Optimal privacy protection of mobility data: a predictive approach

Emilio Molina, Mirko Fiacchini, Sophie Cerf, Bogdan Robu

Optimal privacy protection of mobility data: a predictive approach ⋆

Emilio Molina * **Mirko Fiacchini** * **Sophie Cerf** ** **Bogdan Robu** *

* *Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, 38000, France. (e-mail:* {name.surname}@gipsa-lab.fr*)*
** *Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, Lille, F-59000, France. (e-mail:* sophie.cerf@inria.fr*)*

Abstract:
Location data are extensively used to provide geo-personalized contents to mobile devices users. Sharing such personal data is a major threat to privacy, with risks of re-identification or inference of sensitive information. Location data broadcasted to services can be sanitized, i.e., by adding noise to spatial coordinates. Such protection mechanisms from the literature are widely generic, e.g., not specific to a user and mobility properties. In this work, we advocate that taking into account the specificities of location data (temporal correlation, human mobility patterns, etc.) enables to gain in the privacy-utility trade-off. Specifically, using future mobility prediction greatly improves privacy. We present a novel protection mechanism, based on model predictive control (MPC). The sanitized location is optimally computed so that it maximizes privacy while guaranteeing a utility loss constraint, for present and future locations. Our formulation explicitly takes into account non-constant sampling time, due to moments when no location data is broadcasted. We evaluate experimentally our control on real mobility dataset and compare to the state of the art. Results show that with knowledge of user's future mobility over a few of minutes, we can gain up to 10% of privacy compared to state of the art, while preserving data utility.

Keywords: Security and privacy; Model predictive and optimization-based control; Predictive control.

1. INTRODUCTION

With the generalization of smart mobile devices, such as phones or smart watches, location data are more than ever a goldmine. Geo-located services are flourishing, such as navigation, venue finders or dating apps (Google Play, 2022). Share one's mobility data to a third party presents however threats to privacy, by exposing highly sensitive personal information. Extracting location points of interest, attackers can discover users' identity, social relationships, and even religious, political or sexual orientations (Gambs et al., 2011). In Europe, the General Data Protection Regulation (GDPR) is a law which regulate data privacy issues. The cumulative sum of GDPR fines until December 2022 was €2381309317 (McCarthy, 2023).

Protection mechanisms have been proposed to enhance one's location privacy. Among the vast literature (Primault et al., 2018; Jiang et al., 2021), some works tackle the scenario of an individual continuously sending his or her mobility data. Such protection mechanisms are mainly based on obfuscation: the location data is disturbed with some spatial noise before being transmitted to the service as in Geo-Indistinguishability (Andrés et al., 2013), inspired from the concept of differential-privacy (Dwork, 2006). Geo-I performs blind obfuscation, in that it applies constant noise for all locations and at all times. It has been extended to tackle this first limitation, with location-dependent privacy (Koufogiannis and Pappas, 2016; Chatzikokolakis et al., 2015), correlating the noise to the location population density. The particular case of obfuscation in an Indoor environment has been also studied by Meira-Góes et al. (Meira-Góes et al., 2018). An open challenge remains on the noise adaptation through time. Chatzikokolakis et al.(Chatzikokolakis et al., 2014) take a step in this direction by taking into account the predictability of the user movement to take a binary decision on whether to sanitize or not the transmitted location. In this work, we take into account the human mobility property of predictability, and present a *time-dynamic* protection mechanism that creates a sanitized mobility trace.

Other protection mechanisms focus on finding optimal counter-attacks. Shokri et al. (Shokri et al., 2012) proposed the first optimal formulation, extended for Geo-I (Bordenabe et al., 2014), and later refined by using additional dimensions of privacy than just attackers' performance (Oya et al., 2017). Indeed, optimal protections are often limited to defeating a specific attack, and result in huge data distortion that render the location-based service useless (Krumm, 2007). Most of those works additionally assume a sporadic data transmission, i.e. not a continuous data broadcast. Recent works tackle the dynamical optimal protection challenge (Yu et al., 2017; Xiao and Xiong, 2015), however without making use of mobility prediction over a future time horizon. Zhang et al. (Zhang et al., 2018) propose an information-theoretic approach that exploits the temporal location correlations to reduce the protection mechanism complexity—but not to enhance protection. In contrast to the state-of-the-art, to achieve privacy protection in practice, we (i) consider a *continuous scenario*, (ii) use *mobility prediction*

⋆ This work has been partially supported by MIAI@Grenoble Alpes (ANR19-P3IA-0003).

over a time horizon to enhance protection, and (iii) rely on a *privacy metric* based on the established notion of points of interests (rather than on defeating specific attacks).

In this paper, we present a novel protection mechanism, called `p mpc-H`, based on non-convex optimal predictive control. The obfuscated position to transmit to the service is computed at each timestep as the solution of an optimization problem. Privacy is defined based on the extraction of users' points of interest, and utility as data distortion. Model predictive control is used to maximize privacy while ensuring a minimal utility level. Future mobility over a limited horizon is taken into account in the optimization, so that the protection can anticipate on the user next move and foster privacy protection. The main challenges of our approach rely in (i) the non-convexity of the problem, and (ii) in the non-constant sampling time of data broadcast—requiring a novel formulation.

We show that including knowledge on the user's next moves allows significant privacy protection improvements. This approach let you compute off-line and on-line instances. In the examples show in this paper, it is assumed the availability of a prediction of the user mobility computing an off-line obfuscation. In a practical scenario, such information can come from the user input or from a prediction algorithm (Yavaş et al., 2005; Gambs et al., 2012): this aspect is the objective of one future research. Analysis of the robustness of our proposed control for mobility prediction with reduced accuracy is let as future work.

The remaining of the paper is organized as follows. Section 2 formalizes the problem, defining metrics and the optimal problem. Section 3 presents our control solution for this non-convex problem. Section 4 details the experimental setup and the evaluation on three aspects: privacy gains, benefit of future knowledge and overhead considerations. Conclusion ends this paper in Section 5.

Notation Through this paper, we will use overline bar to reference parameters associated to a sanitized position. We use bold notation with vectors and sub-index representing the coordinate of a vector, for example, $\mathbf{x}_i$ correspond to coordinate i of the vector $\mathbf{x}$. On the other hand, to express dependence on time t, we use parenthesis $x(t)$.

Finally, for $a < b$ integers, we introduce the following notation $X_a^b := \{X(k)\}_{k=a}^b = \{X(a), X(a+1), ..., X(b-1), X(b)\}$.

2. PROBLEM FORMULATION

We consider the problem of obfuscation of a user mobility trace. The objective is to transmit modified positions which preserve certain privacy levels.

We denote by $l(t) = (x(t), y(t)) \in \mathbb{R}^2$ the actual position of a user at time t, and $\bar{l}(t) \in \mathbb{R}^2$ the sanitized position transmitted at time t to a third-party service by some method. An example of this method is presented in Fig. 1. In the image the actual positions of a user on a map is represented in dark blue, and in light orange positions the obfuscated trace obtained using a state of the art protection mechanism called GeoI (Andrés et al., 2013). That mechanism consists in adding random noise to the actual position, further details are given in Section 4.1.

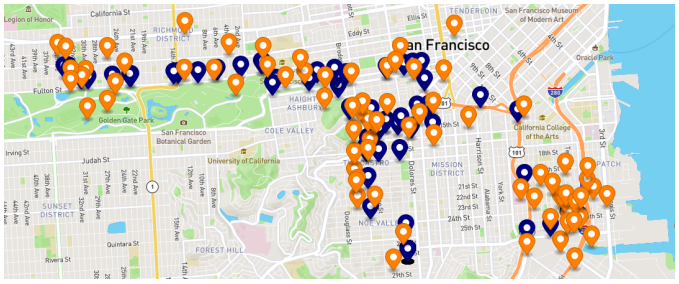

Fig. 1. Illustration of a mobility trace on a map, comparison of actual position (blue) and sanitized one using Geo-I (orange). User `oilrag`.

2.1 Privacy and Utility Measures

For a time t and a duration T, we denote by $N_T(t)$ the number of times in $[t-T, t]$ where the position was transmitted and:

$$t_1 < t_2 < ... < t_{N_T(t)-1} < t_{N_T(t)}$$

the respective transmission times in increasing order. This times depend on t but we omit it to maintain a more pleasant notation.

At time t, the centroid $c(t) \in \mathbb{R}^2$ of the mobility trace l over a past window of length T is

$$c(t) = (x_c(t), y_c(t)) = \frac{1}{N_T(t)} \sum_{k=1}^{N_T(t)} l(t_k). \tag{1}$$

Therefore, the privacy level at time t can be defined as:

$$p(t) = \frac{1}{N_T(t)} \sum_{k=1}^{N_T(t)} \|l(t_k) - c(t)\|_2. \tag{2}$$

This metric is an online measure of spatial distortion, reflecting the exposure of one's points of interests (Cerf et al., 2018). In practice, to have privacy value 0 means keeping at the same position for a period of time T, That position is a possible point of interest to be tracked. Note that we use the mean dispersion for simplification reasons, maximum (Primault et al., 2014) or median (Cerf, 2019) could also be used. The two previous expressions hold also in the case where a sanitized position $\bar{l}(t)$ is used, in this case they will be denoted by $\bar{c}$ and $\bar{p}$.

The utility loss function is given by the distance between the real position denoted by $l(t)$ and the transmitted one $\bar{l}(t)$ (Oya et al., 2017), that is :

$$q(t) = \|l(t) - \bar{l}(t)\|_2. \tag{3}$$

2.2 Optimal problem

Privacy is computed using past and present location data. However, at a given time t, only the current position can be controlled since the previous one have already been transmitted. The obfuscation problem thus optimizes only the current position $\bar{l}(t)$. We then write the privacy as a function depending only on the current time and the variable $\bar{l}(t)$:

$$\bar{p}(t, \bar{l}(t)) = \frac{1}{N_T(t)} \sum_{k=1}^{N_T(t)} \|\bar{l}(t_k) - \bar{c}(t, \bar{l}(t))\|_2 \tag{4}$$

with

$$\bar{c}(t, \bar{l}(t)) = \frac{1}{N_T(t)} \sum_{k=1}^{N_T(t)} \bar{l}(t_k).$$

where $\bar{l}(t_1),...,\bar{l}(t_{N_T(t)-1})$ are fixed and $\bar{l}(t_{N_T(t)}) = \bar{l}(t)$ to be chosen.

A optimization method will consist in applying the following obfuscated position

$$\bar{l}^*(t) = \arg\max_{\bar{l}\in\mathbb{R}^2} \bar{p}(t,\bar{l}) \quad \text{s.t. } \|l(t)-\bar{l}\|_2^2 \leq \Delta^2, \tag{5}$$

whose objective is to compute the position to be transmitted that maximize the privacy, among those that guarantee a bound on the utility loss. Note that we use the square of the utility loss, this constraint is equivalent to $\|l(t)-\bar{l}\|_2 \leq \Delta$ and more suitable to optimization methods used in the following sections. This problem is non-convex, then we could have many local solutions.

3. CONTROL SOLUTION

Consider a time interval $[0,\tau]$, where we will apply an obfuscation method, discretized in M points noted $\{s_k\}_{k=1}^M$. We consider N points with $N < M$ representing the number of points in a mobile window $[s_k - T, s_k]$ except when $s_k < T$, in that case we adjusted its length.

3.1 Transition system

In order to keep notation simple, let us denote the real position at time s_k by $(x(k),y(k))$. To deal with non-constant sampling time, we introduce the binary variable $n(k)$ which takes value 1 if the position at time s_k is transmitted and 0 otherwise.

We write the state variable $z(k) = (\mathbf{x}(k),\mathbf{y}(k),\mathbf{n}(k)) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$, which will have the information of the time windows $[s_k - T, s_k]$. $\mathbf{x}$ and $\mathbf{y}$ are vectors representing the part associated to the position, and the vector $\mathbf{n}$ related to the transmission. The transition system for these variables is:

$$\begin{array}{llll} \mathbf{x}_1(k+1) & = \mathbf{x}_2(k), & \mathbf{y}_1(k+1) & = \mathbf{y}_2(k), \\ \mathbf{x}_2(k+1) & = \mathbf{x}_3(k), & \mathbf{y}_2(k+1) & = \mathbf{y}_3(k), \\ & \vdots & & \vdots \\ \mathbf{x}_{N-1}(k+1) & = \mathbf{x}_N(k), & \mathbf{y}_{N-1}(k+1) & = \mathbf{y}_N(k), \\ \mathbf{x}_N(k+1) & = x(k+1), & \mathbf{y}_N(k+1) & = y(k+1). \end{array}$$

and the transmission part:

$$\begin{array}{ll} \mathbf{n}_1(k+1) & = \mathbf{n}_2(k), \\ \mathbf{n}_2(k+1) & = \mathbf{n}_3, \\ & \vdots \\ \mathbf{n}_{N-1}(k+1) & = \mathbf{n}_N(k), \\ \mathbf{n}_N(k+1) & = n(k+1). \end{array} \tag{6}$$

From above equations it is possible to note that $\mathbf{x},\mathbf{y}$ and $\mathbf{n}$ are acting like buffers saving the values in the time window $[s_k - T, s_k]$ and updated each time in the last position ($\mathbf{x}_N, \mathbf{y}_N$ and $\mathbf{n}_N$). This system can be equivalently written as

$$\begin{array}{ll} \mathbf{x}(k+1) & = A\cdot\mathbf{x}(k) + b\cdot x(k+1), \\ \mathbf{y}(k+1) & = A\cdot\mathbf{y}(k) + b\cdot y(k+1), \\ \mathbf{n}(k+1) & = A\cdot\mathbf{n}(k) + b\cdot n(k+1), \end{array} \tag{7}$$

with

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

Previous system can be written in a more compact form as:

$$z(k+1) = \mathscr{A}\cdot z(k) + \mathscr{B}\cdot u(k) \tag{8}$$

where

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

and $u(k) = (x(k+1), y(k+1), n(k+1))$. Note that the state of this system corresponds precisely to $z(k) = (x_{k-N}^k, y_{k-N}^k, n_{k-N}^k)$

Now, using the notation $z = (\mathbf{x},\mathbf{y},\mathbf{n})$ we define the privacy function $p : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ in the following way

$$p(z) = \frac{\sum_{i=1}^{N} \sqrt{(\mathbf{x}_i - x_c(z))^2 + (\mathbf{y}_i - y_c(z))^2}\cdot\mathbf{n}_i}{\sum_{i=1}^{N}\mathbf{n}_i}, \tag{9}$$

where

$$x_c(z) = \frac{\sum_{i=1}^{N}\mathbf{x}_i\cdot\mathbf{n}_i}{\sum_{i=1}^{N}\mathbf{n}_i}, \qquad y_c(z) = \frac{\sum_{i=1}^{N}\mathbf{y}_i\cdot\mathbf{n}_i}{\sum_{i=1}^{N}\mathbf{n}_i}. \tag{10}$$

To obtain optimal obfuscation in the sense of (5) we use a Model Predictive Control (MPC) structure presented in the following section.

3.2 Non-convex MPC problem

The key idea of this method is to use the knowledge of future locations over an horizon of H future steps, in order to compute an optimal sanitized location in each time. The procedure computes iteratively over $k \in \{0,..,M-H\}$ the optimal obfuscation, denoted by $(\bar{\delta}x(k), \bar{\delta}y(k)) = (\bar{x}(k) - x(k), \bar{y}(k) - y(k))$, where $(\bar{x}(k),\bar{y}(k))$ is the optimal position to transmit. We suppose that the utility loss is bounded in norm, that is, there is a $\Delta(k)$ such that

$$\|(\bar{\delta}x(k), \bar{\delta}y(k))\|_2^2 \leq \Delta(k)^2, \tag{11}$$

where $\Delta(k)$ is the maximum utility loss allowable. Equation (11) is equivalent to constraint in (5). We also denoted by $\bar{z}_H(k) = (\bar{\mathbf{x}}(k), \bar{\mathbf{y}}(k), \bar{\mathbf{n}}(k))$ the state variable introduced in the previous section but associated to the sanitized position.

For a given $k \in \{0,..,M-H\}$, the method uses the N previous obfuscated positions $(\bar{z}_H(k))$ and information on the H future steps $(x_k^{k+H}, y_k^{k+H}, n_k^{k+H})$. Eventually, the following optimization problem, named $P(\bar{z}_H(k), x_k^{k+H}, y_k^{k+H}, n_k^{k+H}, \Delta_k^{k+H})$, is solved

$$\max_{(\delta x_i,\delta y_i)_{i=1}^H \in \mathbb{R}^H \times \mathbb{R}^H} \quad \sum_{j=1}^{H} p(\tilde{z}(k+j))$$

$$\tilde{z}(k+i) = \mathscr{A}\tilde{z}(k+i-1) + \mathscr{B}\bar{u}(k+i-1),\ i \in \{1,..,H\},$$

$$\bar{u}(k+i-1) = \begin{pmatrix} x(k+i)+\delta x_i \\ y(k+i)+\delta y_i \\ n(k+i) \end{pmatrix}, \qquad i \in \{1,..,H\},$$

$$\delta x_i^2 + \delta y_i^2 \leq \Delta^2(k+i), \qquad i \in \{1,..,H\},$$

$$\tilde{z}(k) = \bar{z}_H(k),$$

It consists in minimizing the average of the future privacy, respecting a maximal utility loss. Here the obfuscation $(\delta x_i, \delta y_i)_{i=1}^H$ plays the role of controls. With this objective function, a same weight is given to each future privacy. A different choice could be made, for example giving different weights, or maximizing the last privacy $p(\tilde{z}(k+H))$.

After solving the problem $P(\bar{z}_H(k), x_k^{k+H}, y_k^{k+H}, n_k^{k+H}, \Delta_k^{k+H})$, we finally assign

$$\bar{\delta x}(k) = \delta x_1^*,\ \bar{\delta y}(k) = \delta y_1^*, \text{ and } \bar{z}_H(k+1) = \tilde{z}^*(k+1),$$

where δx_i^* and δy_i^* are the optimal solution and $\tilde{z}^*(k+1)$ is obtained using that optimal control. The procedure is repeated for $k+1$ until $M-H$. Note that we finish in $M-H$, as after this value it is impossible to use H future points.

4. EXPERIMENTAL EVALUATION

This section presents our evaluation framework, called p mpc−H when it uses an horizon of H steps, the competitor to whom we compare, and the real dataset used. Results highlighting the advantage of optimization are presented, and an analysis on computing overhead is given.

4.1 Experimental Setup

Experiments and analysis presented are replicable using our openly available Python code [1]. Location data are taken from the Cabspotting dataset, collecting real mobility traces of approximately 500 taxis collected over 30 days in the San Francisco Bay Area (Piorkowski et al., 2009). It consists of GPS coordinates (*lat*, *lng*) recorded at a non-constant sampling time (median is 30 s).

We compare our p mpc−H with the popular state-of-the-art protection mechanism Geo-I (Andrés et al., 2013). Geo-I protects user's positions by adding spatial noise:

$$\bar{l}(t) = l(t) - \frac{1}{\varepsilon}\left[W_{-1}\left(\frac{\alpha(t)-1}{e}\right)+1\right]\begin{pmatrix} cos\ \theta(t) \\ sin\ \theta(t) \end{pmatrix} \qquad (12)$$

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 $[0,\ 2\pi)$. Geo-I realizes the established differential privacy model Dwork (2006), hence the non-Gaussian distribution, enabling to derive mathematical privacy guarantees on the sanitized data.

In details, we evaluate the privacy and utility of several mobility traces: (i) p the actual user location, (ii) p obf locations sanitized using Geo-I ($\varepsilon = 5 \cdot 10^{-3}$), and (iii) p mpc−$H$ optimal obfuscation using our presented MPC, with a prediction horizon of H future location samples. Note that $H = 1$ means that we just use the current position, in other words, no future information is used to optimize the privacy.

In the following instance, we will consider $\tau = 5000s$ the length of the window of time to analyze, while the length of the mobile window is $T = 450s$. We discretize the time in intervals of $30s$. As real transmission is not necessary reported each $30s$, we re-sampled it to have it in that format. In order to fairly compare MPC and Geo-I methods, we use the same maximum utility loss in both cases. Thus, first we compute the obfuscation for each time using Geo-I, then, we compute the utility loss for that obfuscation (see figure 4) and finally we use those values as the upper bound for the utility loss in MPC method, i.e., we assign these values to the parameters $\Delta(k)$.

4.2 Gains on privacy with p mpc−H

Fig. 2 compares the actual positions and the sanitized ones obtained using GeoI and our MPC approach with $H = 15$. Mains differences can be observed next to the position $(3, 0.5)$, where the taxi spends a large time in the same position (there is few l points in that region). GeoI's random noise can be observed, while a more regular trace is computed by the MPC, enhancing user's protection by sending a fake realistic movement.

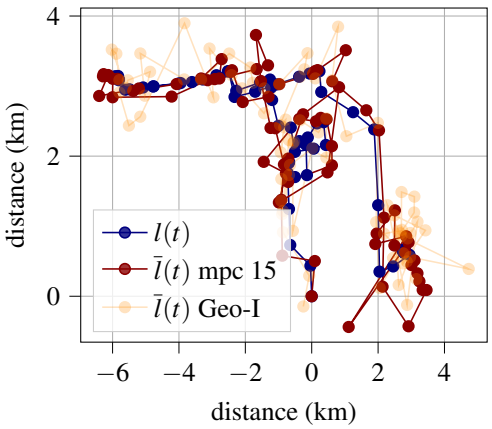

Fig. 2. Illustration of a mobility trace on a map, comparison of actual position and sanitized one p mpc-15. User oilrag.

Fig. 3 presents the comparison of privacy protection levels through time for a selected user trace using at each time the maximal utility loss obtained from Geo-I shown in Fig. 4. The average of that loss is around 206 meters.

p obf allows reaching a higher privacy than p, without protection. It is especially the case when p is low, i.e., user is in a privacy-sensitive situation, as after 4000 s. When the user has an inherently protecting move (p is high, for instance around 2000 s), p obf does not offer more privacy protection. p mpc−H reaches higher—sometimes similar—privacy levels than its competitors. Gains are particularly significant when prediction is valuable, i.e., when the user changes its mobility pattern. In Fig. 3 it corresponds to moments when p varies (e.g., from 1500 s to 2000 s or from 2500 s to 3000 s): a clear advantage for p mpc−H can be seen. When p is constant (e.g., from 550 s to 1500 s p mpc−H can also perform better that its

[1] https://gitlab.inria.fr/scerf/optimal-privacy

competitors. Overall, aggregated over the whole trace duration, the total gain of using p mpc−H over p obf taking $H = 15$ is of 14%.

4.3 Longer horizons H for better protection

p mpc−H better protects users locations thanks to future mobility prediction. For each method, we compute the average of the privacy values obtained at each time and we compute the gain with respect to the average of the real data privacy values p. Fig. 5 and Table 1 present privacy gain of using p mpc−H compared with p obf Geo-I for several horizons. The longer the horizon, the better the privacy protection. The privacy gain seems linear and increasing with respect to the horizon duration. In the most favorable scenario, with a horizon of 15 steps, the privacy gain is of +22%.

4.4 Limited overhead

The longer the prediction horizon, the better the protection. However, a long horizon duration has a computing cost in the p mpc optimization. We evaluate this overhead by collecting the distribution of execution time of the solver on one data point. In our experiments, we used a laptop Dell with Intel processor i5-8265U and CasADi optimization software in Python Andersson et al. (2019).

Results are collected in Table 2, with a horizon of 15 steps, the optimization takes 2.349 s on mean, that is about 25 times longer than with an horizon of 1. Moreover comparing the worst case (max time) the difference is even more important, being 32 times longer. While the overhead is non-negligible, it is however not significant when compared to the dataset sampling time of about 30 s. Therefore an horizon of 1 or 15 are both possible to be used in on-line cases.

H	(p mpc-p)/p	(p mpc-p obf)/p obf
1	0.189	0.103
2	0.190	0.104
5	0.195	0.109
10	0.206	0.120
15	0.220	0.133

Table 1. Performance of p mpc−H compare with real data pand Geo-I p obf

H	min time	mean time	max time
1	0.063 s	0.095 s	0.296 s
2	0.086 s	0.170 s	0.433 s
5	0.171 s	0.311 s	2.147 s
10	0.303 s	1.290 s	4.889 s
15	0.565 s	2.349 s	9.455 s

Table 2. Execution times of p mpc−H

5. CONCLUSION

We propose a new method, denoted $mpc - H$, to compute an optimal obfuscation using future information. This method is based in model predictive control tools and could be use for on-line and off-line instances. In our simulations we can see an improvement in the privacy using $mpc - H$, which is increasing with respect to H. This overcome by about 20% the real privacy, and about 12% the one obtained using Geo-I. The execution time reported (lower than 1 s in average for $H \leq 5$) tell us that an online implementation of the method is totally feasible. The next step of this work will be to explore analytically the performance of the proposed *MPC* scheme. In order to implement it in real time instances, we will study how to deal with the lack of information about the future. We think that a good strategy would be to mix the scheme with machine learning techniques which predict the future steps. Finally, the structure of the method let open opportunities to include other constraints that we might consider in a future work. For example, looking for an optimal obfuscated position preserving the user speed, useful in some services.

REFERENCES

Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., and Diehl, M. (2019). CasADi – A software framework for nonlinear optimization and optimal control. Mathematical Programming Computation, 11(1), 1–36. doi:10.1007/s12532-018-0139-4.

Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K., and Palamidessi, C. (2013). Geo-indistinguishability: Differential Privacy for Location-based Systems. In CCS, 901–914. ACM.

Bordenabe, N.E., Chatzikokolakis, K., and Palamidessi, C. (2014). Optimal geo-indistinguishable mechanisms for location privacy. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security, 251–262. ACM.

Cerf, S. (2019). Control Theory for Computing Systems: Application to big-data cloud services & location privacy protection. Theses, UNIVERSITÉ GRENOBLE ALPES.

Cerf, S., Robu, B., Marchand, N., Mokhtar, S.B., and Bouchenak, S. (2018). A control-theoretic approach for location privacy in mobile applications. In 2018 IEEE Conference on Control Technology and Applications (CCTA), 1488–1493. IEEE.

Chatzikokolakis, K., Palamidessi, C., and Stronati, M. (2014). A predictive differentially-private mechanism for mobility traces. In International Symposium on Privacy Enhancing Technologies Symposium, 21–41. Springer.

Chatzikokolakis, K., Palamidessi, C., and Stronati, M. (2015). Constructing elastic distinguishability metrics for location privacy. In PETS, volume 2015, 156–170.

Dwork, C. (2006). Differential Privacy. In Automata, Languages and Programming, volume 4052 of Lecture Notes in Computer Science, 1–12. Springer Berlin Heidelberg.

Gambs, S., Killijian, M.O., and del Prado Cortez, M.N. (2011). Show Me How You Move and I Will Tell You Who You Are. Transactions on Data Privacy, 4(2), 103–126.

Gambs, S., Killijian, M.O., and del Prado Cortez, M.N. (2012). Next place prediction using mobility markov chains. In Proceedings of the First Workshop on Measurement, Privacy, and Mobility, 3. ACM.

Google Play (2022). Travel & Local - Android Apps on Google.

Jiang, H., Li, J., Zhao, P., Zeng, F., Xiao, Z., and Iyengar, A. (2021). Location privacy-preserving mechanisms in location-based services: A comprehensive survey. ACM Computing Surveys, 54(1), 1–36.

Koufogiannis, F. and Pappas, G.J. (2016). Location-dependent privacy. In Decision and Control (CDC), 2016 IEEE 55th Conference on, 7586–7591. IEEE.

Krumm, J. (2007). Inference attacks on location tracks. In International Conference on Pervasive Computing, 127–143. Springer.

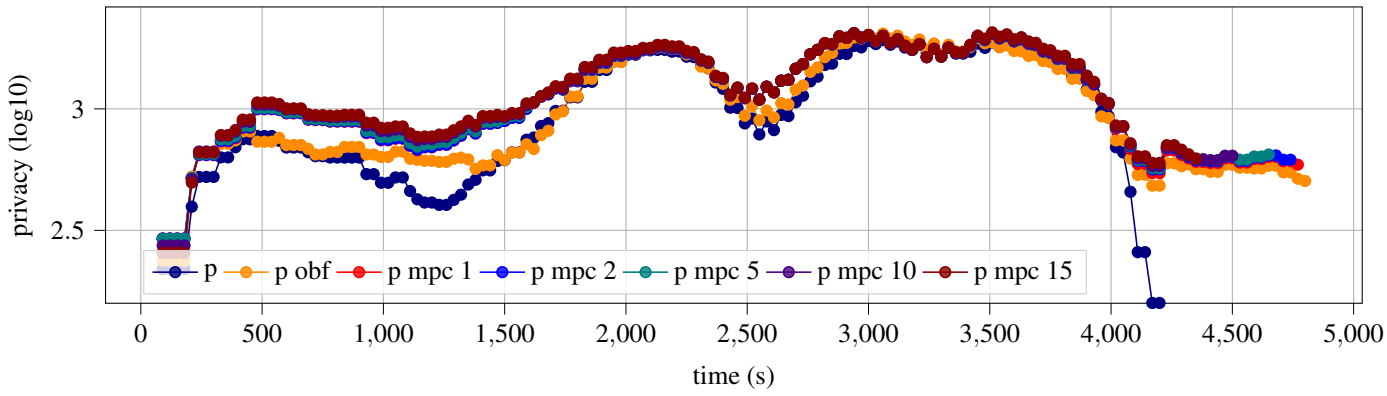

Fig. 3. Future prediction allows for privacy gains. Comparison of `p mpc-`H for different prediction horizons with competing protection algorithm, evolution of privacy through time. User `oilrag`.

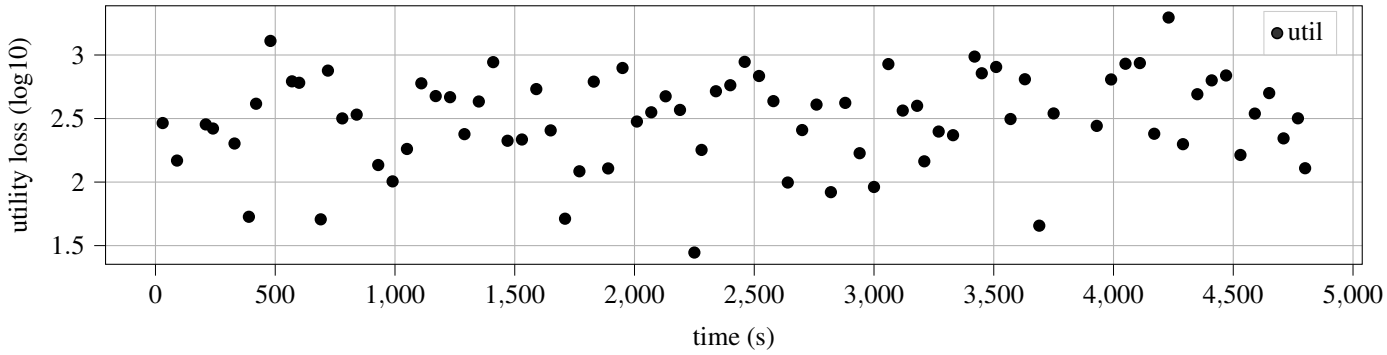

Fig. 4. Energy used each time for this instance.

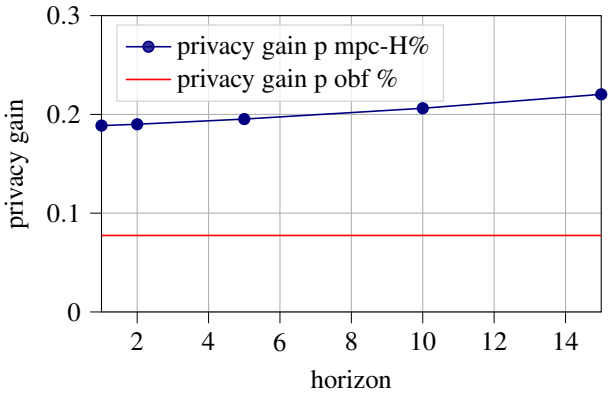

Fig. 5. Longer `p mpc-`H horizons reach higher privacy: gains in privacy according to horizon duration H. User `oilrag`.

McCarthy, N. (2023). The biggest GDPR fines of 2022. URL `https://www.eqs.com/compliance-blog/biggest-gdpr-fines/`.

Meira-Góes, R., Rawlings, B.C., Recker, N., Willett, G., and Lafortune, S. (2018). Demonstration of indoor location privacy enforcement using obfuscation. IFAC-PapersOnLine, 51(7), 145–151.

Oya, S., Troncoso, C., and Pérez-González, F. (2017). Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1959–1972.

Piorkowski, M., Sarafijanovic-Djukic, N., and Grossglauser, M. (2009). CRAWDAD dataset epfl/mobility (v. 2009-02-24). Downloaded from `http://crawdad.org/epfl/mobility/20090224`. doi:10.15783/C7J010.

Primault, V., Ben Mokhtar, S., Lauradoux, C., and Brunie, L. (2014). Differentially Private Location Privacy in Practice. In MoST'14. San Jose, US.

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

Shokri, R., Theodorakopoulos, G., Troncoso, C., Hubaux, J.P., and Le Boudec, J.Y. (2012). Protecting location privacy: optimal strategy against localization attacks. In Proceedings of the 2012 ACM conference on Computer and communications security, 617–627.

Xiao, Y. and Xiong, L. (2015). Protecting locations with differential privacy under temporal correlations. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 1298–1309. ACM.

Yavaş, G., Katsaros, D., Ulusoy, Ö., and Manolopoulos, Y. (2005). A data mining approach for location prediction in mobile environments. Data & Knowledge Engineering, 54(2), 121–146.

Yu, L., Liu, L., and Pu, C. (2017). Dynamic differential location privacy with personalized error bounds. In NDSS.

Zhang, W., Li, M., Tandon, R., and Li, H. (2018). Online location trace privacy: An information theoretic approach. IEEE Transactions on Information Forensics and Security, 14(1), 235–250.