar, "Location privacy-preserving mechanisms in location-based services: A comprehensive survey," ACM Computing Surveys, vol. 54, no. 1, pp. 1–36, 2021.

[4] C. Dwork, "Differential Privacy," in Automata, Languages and Programming, vol. 4052 of Lecture Notes in Computer Science, pp. 1–12, Springer Berlin Heidelberg, 2006.

[5] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," IEEE Transactions on Mobile Computing, vol. 7, no. 1, pp. 1–18, 2007.

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

[7] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in Proceedings of the 2012 ACM conference on Computer and communications security, pp. 617–627, 2012.

[8] H. Brenner and K. Nissim, "Impossibility of differentially private universally optimal mechanisms," SIAM Journal on Computing, vol. 43, no. 5, pp. 1513–1540, 2014.

[9] V. Primault, A. Boutet, S. B. Mokhtar, and L. Brunie, "The long road to computational location privacy: A survey," IEEE Communications Surveys & Tutorials, 2018.

[10] S. Cerf, B. Robu, N. Marchand, and S. Bouchenak, "Privacy protection control for mobile apps users," Control Engineering Practice, 2023.

[11] E. Molina, M. Fiacchini, S. Cerf, and B. Robu, "Optimal privacy protection of mobility data: a predictive approach," in IFAC WC 2023 - 22nd IFAC World Congress, 22nd IFAC World Congress, (Yokohama, Japan), July 2023.

[12] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "Crawdad epfl/mobility." Downloaded from `https://dx.doi.org/10.15783/C7J010`, 2022.

[13] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential Privacy for Location-based Systems," in CCS, pp. 901–914, ACM, 2013.