Spatial Privacy Pricing: The Interplay between Privacy, Utility and Price in Geo-Marketplaces

Kien Nguyen
kien.nguyen@usc.edu
University of Southern California
Los Angeles, California

John Krumm
jckrumm@microsoft.com
Microsoft Research AI
Redmond, Washington

Cyrus Shahabi
shahabi@usc.edu
University of Southern California
Los Angeles, California

ABSTRACT

A geo-marketplace allows users to be paid for their location data. Users concerned about privacy may want to charge more for data that pinpoints their location accurately, but may charge less for data that is more vague. A buyer would prefer to minimize data costs, but may have to spend more to get the necessary level of accuracy. We call this interplay between privacy, utility, and price *spatial privacy pricing*. We formalize the issues mathematically with an example problem of a buyer deciding whether or not to open a restaurant by purchasing location data to determine if the potential number of customers is sufficient to open. The problem is expressed as a sequential decision making problem, where the buyer first makes a series of decisions about which data to buy and concludes with a decision about opening the restaurant or not. We present two algorithms to solve this problem, including experiments that show they perform better than baselines.

CCS CONCEPTS

• **Security and privacy** → **Economics of security and privacy**;
• **Information systems** → **Geographic information systems**;
• **Computing methodologies** → *Planning under uncertainty*.

KEYWORDS

geo-marketplace, location privacy, privacy pricing

ACM Reference Format:
Kien Nguyen, John Krumm, and Cyrus Shahabi. 2020. Spatial Privacy Pricing: The Interplay between Privacy, Utility and Price in Geo-Marketplaces. In *28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20), November 3–6, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3397536.3422213

1 INTRODUCTION

The current business model of many major technology companies is to provide free services to users in return for their data. Often the services and data are based on location. These freely acquired datasets are then used for various targeted advertising purposes, which in turn pay for the services developed and offered by the company. In addition, sometimes the companies simply sell the users' data to third parties and make a profit. These third parties may be interested in gathering information from certain locations. For example, public health authorities can use the data to identify potential pandemic clusters; city authorities are interested in travel patterns during heavy traffic; or advertisers are interested in the popularity of various locations at different times. The Location Privacy Protection Act of 2012 [1] tried to address this data collecting and sharing practice. It requires any company that obtains location information from a customer's smartphone to get that customer's express consent before collecting the location data, as well as before sharing the location data with third parties. However, the current practice is that if a user wants to hide her location data from a service provider, she has to turn off her location-detection device and (temporarily) unsubscribe from the service.

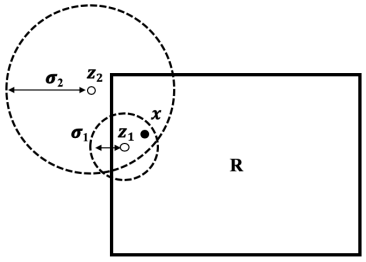


Figure 1: The true location x (black dot) can be sold as one of the noisy data points z_1, z_2 (white dots). The standard deviation σ_1 of noise of z_1 is smaller than that of z_2. Hence, z_1 is more expensive than z_2.

Recently, an alternative framework is emerging to create data marketplaces through which data owners offer their location data to potential buyers [10], dubbed a geo-marketplace [13]. Data marketplaces raise a number of interesting issues about data ownership, utility, pricing and privacy. Focusing on geo-marketplaces in this paper, we study the interplay between location data utility, privacy and value (i.e., pricing).

To illustrate, consider our running example scenario: a buyer is interested in checking if the number of people inside a target region R is large enough to, say, open a restaurant in R (utility). Determining whether the population is sufficiently large in a given area is also important for a number of decision making applications such as identifying pandemic hotspots, opening COVID-19 test centers, expanding public transportation, opening new community services (e.g. youth centers) and representative government.

Suppose the geo-marketplace has locations of many users, where each user's location can be sold at different levels of accuracy corresponding to her privacy vs. price preferences. There are different

SIGSPATIAL '20, November 3–6, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-8019-5/20/11…$15.00
https://doi.org/10.1145/3397536.3422213

approaches to capture privacy or inaccuracy. For our purpose, and to simplify the discussion, we assume privacy or inaccuracy is captured as a data point with a Gaussian noise level represented by the standard deviation of the noise distribution. [1] In Figure 1, for example, a user's true location at $\mathbf{x}$ can be sold as noisy points with mean and standard deviation (z_1, σ_1) for \$1 or (z_2, σ_2) for \$2.

The buyer can purchase data from multiple users and/or multiple times from one user at different levels of accuracy. After making a decision of "open" the restaurant or "cancel" (i.e. not open), the buyer would receive a net profit equal to the corresponding revenue of the decision *minus* the cost of purchasing data. The buyer's objective is to maximize this profit.

Maximizing the profit is challenging for the buyer for several reasons. First, the locations of users are uncertain, and the buyer can only reduce this uncertainty by purchasing more data. Second, the purchasing actions are irrevocable (i.e. the buyer cannot ask for reimbursement after already buying the data), so the purchasing action may not be optimal in hindsight. And third, although the problem can be modelled as a partially observable Markov decision process (POMDP), with locations as state and observations and possibly prices as actions, the large number of users and the continuous nature of the domain spaces render the standard POMDP solutions impractical due to the explosion of the number of states.

Some seemingly obvious solutions are not always effective. For example, buying the most accurate data from all users may exceed the payoff for making the correct open/cancel decision, resulting in a negative profit. Alternatively, spending a fixed, prespecified amount on purchasing raises the issue of predetermining that fixed amount. Therefore, there is a need for an adaptive approach to this problem. Other studies have attempted to find adaptive solutions to similar problems. However, they either considered a simpler version of this problem [16] or had different objectives [6, 8, 9, 14], thus, resulting in inapplicable solutions. Finally, the approaches for location privacy protection such as spatial cloaking [18], differential privacy [5] or Geo-Indistinguishablity [3] are relevant but orthogonal to our work as we discuss in Section 2.2. To the best of our knowledge, we are the first to consider the interplay between privacy, utility and price in a data marketplace, particularly in geo-marketplaces, with the focus on the profit of buyers.

We develop two adaptive algorithms to help buyers optimize the buying actions to obtain necessary data for a decision while striving to reduce the data acquisition cost, called the spatial information probing (SIP) algorithm and the SIP algorithm with terminals (SIP-T). Our algorithms take into account the uncertainty in the data, the irrevocability of the collection process and the large number of users' possible locations. Both algorithms start by buying data at a lower price (dubbed *probing*) in order to gather information, then continue to buy at higher prices the data points that have high potential to give high profit. These algorithms use the *expected incremental profit* (EIP), which intuitively is the expected increment of the expected profit when purchasing a data point at a price, to choose the next data point and the next price at which to buy. SIP-T enhances SIP by taking into account the distance of purchased data from the target region and focusing more on distinguishing whether a data point is inside or outside the target region.

Specifically, our contributions are:

- Proposing the problem of balancing the benefits for users and buyers in data marketplaces in terms of privacy, utility and price. In the spatial data marketplace, we called this problem *spatial privacy pricing*.
- Presenting a specific application in which buyers optimize their purchasing actions to obtain necessary data for a decision while trying to reduce the data acquisition cost.
- Developing adaptive algorithms which take into account multiple obstacles such as uncertainty of the data, the irrevocability of the collection process and the large range of possible locations of users. Our algorithms adaptively buy different data points at difference prices based on the EIP and the geometry of the purchased data.
- Extensive experiments comparing our algorithms to baselines over different settings for users' data, buyers' decisions and algorithmic parameters.

2 PROBLEM SETTING

In this section, we formalize the problem of the buyer deciding to *open* or *cancel* a restaurant in a target region R in order to demonstrate different aspects of the spatial privacy pricing problem: privacy, utility and value (i.e. pricing). We start by defining the notion of privacy valuation of users. We then introduce privacy pricing, which serves as the fundamental mechanism to balance the benefits for users and buyers. Then, we present the decision problem of the buyer, which involves purchasing data from users using the privacy pricing mechanism. Finally, we formalize our problem setting.

2.1 Users' Privacy Valuation

In a data marketplace, the privacy concern or valuation of users can be quantified with three components: their general privacy concern, their concern for a specific data point and their concern about a specific buyer. We present a simple model that captures these three aspects in a straightforward way.

For general privacy valuation, we assume each user u_i has an overall *privacy level* $\rho_i \geq 0$ which reflects the user's own valuation of their privacy and is independent between users. This reflects Westin's series of privacy surveys, where he categorized privacy concerns of people as high, moderate and low [11].

Each data point $\mathbf{x}_{i,j}$ of user u_i can have its own *sensitivity* $v_{i,j} \geq 0$, which is independent of ρ_i and independent between data points. Sensitivity $v_{i,j}$ reflects how sensitive u_i feels about $\mathbf{x}_{i,j}$, e.g. a gas station *vs.* a hospital.

For different buyers, users might also have a different level of *trust* ζ_b, e.g., an unknown developer may deserve less trust than the Centers for Disease Control and Prevention. In this work, since we start with a single buyer with a specific query, we consider $\zeta_b = 1$.

Subsequently, the total *privacy valuation* $\lambda_{i,j,b}$ of user u_i for their data $\mathbf{x}_{i,j}$ for a buyer b would be a function of ρ_i, $v_{i,j}$ and ζ_b

$$\lambda_{i,j,b} = \rho_i v_{i,j} \zeta_b \tag{1}$$

When the user has only one data point, we simply use $\lambda_i = \lambda_{i,j,b}$. For simplicity, we use λ_i in the remaining discussion. A user with

[1] This can be simply extended to use more sophisticated location privacy mechanisms such as Geo-Indistinguishablity [3].

a higher privacy valuation would expect to receive higher value for their location information. Next we discuss how λ_i affects the price of the users' location data.

2.2 Privacy Pricing

One popular approach to protect users' privacy when releasing data is by adding noise, either directly to the data or to some components of the data release process. In this work, Gaussian noise is added with the magnitude represented by the standard deviation σ dependent on the price paid. The noise magnitude or standard deviation σ is considered as the *noise level*.

When a data point $\mathbf{x}_i$ is traded at a price q_i, the noise magnitude σ_i is given as a function of q_i and λ_i as follows

$$\sigma_i = \begin{cases} \frac{\lambda_i}{q_i} k_\sigma & : \quad 0 < q_i < \lambda_i \\ 0 & : \quad \lambda_i < q_i \end{cases} \tag{2}$$

where k_σ is a scaling factor to make the resulting σ_i match with the real-world values of the noise magnitude, discussed in Section 5.4. This is one possible model that captures the essence of privacy and price. A higher price q_i reduces σ_i, while a higher privacy valuation λ_i increases σ_i. The privacy valuation λ_i can be viewed as the price the user u_i wants in order to sell their unperturbed data.

Although our users' pricing and privacy models need real-world evaluation, they capture the essence of our model and permit the full development and testing of our proposed algorithms. The functioning of the algorithms is unaffected by the particular pricing and privacy models. For example, other privacy preserving frameworks, such as spatial cloaking [18], differential privacy [5] or Geo-Indistinguishability [3], could also be used instead of the Gaussian noise. In that case, the parameters in those frameworks, e.g. ϵ in differential privacy, can be derived from a function $\epsilon = f_1(q_i)$ of the price q_i with a similar or more complicated pricing model. The noise level σ_i then can be derived from a function $\sigma_i = f_2(\epsilon)$ of those parameters. Eventually the noise level σ_i can still be considered as a function $\sigma_i = f(q_i) = f_1\big(f_2(q_i)\big)$ of the price. The work on specific privacy-preserving techniques or pricing models is orthogonal to our work. Consequently, in this presentation, we assume the privacy valuation of users and the pricing function are fixed, and we focus on the problem of the buyers maximizing their total reward (i.e. profit) with the strategy to make the final *open* or *cancel* decision described in the next section.

2.3 The Buyer's Profit Maximization Problem

In order to finally decide to *open* or *cancel* (i.e. not open) a restaurant in the target region R, the buyer can take a sequence of actions $\mathbf{a} = \{a_1, a_2, \ldots, a_T\}$, where taking an action a would give the buyer a reward or profit $r(a)$. The goal of the buyer is to maximize the expected total profit of the chosen actions

$$\max \mathbb{E}[r(\mathbf{a})] = \max \mathbb{E}[\sum_{t=1}^{T} r(a_t)] \tag{3}$$

where the expectation is over the (possibly) noisy data from users and the randomness of the process of choosing actions.

There are two types of actions that a_t can represent: open/cancel a^o actions and buying a^b actions. The buyer can take multiple buying actions to gather information before ultimately deciding to take an open/cancel action. Therefore, the last action a_T must be an open/cancel action a^o and other actions $\mathbf{a}_{1:T-1} = \{a_1, a_2, \ldots, a_{T-1}\}$ must be buying actions a^b.

An open/cancel action a^o can take one of the two possible values in the set $A^o = \{open, cancel\}$, which mean open or not open a restaurant, respectively. The profit of an opening action a^o is

$$r(a^o) = \begin{cases} \beta n - c & if \quad a^o = open \\ 0 & if \quad a^o = cancel \end{cases} \tag{4}$$

where β is the profit per user (or the gross margin), n is the number of users in the region R that may visit the restaurant and c is some fixed cost for opening the restaurant, for example, the monthly rent or the cost of operation. Naturally, more users leads to a higher profit. This form of the profit function in Equation 4 can capture other variations of the buyer's profit in our problem. For example, the buyer may decide to open only if $r(open) > r_0$, then c can be set as $c \leftarrow c + r_0$. In another example, when the buyer decides to cancel, one may consider the buyer losing a fraction $-\beta/k$ of the profit for each user inside R as some opportunity cost. This variation can be captured by setting $\beta \leftarrow (1 + \frac{1}{k})\beta$.

Each buying action a^b requires the buyer to purchase a data point $\mathbf{x}_i$ at some price $q > 0$. For clarity, we denote this buying action as $a^b(i, q)$. The set A^b of all possible buying actions includes all data points at all possible prices. The profit for this action $a^b(i, q)$ is the negative of the price the buyer needs to pay, which means

$$r(a^b(i, q)) = -q \tag{5}$$

Thus $r(\mathbf{a}_{1:T-1})$ can be considered as the cost of buying data.

In addition to $r(a^b(i, q))$, after taking $a^b(i, q)$, the buyer will also receive a noisy data point z_i whose coordinates are the true coordinate of $\mathbf{x}_i$ perturbed by independent Gaussian noise with σ_i derived from the price q using Equation 2, which means

$$z_i = \mathbf{x}_i + \eta, \quad \eta \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 I) \tag{6}$$

where I is the 2×2 identity matrix .

In this work, the number of actions T the buyer can take is not restricted. In addition, the buyer focuses on their business problem and is honest with the use of data. The buyer is also not restricted by any budget for purchasing data because an excessive cost of purchasing data will eventually decrease the net profit.

2.4 Formal Definition

For simplicity, we consider a snapshot in time of users' locations and assume each user has only one data point $\mathbf{x}_i$ at that time with sensitivity v_i. (Note that this same point could be sold multiple times at different levels of accuracy.) The setting would be similar to the case of multiple data points with minor changes. With all the quantities defined, we can formalize the problem setting as follows:

Given a snapshot in time of locations of N users where each user u_i is at location $\mathbf{x}_i \in \mathbb{R}^2$, has privacy valuation λ_i and is willing to trade $\mathbf{x}_i$ at different prices q_i, the buyer's objective is to maximize the expected total profit of an action sequence $\mathbf{a} = \{a_1, a_2, \ldots, a_T\}$ where $a_T \in A^o, a_i \in A^b \ \forall i = 1, .., T-1$:

$$\max \mathbb{E}[r(\mathbf{a})] = \max \mathbb{E}[\sum_{t=1}^{T} r(a_t)] \tag{7}$$

3 RELATED WORK

The concept of marketplaces for geosocial data was proposed in [10]. In [2], the authors investigated the value of spatial information to guide the purchases of the buyer. In [13], a geo-marketplace was proposed where location data are protected using searchable encryption. However, their setting only consider buying data points at full price, while in our setting one data point can be sold multiple times at different levels of accuracy for different prices. Singla [16] also studied the problem of maximizing the buyer's profit when information is available at a price, but also can only be purchased at full price, which can be considered as a simpler form of our problem and is used as a baseline in our experiments. In [9], Gupta *et. al* extends on the capability of purchasing the same data multiple times. However, they focus on maximizing the total profit, while the objective in our problem is to make a binary decision. In addition, although their method provides some theoretical guarantees of the performance, it relies on the *grades* of states of Markov decision processes that are infeasible to compute in our problem.

A partially observable Markov decision process (POMDP) is a powerful framework for modelling sequential decision making processes under uncertainty with the goal of maximizing total reward. POMDPs have been an active research area and considerable progress has been made on solving POMDPs [4, 15, 17]. While our problem setting can be considered as a POMDP with continuous state, observation and action spaces, our problem also includes a large number of users. Since the number of possible states in this POMDP increases exponentially with the number of users, standard POMDP algorithms are computationally infeasible for our problem.

There are also several studies on selling privacy [6, 8, 14]. However, they focus on the accuracy of the inferred number of users compared to the true number, while our focus is on the sufficiency of that number for making a binary decision which may incur some cost itself. In their setting, data points are also only purchased once, compared to possibly multiple times in our setting.

Many privacy-preserving techniques have been proposed for location privacy such as spatial cloaking [18], differential privacy [5, 20] or Geo-Indistinguishablity [3]. These techniques can be applied as the privacy protection mechanism for users' locations in our framework and a pricing model can be derived based on the parameters of these mechanisms, as discussed in Section 2.2. Therefore, while relevant, these techniques are orthogonal to our work.

4 METHODOLOGY

In this section, we first describe the strategy the buyer would employ to make the open/cancel decision. Given such an open/cancel strategy, we define the *expected incremental profit* (EIP) which serves as the basis for our spatial information probing techniques, the SIP and SIP-T algorithms.

4.1 The Buyer's Strategy to Make Open/Cancel Decision

Recall that each buying action $a^b(i, q)$ gives the buyer a noisy data point z_i. The noisy data obtained from the buying actions $\mathbf{a}_{1:T-1}$ can help the buyer to make the open/cancel action a_T as follows. Assume that after taking actions $\mathbf{a}_{1:T-1}$, the buyer owns a set of noisy data $\mathcal{Z}$. If there is more than one version of z_i from user i, the buyer keeps only the most accurate. Then, from the buyer's perspective, given z_i, the true location $\mathbf{x}_i$ has the distribution

$$\mathbf{x}_i \sim \mathcal{N}(z_i, \sigma_i^2 I) \tag{8}$$

The probability p_i of the user u_i being inside R is

$$p_i = \int_R P(\mathbf{x}_i) d\mathbf{x}_i \tag{9}$$

Considering each p_i as the success probability of an independent Bernoulli trial, the probability distribution $P(n|\mathcal{Z})$ of the number of users inside region R would be a Poisson binomial distribution [19] with mean as follows:

$$\mu_n = \sum_i p_i \tag{10}$$

Given this distribution, the current expected profit for open/cancel actions are $\mathbb{E}[r(open)] = \beta\mu_n - c$ and $\mathbb{E}[r(cancel)] = 0$. If the buyer decides to make an open/cancel decision at time T, i.e. a_T is an open/cancel action a^o, the best action a_T would be

$$a_T = \begin{cases} open & \text{if} \quad \beta\mu_n - c > 0 \\ cancel & \text{if} \quad \text{otherwise} \end{cases} \tag{11}$$

Assume that the buyer starts with an empty set of data, which gives $\mu_n = 0$. Then, $\mathbb{E}[r(open)] = -c$. Adding an additional p_i to μ_n in Equation 10 increases μ_n. Thus by buying more data, the estimate of μ_n increases, thus increasing $\mathbb{E}[r(open)]$. Therefore, the buyer would want to buy data to increase $\mathbb{E}[r(open)]$ while also trying to decrease the cost $\sum_{t=1}^{T-1} r(a_t)$. The buyer would stop buying when they can be confident of making the *open* decision, i.e. $\beta\mu_n - c > 0$ or when buying more data, in expectation, would not give more benefit. Thus, $\beta\mu_n - c > 0$ can be considered as the *open* trigger for the buyer. We call $\beta\mu_n - c > 0$ the *opening condition*.

4.2 The Expected Incremental Profit (EIP)

With the strategy to make the open/cancel decision established, we introduce the *expected incremental profit* $\text{EIP}(\mathcal{Z}, a)$ of taking a buying action a with the current set $\mathcal{Z}$ of purchased noisy data. $\text{EIP}(\mathcal{Z}, a)$ would then serve as the criteria for the buyer to choose a buying action in our algorithms.

To develop $\text{EIP}(\mathcal{Z}, a)$, we first calculate the expected profit if the buyer takes $a^o = open$ immediately given the current noisy data $\mathcal{Z}$

$$\mathbb{E}[r(open)|\mathcal{Z}] = \beta \sum_i p_i - c \tag{12}$$

If the buyer takes a buying action $a = a^b(i, q)$ and obtains a new noisy data point z_i' from the same user with the new probability p_i' of $\mathbf{x}_i \in R$, replacing the current noisy data z_i in $\mathcal{Z}$ with z_i', the expected profit of *open* is

$$\mathbb{E}[r(open)|\mathcal{Z}, z_i'] = \beta\big(p_i' + \sum_{j \neq i} p_j\big) - c - q \tag{13}$$

and the incremental profit is:

$$\text{IP}(\mathcal{Z}, a, z_i') = \mathbb{E}[r(open)|\mathcal{Z}, z_i'] - \mathbb{E}[r(open)|\mathcal{Z}] = \beta(p_i' - p_i) - q$$

If z_i does not exist in $\mathcal{Z}$ (i.e. $\mathbf{x}_i$ was never purchased before), $p_i = 0$.

Since p_i' is unknown before we actually perform the buying action $a = a^b(i, q)$, we can calculate the expected incremental profit

for taking the buying action $a = a^b(i, q)$ given the current noisy data $\mathcal{Z}$ as follows:

$$\begin{aligned}\text{EIP}(\mathcal{Z}, a) &= \mathbb{E}_{z_i'}[\text{IP}(\mathcal{Z}, a, z_i')] && (14)\\ &= \mathbb{E}_{z_i'}[\beta(p_i' - p_i) - q] && (15)\\ &= \beta\mathbb{E}_{z_i'}[p_i'|z_i, \sigma_i, q] - \beta p_i - q && (16)\\ &= \beta \int_{\mathbb{R}^2} P(z_i'|z_i, \sigma_i, q) p_i'(z_i') dz_i' - \beta p_i - q && (17)\end{aligned}$$

where $\mathbb{R}$ is the real domain and

$$\begin{aligned}p_i'(z_i') &= \int_R P(\mathbf{x}_i) d\mathbf{x}_i && (18)\\ \mathbf{x}_i &\sim \mathcal{N}(z_i', \sigma'^2_i I) && (19)\end{aligned}$$

The distribution $P(z_i'|z_i, \sigma_i, q)$ of the new noisy point z_i' if we takes action $a = a^b(i, q)$ is the convolution of two distributions: the distribution $\mathbf{x}_i|z_i, \sigma_i \sim \mathcal{N}(z_i, \sigma_i^2 I)$ of the true location $\mathbf{x}_i$ given the current noisy data z_i and the distribution $z_i'|\mathbf{x}_i, q \sim \mathcal{N}(\mathbf{0}, \sigma'^2_i I)$ of the next noisy data z_i' generated from the true location $\mathbf{x}_i$ where σ_i' is the noise magnitude for $\mathbf{x}_i$ at the price q. This means the distribution $P(z_i'|z_i, \sigma_i, q)$ is

$$z_i'|z_i, \sigma_i, q \sim \mathcal{N}\big(z_i, (\sigma_i^2 + \sigma'^2_i) I\big) \quad (20)$$

The buyer can then choose the next buying action as the action that maximizes the expected incremental profit:

$$a^*(\mathcal{Z}) = \underset{a}{\text{argmax}}\, \text{EIP}(\mathcal{Z}, a) \quad (21)$$

With $\text{EIP}(\mathcal{Z}, a)$, we develop two algorithms to help the buyer maximize their profit while maintaining a low purchasing cost.

4.3 The Spatial Information Probing Algorithms

We present two greedy algorithms that are based on the *probing* (or *information gathering*) technique. The general idea is to utilize the continuity of the price to obtain noisy data at low cost first in order to quickly eliminate uninteresting data before paying a higher price for more accurate data. The two algorithms are called *spatial information probing* (SIP) and SIP with terminals (SIP-T).

4.3.1 The SIP Algorithm. The pseudo-code for SIP is shown in Algorithm 1. SIP has two phases: the first phase is a *pure exploration* phase and the second phase is an *exploration-exploitation* phase. In the first phase, the buyer would buy all available data points in a large bounding box around the target region R at a small *starting price* q_0 (lines [2-9]). Then in the second phase (lines [10-30]), the buyer repeatedly calculates EIPs (lines [10-13]), takes a buying action based on the high potential EIPs (line [15, 27]) and uses the newly purchased noisy data, if any, to guide the next action (line [19, 29]). After each purchase, the buyer also checks the *opening condition* described in Section 4.1. The buyer would stop buying if the best EIP value is negative (line 16) which means that in expectation, buying more data would not increase the profit.

Algorithm 1 The SIP Algorithm

Input: Users $\mathcal{U}$; region R; β; c; starting price q_0; price increment factor h;
Output: *open* or *cancel*

1: $\mathcal{Z} \leftarrow \emptyset$; $V \leftarrow 0$; $\mathcal{E} \leftarrow \{\}$;
2: **for** $\mathbf{x}_i \in \mathcal{U}$ **do**
3: $z_i \leftarrow$ Buy $\mathbf{x}_i$ at price q_0
4: $\mathcal{Z} \leftarrow \mathcal{Z} \cup z_i$; $p_i \leftarrow P(\mathbf{x}_i \in R|z_i)$
5: $V \leftarrow V + \beta p_i$
6: **if** $V - c > 0$ **then**
7: **return** *open*
8: **end if**
9: **end for**
10: **for** $z_i \in \mathcal{Z}$ **do**
11: $a \leftarrow \underset{a}{\text{argmax}}\, \text{EIP}(z_i, a)$ (Algorithm 2)
12: $\mathcal{E}[a] \leftarrow \text{EIP}(z_i, a)$
13: **end for**
14: **while** true **do**
15: $a^b(i, q) \leftarrow \text{argmax}_a\, \mathcal{E}[a]$
16: **if** $\mathcal{E}[a^b(i, q)] \leq 0$ **then**
17: **return** *cancel*
18: **end if**
19: Delete $\mathcal{E}[a^b(i, q)]$
20: $z_i' \leftarrow$ Buy $\mathbf{x}_i$ at price q
21: $\mathcal{Z} \leftarrow (\mathcal{Z} \setminus z_i) \cup z_i'$
22: $p_i \leftarrow P(\mathbf{x}_i \in R|z_i)$
23: $p_i' \leftarrow P(\mathbf{x}_i \in R|z_i')$
24: $V \leftarrow V - \beta p_i + \beta p_i'$
25: **if** $V - c > 0$ **then**
26: **return** *open*
27: **end if**
28: $a \leftarrow \underset{a}{\text{argmax}}\, \text{EIP}(z_i, a)$ (Algorithm 2)
29: $\mathcal{E}[a] \leftarrow \text{EIP}(z_i, a)$
30: **end while**

Algorithm 2 Find The Best Next Buying Action

Input: z; λ_i; current price q; price increment factor h;
Output: Best next buying action a^*

1: $v \leftarrow -\infty$
2: **while** $q < \lambda_i$ **do**
3: $q \leftarrow \min(q \times h, \lambda_i)$
4: **if** $v < \text{EIP}(z, a^b(i, q))$ **then**
5: $a^* \leftarrow a^b(i, \lambda_i)$
6: $v \leftarrow \text{EIP}(z, a^*)$
7: **end if**
8: **end while**
return a^*

Because $\text{EIP}(\mathcal{Z}, a)$ only depends on z_i, not any other noisy points, we use $\text{EIP}(z_i, a) = \text{EIP}(\mathcal{Z}, a)$ as $a^b(i, q)$ clearly specifies z_i. Also, to reduce computation time, whenever the buyer calculates EIPs, the buyer keeps track of the best next buying action a for each noisy data point z_i and its corresponding value of EIP (line 12 and 29). This best next action is calculated using Algorithm 2.

Algorithm 2 chooses the best next buying action for a noisy data point z_i given the most recent purchased price q and a *price increment factor* h. Even though one can try to calculate the $\text{EIP}(z_i, a)$ for all possible prices in $[q, \lambda_i]$, it would be computationally impractical. In addition, the next price which gives the highest $\text{EIP}(z_i, a)$ may be too close to q to be informative enough (i.e. significant change in p_i) to make good progress. To overcome these obstacles, Algorithm 2 operates in multiple rounds and increases the potential price in each round by a factor h until the potential price is at least λ_i. The factor h allows the next purchase for $\mathbf{x}_i$ to be significantly more informative than the current z_i, besides reducing computation.

4.3.2 The SIP-T Algorithm. One issue with SIP is that, for $\mathbf{x}_i$ that is actually inside R but closer to the edges, the first low-price purchases may give unexpectedly low values for the probability of $\mathbf{x}_i$ being inside R, making the increment in the expected profit become smaller than the next price. This issue may cause the highest value of $\text{EIP}(z_i, a)$ to be negative, thus, SIP may stop buying earlier than expected. We propose the SIP-T algorithm to address this issue.

SIP-T defines a *terminal* belief (or terminal) τ_i for each data point $\mathbf{x}_i$ and only stops buying when either all data points are in their terminals or the *opening condition* is satisfied. The terminal τ_i specifies that the buyer can be certain about whether $\mathbf{x}_i$ is inside R or not. The *terminating* condition determines if z_i is at τ_i.

Although a threshold for p_i can be used for the terminating condition (e.g. $p_i < 0.05$), the high-magnitude noise when purchasing data at low prices make this approach ineffective. Instead, SIP-T uses the standard deviation σ_i of the noise to check for the terminating condition. More specifically, for a noisy data point z_i with the standard deviation σ_i of the noise, z_i is at τ_i if in each dimension z_i is either (1) inside R with the distance to each edge at least $k\sigma_i$ or (2) outside R with the distance to the closest edge at least $k\sigma_i$. We arbitrarily choose $k = 2$.

With the terminal condition defined as above, SIP-T is SIP with two changes. The first change is that, in line 16, it returns *cancel* when $\mathcal{E} = \emptyset$ instead of $\mathcal{E}[a^b(i, q)] \leq 0$. The second change is that after buying a data point and updating the expected profit in line 27, it checks for the terminating condition, and if the condition holds, it immediately continues to the next purchase round (line 14) instead of calculating next best buying action for that data point.

5 EXPERIMENTAL EVALUATION

We experimentally evaluate our probing algorithms on a real-world dataset with various settings for users' data, the buyer's decisions and algorithmic parameters.

5.1 Datasets

We experiment on the Gowalla dataset from the SNAP project [2] with 6,442,890 check-ins of 196,591 users over the period of February 2009 to October 2010. Each check-in includes a user id, check-in time, latitude and longitude.

We collect check-ins within a Los Angeles boundary defined by a bounding box from a southwest corner at (-118.684687, 33.699675) to northeast corner (-118.144458, 34.342324) in degrees of latitude and longitude. We then keep only one random check-in per user, which results in a subset of 5827 check-ins.

The radian (lat, long) coordinates are then converted to a locally planar Cartesian coordinate system with the mid-point of the latitude/longitude bounding box as the reference point. The bounding box in local Euclidean coordinates (x, y) is from the southwest corner at (-25,000, -35,000) to the the northeast corner at (25,000, 35,000) in meters. Figure 2 shows the check-ins in Los Angeles area within the bounding box.

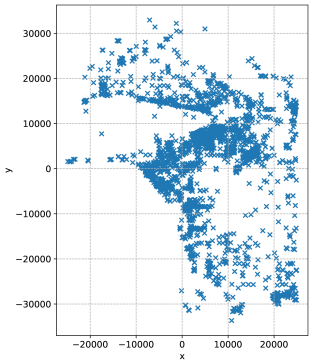


Figure 2: Check-ins of Gowalla users in Los Angeles converted to local Euclidean coordinates, 1 check-in per user.

5.2 Evaluation Metrics

To avoid any bias towards any position of the target region R, we evaluate performance of each algorithm over a collection of regions $\mathcal{R}$. Our collection of regions is a grid over the bounding box. Each cell of the grid is considered as a target region for making an open/cancel decision. The evaluation metrics includes the *Average Realized Profit* (ARP), the *Median Realized Profit* (MRP) and the *Recall* as explained in the following.

The ARP and MRP metrics. For each target region $R \in \mathcal{R}$, the net profit of the buyer's action sequence $\mathbf{a}_R$ is $r(\mathbf{a}_R)$, calculated using Equation 7. Then ARP and MRP are defined as the mean and median values of $\{r(\mathbf{a}_R) | R \in \mathcal{R}\}$. The reason we use MRP, in addition to ARP, is because there are often a few popular regions that have many users and become outliers in term of ARP since opening a restaurant there can yield extremely high profit compared to other regions. MRP is more robust to the effect of outliers and, as discussed later, can better reflect the cost spent by each algorithm.

Recall. Given the ground truth data, for each target region R, if the opening condition holds/does not hold for R, R is considered as a *positive/negative* region. Similarly, the value *open*/*cancel* of the open/cancel action a^o_R is considered as *positive/negative* decision. Consequently, recall can be calculated for the set $\{a^o_R | R \in \mathcal{R}\}$ as the ratio of the positive regions the algorithm can find.

5.3 Baselines

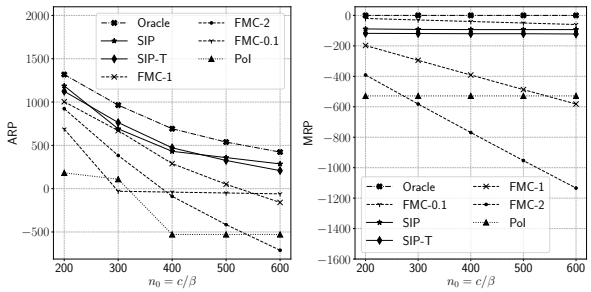
We compare our algorithms to several baselines. As mentioned in Section 3, since standard POMDP algorithms are computationally infeasible for our problem, we do not consider them as baselines.

[2] https://snap.stanford.edu/data/loc-gowalla.html

The first baseline is the *Oracle* algorithm which simply knows the most accurate location data of all users. Oracle is used as a benchmark to show the maximum value of each metric that any algorithm can achieve, even though this is unlikely since an algorithm usually needs to purchase data in order to make an open/cancel decision and the price of purchasing data may surpass the profit.

The second baseline is the utility maximization algorithm in [16], referred here as the *PoI algorithm*. With PoI, each point $\mathbf{x}_i$ is assigned a value (called a "grade"), which is the extra expected profit v_e such that the expected profit the buyer may earn by buying $\mathbf{x}_i$ at a certain price to obtain new noisy data is the same as the expected profit from stopping buying $\mathbf{x}_i$ and receiving v_e. The grade v_e is calculated based on the price of $\mathbf{x}_i$, the profit per user β and a uniform probability of $\mathbf{x}_i$ being inside the region R. PoI then ranks data points based on their grades and then repeatedly buys data points *at full price* until the maximum grade is non-positive or the opening condition holds.

The other baseline is the *Fixed Maximum Cost* (FMC) algorithm that spends a fixed amount to buy all data points at the same price. As mentioned, the challenge with the FMC is how to predict the fixed amount. We used 0.1%, 1% and 2% of the fixed cost c to derive three variations, called *FMC-0.1, FMC-1* and *FMC-2*, respectively.



(a) ARP (b) MRP

	200	300	400	500	600
FMC-0.1	0.17	0	0	0	0
FMC-1	0.83	1	0.67	1	1
FMC-2	0.83	1	0.67	1	1
PoI	0.17	0.25	0	0	0
SIP	0.83	0.5	0.33	0.5	1
SIP-T	0.83	1	0.67	1	1

(c) Recall

Figure 3: The effect of the minimum user threshold n_0

5.4 Parameter Setup

The gross margin per user β is set to $\beta \in \{50, \mathbf{100}, 150, 200\}$ in US dollars. The values are chosen based on the annual gross margin of some fast food chains such as Starbucks, McDonald's or Del Taco, divided by their total number of customers, which can be found in their annual reports. The bold value indicates the default value used in the experiments where this parameter is fixed.

Instead of fixing values for the fixed cost c, which is difficult to know in the real world, we consider the ratio $n_0 = c/\beta$ which can be seen as the minimum number of users that the buyer would need to *open*, because if $n > n_0$, then $\beta n > c$, which means the opening condition holds. We call this ratio the *minimum user threshold* and set it to $n_0 \in \{200, 300, \mathbf{400}, 500, 600\}$. These values are derived from [7] where the authors showed that the average number of restaurants was 2.3 (SD, 1.8) per 1,000 residents in Los Angeles County, yielding about 435 residents per restaurant. When β is fixed at 100, these values yield $c \in \{20{,}000, 30{,}000, 40{,}000, 50{,}000, 60{,}000\}$.

For users' privacy valuation, we simulate the privacy level and sensitivity values from two uniform distributions $\rho_i \sim Unif(0, d)$ and $v_{i,j} \sim Unif(0, d)$ then multiply them to get $\lambda_i = \rho_i v_{i,j}$. These distributions mean that each user has general privacy levels from 0 to d, and their own data points have sensitivity levels from 0 to d. The scale d is set to $d \in \{1, 2, \mathbf{3}, 4, 5\}$.

For the size of the regions R, we experimented with different $L \times L$ region sizes with $L \in \{2{,}500, \mathbf{5{,}000}, 7{,}500, 10{,}000\}$ meters, which creates from 35 to 560 regions in the test grid. These values are derived from the study [12].

For the parameters of SIP and SIP-T, the starting price q_0 is set to $q_0 \in \{0.0001, \mathbf{0.001}, 0.002, 0.003, 0.004\}$, and the price increment factor h is set to $h \in \{1.5, \mathbf{2}, 3.5, 5\}$ to reflect small and large increment factors. The scaling factor k_σ is fixed at $k_\sigma = 20$, which would render the standard deviation of noise to be $\{20, 40, 80, \dots\}$ for the ratio $\frac{\lambda_i}{q} = \{1, 0.5, 0.25, \dots\}$. These values are close to the noise magnitude of real-world location noise such as GPS, Wi-Fi or cell towers. Finally, the experiments for each region is executed in a single core of an Intel® Core™ i9-9980XE CPU.

5.5 Experimental Results

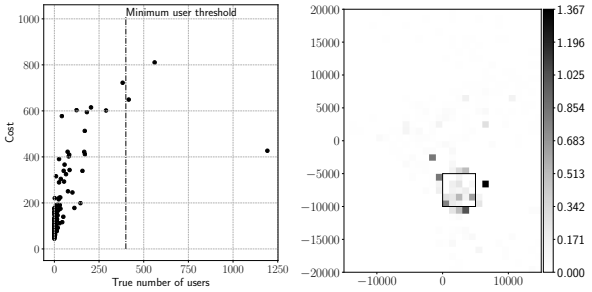
5.6 The effect of profit model parameters

5.6.1 The minimum customer threshold n_0. We first show in Figure 3 the effect of the minimum customer threshold n_0. SIP and SIP-T consistently outperform other algorithms in most cases.

The average realized profit (ARP) decreases when n_0 increases, since there are fewer regions that have enough users to open a restaurant (Figure 3(a)). However, while the ARP of other algorithms sharply decreases, SIP and SIP-T maintain their performance compared to Oracle. Note that Oracle is assumed to know the accurate data of all users. The actual cost to obtain such accurate data is about *1.8 million* USD. Besides, although ARP of *FMC-1* is sometimes similar to that of SIP and SIP-T, we emphasize that FMC comes with no principles for deciding how much to spend. Therefore, in some scenarios, it just got lucky. For example, FMC can easily suffer from underspending (such as *FMC-0.1*) or overspending (such as *FMC-2*).

The median realized profit (MRP) further demonstrates the superiority of SIP and SIP-T compared to other algorithms (Figure 3(b)). Since the decision for the majority of the regions in the grid should be *cancel*, Oracle has a zero MRP and other algorithms have negative MRPs, which are the median amount they spent on purchasing data. MRP values of SIP and SIP-T are stable and several times higher than those of other algorithms, except for *FMC-0.1* which was underspending.

While often spending much less than the other algorithms, SIP and SIP-T can still make correct *open* decisions for the regions where the buyer should decide *open*. This is shown in Table 3(c) where recall of SIP and SIP-T is comparable to *FMC-1* and *FMC-2* which spent significantly more.



(a) Amount spent per region (b) Amount spent per data point

Figure 4: Illustrations for the amount spent by the SIP

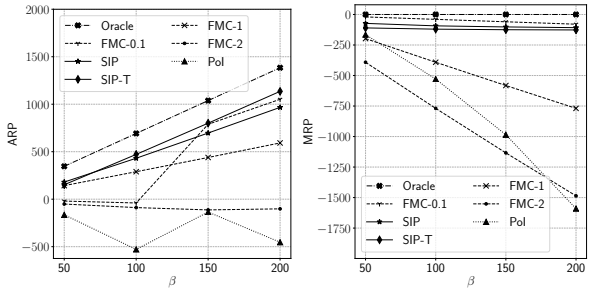
The reason for the superior performance of SIP and SIP-T are their highly adaptive nature: adaptive to the number of users in the target region and to the position of each data point relative to the target region. Figure 4 illustrates the amount SIP spent for each region against the true number of users (Figure 4(a)) and for each data point (Figure 4(b)). The amount spent tends to be higher for regions that have the true number of users closer to the minimum user threshold and higher for data points closer to the edges of a target region. This is because the true state being inside or outside a region is harder to identify for users closer to the edge, thus requiring more accurate data, which costs more. Although PoI dynamically decides which data point to buy, it does not perform well compared to SIP and SIP-T because it bought data at full price.

While SIP and SIP-T have comparable ARP, SIP has a higher MRP but a lower recall than that of SIP-T. This is because SIP-T continues to buy a data point at higher accuracy even though the EIP of the data point can be negative. Hence SIP-T would spend more than SIP but have more accurate location information.

The general trade-off is that spending more money buys more accurate data, which helps make more accurate open/cancel decisions. However, spending excessively may decrease the final profit, because the high cost of purchasing data may surpass the profit. SIP and SIP-T are two alternatives that balance this trade-off with different foci: on the expected profit per data point vs. distinguishing whether a data point is inside or outside the target region.

5.6.2 The profit per user β. The effect of the gross margin (or gross profit) per user β is shown in Figure 5. As β increases, the buyer can gain more per user, thus, gaining higher ARP. However, as minimum user threshold n_0 is fixed, when β increases, c also increases, thus, increasing the amount the FMC-based algorithm spends. That is why MRP decreases for the FMC-based algorithms.

With a higher value of β, the EIP of a data point at a price also increases. Increasing EIP leads to a higher chance that such data would be purchased at a higher price. This explains the slight decrease in MRP of SIP. Since SIP-T does not rely on EIP to stop buying a data point, its MRP does not change. Recall of SIP and SIP-T remain comparable to *FPC-1* and *FMC-2* while spent significant less. The result from the minimum customer threshold n_0 and the profit per user β show that SIP and SIP-T can give consistent and high results for different profit model parameters.



(a) ARP (b) MRP

	50	100	150	200
FMC-0.1	0	0	0.33	0.33
FMC-1	0.67	0.67	0.67	0.67
FMC-2	0.67	0.67	0.67	1
PoI	0	0	0.33	0.33
SIP	0.33	0.33	0.33	0.33
SIP-T	0.67	0.67	0.67	0.67

(c) Recall

Figure 5: The effect of the gross margin per user β

5.7 The effect of query parameters and user's data

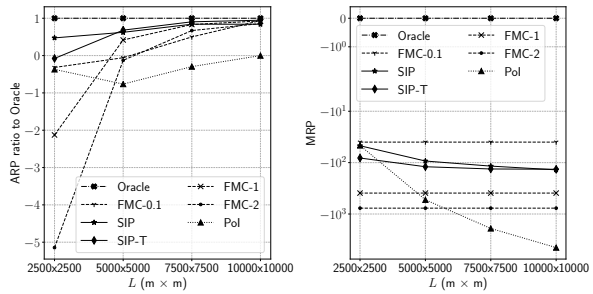
5.7.1 The target region L. Figure 6 shows the effect of the size L of the target region. Because the difference between values of ARP and MRP for different values of the size L is large, ARP is shown as the ratio to ARP of Oracle and MRP is shown on a logarithmic scale. With a larger region (i.e. a higher value of L), the buyer can gain more users, resulting in a higher profit in general. When the region size is very large (e.g. L = 10,000), all algorithms achieve comparable ARP. However, when L becomes smaller, SIP and SIP-T can maintain good performance while other methods suffer. Also, when L gets smaller, with a uniform probability, the probability of a data point being inside a region also becomes smaller. This makes PoI exclude many data points from buying, which may eventually not buy any data points and have a 0 recall. SIP and SIP-T, again, can maintain good ARP and recall while spent significantly less.

5.7.2 The scale d of users' privacy distributions. Figure 7 shows the effect of the scale d of the users' privacy distributions. In all metrics, performance tends to decrease when the scale increases, because the FMC-based algorithms would obtain noisier data for the same price, and SIP and SIP-T would need to spend more to obtain the same level of accuracy of a data point. For PoI, a lower value of d make it more likely to buy a data point, since the expected profit gain would be higher, hence, it would spend more to buy data, resulting in a lower MRP; and vice versa.

5.8 The effect of algorithmic parameters

5.8.1 The starting price q_0. Figure 8 shows the effect of the starting price q_0. Because it is not preferable to spend a large amount during the pure exploration phase, the starting price is set to small values.

While SIP-T uses the starting price, it eventually aims to distinguish whether a data point is inside or outside the target region.



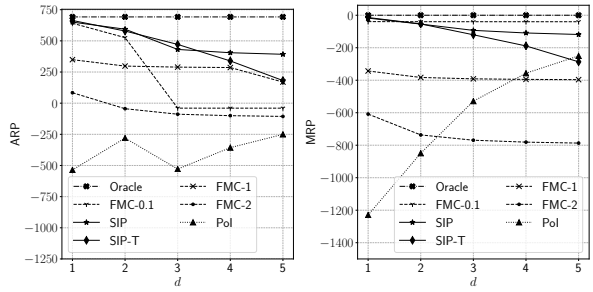
(a) ARP (b) MRP

	2500	5000	7500	10000
FMC-0.1	0	0	0.2	0.8
FMC-1	1	0.67	1	1
FMC-2	1	0.67	1	1
PoI	0	0	0.2	0.4
SIP	1	0.33	0.6	0.6
SIP-T	1	0.67	1	1

(c) Recall

Figure 6: The effect of the size L of the target region

Therefore, as long as the starting price is relatively small, the change of the starting price does not have significant effect to the performance of SIP-T. However, for SIP, there is a small change in ARP since with a higher starting price, it can obtain more accurate data to make more accurate open/cancel decisions. However, it would decrease its MRP, which reflects the total cost of buying data. On the other hand, a very small value of q_0 may result in too noisy data points at the beginning and can negatively effect the performance of SIP. Since SIP-T is more stable to the change of q_0, it is more preferable if one is willing to spend more budget for buying data.

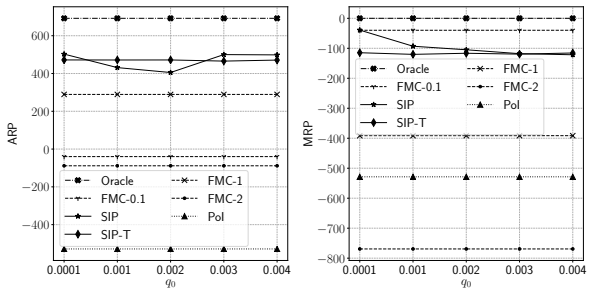


(a) ARP (b) MRP

	1	2	3	4	5
FMC-0.1	0.67	0.33	0	0	0
FMC-1	1	0.67	0.67	0.67	0.33
FMC-2	1	1	0.67	0.67	0.67
PoI	0.67	0.33	0	0	0
SIP	0.67	0.67	0.33	0.33	0.33
SIP-T	1	0.67	0.67	0.67	0.67

(c) Recall

Figure 7: The effect of the scale d of the privacy distributions



(a) ARP (b) MRP

FMC-0.1	0	0	0	0	0
FMC-1	0.67	0.67	0.67	0.67	0.67
FMC-2	0.67	0.67	0.67	0.67	0.67
PoI	0	0	0	0	0
SIP	0.33	0.33	0.33	0.67	0.67
SIP-T	0.67	0.67	0.67	0.67	0.67

(c) Recall

Figure 8: The effect of the starting price q_0

5.8.2 The price increment factor h. Figure 9 shows the effect of the price increment factor h. For both SIP and SIP-T, a higher value of h means that they would skip calculating EIP of a potential price more often. This can decrease ARP but probably for different reasons: SIP may incorrectly stop buying a data point, resulting in a higher MRP because it might spend less; on the other hand, SIP-T may purchase data points at a too high price, resulting in a lower MRP because it might spend more.

With a smaller value of h, SIP may spend more on uninformative purchases, because the standard deviation of the next purchase may not be different enough from the current purchase, thus, increasing the total amount spent (i.e. a lower MRP). While both SIP and SIP-T spend more, SIP-T performs better than SIP with a small value of the factor h, because it can make more accurate open/cancel decisions.

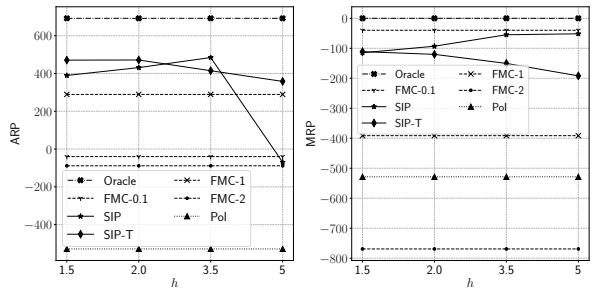
	1.5	2	3.5	5
SIP	265	124	54	38
SIP-T	163	100	59	48

Table 1: The average execution time (in seconds) of SIP and SIP-T for different values of the price increment factor h

A smaller value of h would also result in a longer execution time of SIP and SIP-T since there would be more potential prices for which they need to calculate EIPs. Table 1 shows the average execution time of SIP and SIP-T for different values of h.

6 DISCUSSIONS AND FUTURE WORK

To ease discussion, several simplifying assumptions were made for the buyer's profit maximization problem. For example, the privacy valuation was assumed to be known to the users or parameters of the profit model of the buyer were known; only a single snapshot of locations of users or only one buyer with one query was considered. Even with these simplifications, the buyer's profit maximization



(a) ARP (b) MRP

	1.5	2	3.5	5
FMC-0.1	0	0	0	0
FMC-1	0.67	0.67	0.67	0.67
FMC-2	0.67	0.67	0.67	0.67
PoI	0	0	0	0
SIP	0.33	0.33	0.33	0
SIP-T	0.67	0.67	0.67	0.67

(c) Recall

Figure 9: The effect of the price increment factor h

problem remains a challenging problem. For example, it can be seen as a particularly hard instance of POMDP. Future work should look at other scenarios that may illustrate new nuances to the central problem. In addition, while the proposed algorithms, SIP and SIP-T, consistently outperformed the baselines, other algorithms could also be explored to further optimize the buyer's decisions.

We also emphasize that the buyer's profit maximization problem is only one specific problem used to concretely demonstrate the broader problem: the spatial privacy pricing problem. For future work, we would explore other aspects of the spatial privacy pricing problem. For example, users' valuation of their data may change when they can observe selling and buying actions in a data marketplace, so we would explore the dynamic of users' valuation given the interactions in the marketplace. Other privacy preserving mechanism or other ways of ensuring privacy beyond adding noise (e.g. encryption) can also be employed. We would also study further the role of the marketplace or other types of queries of the buyers.

7 CONCLUSIONS

In a geo-marketplace, users can charge a price for their location data, and buyers can try to optimize their utility by making intelligent buying decisions. With *spatial privacy pricing* we introduce the element of privacy, where a user can charge different prices depending on how much a particular data point would reveal. We illustrate this interplay between privacy, utility and price with an example scenario of a buyer who is considering opening a restaurant. The restaurant will only be profitable if there are enough people nearby. With this example, we formalize the privacy and pricing considerations of the seller as well as the buying and utility considerations of the buyer. The buyer's reasoning is captured as an incremental expectation maximization algorithm that accounts, in a principled way, for the points' uncertainty, prices and the anticipated profit.

Our formulation results in a new geospatial problem for optimizing a buyer's decision-making process. Using our formula for "expected incremental profit", we introduced two related algorithms for specifying which location points a buyer should buy at which prices. The algorithms look for the next best point to buy. Compared with five baseline algorithms, our SIP and SIP-T algorithms are able to better adapt to the locations and prices of user data.

To the best of our knowledge, this is the first research that considers the privacy, utility and price of location data in a unified framework. This is an important step for creating a real geo-marketplace.

Acknowledgements: This research has been funded in part by NSF grants IIS-1910950 and CNS-2027794, the USC Integrated Media Systems Center, and unrestricted cash gifts from Microsoft. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the sponsors such as the National Science Foundation.

REFERENCES

[1] 112th Congress. 2012. Location Privacy Protection Act of 2012.
[2] Heba Aly, John Krumm, Gireeja Ranade, and Eric Horvitz. 2019. To Buy or Not to Buy: Computing Value of Spatiotemporal Information. *ACM TSAS* 5, 4 (2019), 1–25.
[3] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *the 2013 ACM SIGSAC CCS'13* (Berlin, Germany). ACM, New York, NY, USA, 901–914.
[4] Adrien Couëtoux, Jean-Baptiste Hoock, Nataliya Sokolovska, Olivier Teytaud, and Nicolas Bonnard. 2011. Continuous upper confidence trees. In *Learning and Intelligent Optimization (LION 2011)*. Springer, 433–445.
[5] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.
[6] Lisa K Fleischer and Yu-Han Lyu. 2012. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. 568–585.
[7] Lauren N Gase, Gabrielle Green, Christine Montes, and Tony Kuo. 2019. Understanding the Density and Distribution of Restaurants in Los Angeles County to Inform Local Public Health Practice. *Preventing chronic disease* 16 (2019).
[8] Arpita Ghosh and Aaron Roth. 2015. Selling privacy at auction. *Games and Economic Behavior* 91 (2015), 334–346.
[9] Anupam Gupta, Haotian Jiang, Ziv Scully, and Sahil Singla. 2019. The Markovian Price of Information. In *IPCO'19*. Springer, 233–246.
[10] Yaron Kanza and Hanan Samet. 2015. An online marketplace for geosocial data. In *Proceedings of the 23rd SIGSPATIAL'15*. 1–4.
[11] Ponnurangam Kumaraguru and Lorrie Faith Cranor. 2005. *Privacy Indexes: A Survey of Westin's Studies*. Technical Report CMU-ISRI-5-138. Carnegie Mellon University.
[12] John C Melaniphy. 2007. *The restaurant location guidebook: a comprehensive guide to selecting restaurant & quick service food locations*. International Real Estate Location Institute.
[13] Kien Nguyen, Gabriel Ghinita, Muhammad Naveed, and Cyrus Shahabi. 2019. A Privacy-Preserving, Accountable and Spam-Resilient Geo-Marketplace. In *ACM SIGSPATIAL'19* (Chicago, IL, USA). Chicago, IL, USA, 299–308.
[14] Kobbi Nissim, Salil Vadhan, and David Xiao. 2014. Redrawing the boundaries on purchasing data from privacy-sensitive individuals. In *ITCS'14*. 411–422.
[15] Konstantin M Seiler, Hanna Kurniawati, and Surya PN Singh. 2015. An online and approximate solver for POMDPs with continuous action space. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2290–2297.
[16] Sahil Singla. 2018. The price of information in combinatorial optimization. In *29th Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2523–2532.
[17] Zachary N Sunberg and Mykel J Kochenderfer. 2018. Online algorithms for POMDPs with continuous state, action, and observation spaces. In *ICAPS'18*.
[18] Idalides J Vergara-Laurens and Miguel A Labrador. 2011. Preserving privacy while reducing power consumption and information loss in lbs and participatory sensing applications. In *2011 IEEE GLOBECOM Workshops*. IEEE, 1247–1252.
[19] Yuan H Wang. 1993. On the number of successes in independent trials. *Statistica Sinica* (1993), 295–312.
[20] Yonghui Xiao, Li Xiong, Si Zhang, and Yang Cao. 2017. Loclok: Location cloaking with differential privacy via hidden markov model. *VLDB Endowment* 10, 12 (2017), 1901–1904.