A Privacy Mechanism for Mobile Commerce

Anil Kumar Chorppath and Tansu Alpcan
Technical University of Berlin
Deutsche Telekom Laboratories
10587 Berlin, Germany
Email: anil.chorppath@sec.t-labs.tu-berlin.de ,
alpcan@sec.t-labs.tu-berlin.de

***Abstract*—In mobile commerce, a company provides location based services to a set of mobile users. The users report to the company their location with a level of granularity to maintain a degree of anonymity, depending on their perceived risk, and receive in return monetary benefits or better services from the company. This paper formulates a quantitative model in which information theoretic metrics such as entropy, quantify the anonymity level of the users. The individual perceived risks of users and the benefits they obtain are considered to be linear functions of their chosen location information granularity. The interaction between the mobile commerce company and its users are investigated using mechanism design techniques as a privacy game. The user best responses and optimal strategies for the company are derived under budgetary constraints on incentives, which are provided to users in order to convince them to share their private information at the desired level of granularity.**

***Index Terms*—Game Theory; mobile commerce; privacy; mechanism design; information theoretic metrics**

I. Introduction

We consider a mobile commerce environment in which the users or customers get benefits from a company (service provider) by disclosing their location with certain degree of accuracy. At the same time, disclosing their location information brings users certain risks and compromises their privacy. Therefore, users have a motivation to maintain anonymity by giving less granular information about their location or no information at all. In this paper, we propose a *mechanism design* [1] approach in which the company acts as a designer and properly motivates its users through the benefits in terms of payment [2] provided to them, in order to obtain desired *granularity of location information* from all the users. We refer to the mechanisms in this setting as *privacy mechanisms.*

The benefits offered by the company to the users can be the location based service resources, discount coupons or monetary awards. It is assumed that the more accurate the information, the more valuable it is for the company. For example, street level information leads to contextual advertisements while city level granularity is less valuable. Concurrently, by being less anonymous, the users take a privacy risk. We take a commodity view of the privacy here, where the users can trade their privacy to obtain benefits from the company in an individual risk aware way.

Fair Information Practice Principle (FIPP) is the global standard that addresses consumer privacy risks. There are three main approaches [3] to implement FIPP: government regulation, self regulation by industry and Privacy Enhancing Technologies (PETs). There is tremendous amount of research on Privacy Enhancing Technologies (PETs) which try to preserve the privacy of users while giving targeted advertisements and services using personal data [4]. We consider our approach as complimentary to PETs, rather than as a substitute. The advertising and service provider industry is moving towards more self regulation which will enhance innovation and competition and ensure benefits for users in addition to safe guards provided by the government regulation [5]. The market based approach presented here models the incentive mechanisms behind this trend.

This paper presents an analytical model and a quantitative approach towards the risk-benefit trade-off of users and the goal of the companies. It uses metrics from information theory to quantify the anonymity level of users, concepts from game theory [6] to model the interaction of users among themselves and company, convex optimization techniques for solution and learning theory to learn the user risk level by the designer.

A. Literature Review

A wireless location privacy protecting system is analyzed and an information theoretic approach to define anonymity is proposed in [7]. In [8], the interaction between the local adversary deploying eavesdropping stations to track mobile users and mobile users deploying mix zones to protect their location privacy is studied using a game-theoretic model. MobiAd, a system for personalized, localized and targeted advertising on smart phones is proposed in [9]. Utilizing the rich set of information available on the phone, MobiAd presents the user with local advertisements in a privacy-preserving way by routing the information through a delay tolerant network. In this work they suggest the service provider to give discounts to motivate users to use MobiAd system. In [10] a proposal is made to provide users with rewards such as free "minutes" to motivate them to accept advertisements.

In [11], a game theoretic model of privacy in a community-based social networking mobile applications is proposed, in which the users take decisions on the level of granularity with which they share their location information to others. In that model, there is no service provider and the individual members of the community use their collective knowledge for personal or social goals. A Pareto improvement of the Nash

This work has been supported by Deutsche Telekom Laboratories.

978-1-4577-0351-5/11/$26.00 ©2011 IEEE

equilibrium is also achieved by making the users to contribute more information to the collective knowledge, using a tit-for-tat mechanism.

In this paper, we use an *information theoretic* approach [12] to quantify the *anonymity level* of the individual mobile users. The size of the crowd in which a user prefer to belong can be mapped to the desired anonymity level which can be further mapped to the granularity of location information. Therefore, the users have the power to make decisions on the level of granularity of location information reported to the service provider who gives them benefits based on that. An incentive or pricing mechanism is designed to achieve the company's goal of extracting the desired level of granularity of information. The company tries to move the Nash Equilibrium (NE) point vector of granularity of information in the underlying game to a desirable point as done in [13].

The next section presents the underlying system model and various parameters. Section III analyses privacy mechanism design problem and the solution. Then in Section IV a learning method for learning the risk factors of users by the designer is discussed. Numerical simulations and their results are shown in Section V. The paper concludes with remarks of Section VI.

II. Privacy Mechanism Model

Consider a mobile network composed of a set of mobile users with cardinality N. Around user i at any time t, let a group of $n_i(t)$ mobile users, $\mathcal{A}$, are in close proximity in an area. The service provider gives location based applications to the mobile users. Therefore, it asks for the location information from the mobiles.

We use an information theoretic approach to quantify the anonymity level of the individual mobile users while giving the location information. The uncertainty of service provider about the location information of user i is defined using the entropy term

$$A_i = \sum_{i=1}^{n_i(t)} p_i \log_2 \frac{1}{p_i}.$$

where probability p_i corresponds to the probability that a user is in a location. The parameter A_i concurrently quantifies the anonymity level of a users i. We can see that $p_i = \frac{1}{\log_2 n_i(t)}$. Then A_i simply boils down to,

$$A_i = \log_2 n_i(t).$$

We next define a metric called *granularity of location information*, g_i, for the i^{th} user as

$$g_i = 1 - \frac{A_i}{\log_2 N}.$$

The value of g_i is between zero and one for each user. The anonymity level obtained by user i by reporting with a granularity level g_i is

$$A_i = (1 - g_i) \ \log_2 N.$$

Here, $g_i = 0$ means the user i keeps its location completely private and $g_i = 1$ means the user gives exact location to the mobile company. We can see that the more the value of g, the less anonymous are the users. With a given value of g_i the users specify the size of the crowd it belongs to, i.e., $n_i(t)$. The Table I gives values of g for different combinations of $n_i(t)$ and N. We can see that as the size of the population N increases the more anonymous become the users.

TABLE I
VALUES OF $n_i(t)$, N AND g

$n_i(t)$	N	g
10^1	10^3	$\frac{2}{3}$
10^3	10^6	$\frac{1}{2}$
10^6	10^9	$\frac{1}{3}$

The users decide on the value of g which they report to the company. In the scenario considered in this model, the users have a continuous decision space resulting from a risk-benefit trade-off optimization, i.e. the allowed decisions are not just full or null information. This allows the designer to provide benefit based on the level of information given by the users.

There is a cost of perceived risk c_i associated with the user's privacy when they give location information, which linearly increases with the granularity of information, i.e,

$$c_i = r_i g_i,$$

where r_i is the risk factor. The risk factor may result from disclosing your daily routine or behavior to unknown parties. For example, the users may not like others to know when they are in their office or home or they may simply care about their privacy on principle. The users estimate or learn about their risk level from past experiences or from reliable sources or by exchanging information with users like how much level of g with which they report to the designer.

While gaining on location privacy, each user loses on the benefits of location based applications/services due to the anonymity. For example, while depending on whether users are in office, home or a particular street or city, they might be targeted with different kinds of offers and services. When they give wrong information they are given wrong services and offers. The total benefit obtained by user i can be quantified as

$$s_i = b_i(g) \log(1 + g_i),$$

where $b_i(g) \in R^+$ is the benefit or subsidy factor provided by the company. Note that the benefit factor b_i provided for user i is designed based on the granularity level chosen by all the users. In other words, the company provides benefits based on the total available information in the actual "information market". We model that the total benefit increases logarithmically with the granularity level, since for low granularity level marginal increase in the value of location information is higher. The logarithmic assumption in this paper can be generalized to any nondecreasing, concave function.

We now summarize the definitions of some of the terms

discussed so far.

1) **(Location) Privacy:** (Location) privacy of an individual user refers to how she discloses and controls the dissemination of her personal (location) data.
2) **Anonymity (location):** Anonymity of a user i, A_i, is the uncertainty of the service provider about the users location.
$$A_i = \sum_{i=1}^{n_i(t)} p_i \log_2 \frac{1}{p_i}.$$
3) **Granularity of Information:** Granularity of information is the level of granularity with which a user i reports its location.
$$g_i = 1 - \frac{A_i}{\log_2 N}.$$
4) **Perceived risk (cost):** It is the total cost perceived by user i as a result of reporting her location with a certain level of granularity of information, which is modeled as linear in g_i,
$$c_i = r_i g_i.$$
5) **Benefit:** The total subsidy or reward user i obtains from the mobile commerce company by disclosing her location with a certain level of granularity of information,
$$s_i = b_i(g) \log(1 + g_i).$$

In a mechanism design setting, there is a *designer* $\mathcal{D}$ at the center who influences N *players* participating in a **strategic (non cooperative) game**. Let us define the interaction of the users in the close proximity in the above setting as an N-player strategic game, $\mathcal{G}$, where each player $i \in \mathcal{A}$ has a respective **decision variable** g_i such that

$$g = [g_1, \ldots, g_N] \in \mathcal{X} \subset \mathbb{R}^N,$$

where $\mathcal{X}$ is the decision space of all players. The cost of each mobile user i will be the risk it perceives minus the benefits it obtains from the company, i.e.,

$$J_i(g) = r_i g_i - b_i \log(1 + g_i) \; \forall i.$$

Each mobile user then solves her own optimization problem

$$\min_{g_i} J_i(g). \tag{1}$$

Note that from the user perspective the benefit b_i is a constant designed by the company, since each user has an information constraint to know the granularity level of other users and calculate its benefit. The users just take best response given the benefit provided by the company.

The *Nash equilibrium* (NE) is a widely-accepted and useful solution concept in strategic games, where no player has an incentive to deviate from it while others play according to their NE strategies. The NE x^* of the game $\mathcal{G}$ is formally defined as

$$g_i^* := \arg\min_{g_i} J_i(g_i, g_{-i}^*), \quad \forall i \in \mathcal{A},$$

where $g_{-i}^* = [g_1^*, \ldots, g_{i-1}^*, g_{i+1}^*, \ldots, g_N^*]$. The NE is at the same time the intersection point of players' best responses obtained by solving user problems individually.

The company acts here as the mechanism designer and has the goal of obtaining a desired level of location information granularity from the users. In this work, the designer has an unconventional objective compared to other works in mechanism design where the designer usually looks for social welfare or designer revenue maximization. The designer or company here wants to improve the precision of location information from each user, which is captured by a designer objective function that takes granularity of information of all the users as its argument. The designer objective we consider here is,

$$\max_b V = \max_b \sum_{i=1}^{N} w_i \log(1 + g_i(b_i)), \tag{2}$$

subject to a budget or resource constraint

$$\sum_{i=1}^{N} b_i \leq B$$

where w_i's are the weights given to individual users as desired by the designer and B is the total budget. The weights depend on how much the company values the location information from different types of users.

It is important to note here that the designer (the mobile commerce company) tries to achieves its objective indirectly by providing benefits to users b as it naturally does not have control on their behavior, i.e. g. Essentially, the company tries to move the NE point vector of g of the resulting game to a desirable point by using the benefits provided to the users.

III. Privacy Mechanism

In a privacy mechanism, each user decides on the location privacy level to be reported, i.e., g_i, depending on its risk level perception as a best response to the benefit set by the company by minimizing individual cost. The underlying game may converge to a Nash equilibrium, which may not be desirable to the service provider because the required level of location information not obtained. Therefore, the designer employs a pricing or subsidy mechanism to motivate the users by properly selecting the benefits delivered to each user by solving a global objective.

The best response of the user i from the first order optimality condition of the convex optimization in equation (1) is

$$g_i = \begin{cases} 0, & \text{if } b_i \leq r_i \\ \frac{b_i}{r_i} - 1, & \text{if } r_i \leq b_i \leq 2r_i \\ 1, & \text{if } b_i \geq 2r_i \end{cases}$$

We can observe that the user reports her location with a nonzero granularity of information only when the subsidy factor is greater the risk factor. Also, the designer does not gain anything by giving the users a subsidy greater than twice their risk factor.

To solve the user problems and designer problem concurrently, we substitute the best response of all users given above in the designer objective in (2). Using these substitutions the

designer objective can be written in terms of the vector b and the designer problem becomes

$$\max_b V = \max_b \sum_i w_i \log(\frac{b_i}{r_i}), \tag{3}$$

subject to

$$\sum_i b_i \leq B$$

and

$$r_i \leq b_i \leq 2r_i \; \forall i. \tag{4}$$

The Lagrangian of this convex optimization problem is;

$$\begin{aligned} L &= \sum_i w_i \log(\frac{b_i}{r_i}) + \nu(\sum_i b_i - B) \\ &+ \sum_i \lambda_i(b_i - 2r_i) + \sum_i \mu_i(r_i - b_i), \end{aligned} \tag{5}$$

where ν, λ_i, μ_i are the unique Lagrange multipliers.

The resulting Karush-Kuhn-Tucker (KKT) conditions will give,

$$\frac{w_i}{b_i} = \nu + \lambda_i - \mu_i, \; \forall i \in \mathcal{A}, \tag{6}$$

and

$$\nu(\sum_i b_i - B) = 0,$$

$$\lambda_i(b_i - 2r_i) = 0, \; \forall i,$$

$$\mu_i(r_i - b_i) = 0, \; \forall i.$$

Since the individual concave utility functions are concave and non-decreasing, the optimum point will be a boundary solution. Therefore,

$$\sum_i b_i = B$$

and using the KKT condition in (6),

$$\sum_i \frac{w_i}{\nu + \lambda_i - \mu_i} = B \; \forall i \in \mathcal{A}. \tag{7}$$

We obtain the optimum benefit for each user as,

$$b_i^* = \frac{w_i}{\nu^* + \lambda_i^* - \mu_i^*}, \; \forall i \in \mathcal{A}, \tag{8}$$

where $\nu^*, \lambda_i^*, \mu_i^*$ are solution to (7). Then, the optimal granularity level of each user will be,

$$g_i = \begin{cases} 0, & \text{if } b_i \leq r_i \\ \dfrac{w_i}{(\nu^* + \lambda_i^* - \mu_i^*)r_i} - 1, & \text{if } r_i \leq b_i \leq 2r_i \\ 1, & \text{if } b_i \geq 2r_i. \end{cases}$$

If the solution is inner, i.e., constraints in (4) are satisfied with strict inequality and $\lambda_i = \mu_i = 0, \; \forall i$. We obtain

$$\nu = \frac{\sum_i w_i}{B}$$

and benefit for each user as

$$b_i = \frac{w_i B}{\sum_i w_i}.$$

Thus, the optimal granularity level of each user in the case of an inner solution is,

$$g_i = \frac{w_i B}{r_i \sum_i w_i} - 1 \; \forall i.$$

When all the users are perceived equally by the designer, i.e. $w_i = w_j \; \forall i, j$, the benefits are equally divided among them. In such a symmetric case,

$$b_i = \frac{B}{N},$$

and

$$g_i = \begin{cases} 0, & \text{if } b_i \leq r_i \\ \dfrac{B}{Nr_i} - 1, & \text{if } r_i \leq b_i \leq 2r_i \\ 1, & \text{if } b_i \geq 2r_i \end{cases} \tag{9}$$

The designer can obtain desired granularity of information from each user by properly selecting the functions in the global objective and the weights in the function. Note that to formulate the objective and for imposing the constraints on the global problem, the designer needs to know the user r's. This she can obtain using a learning method which will be considered next.

IV. Learning the Risk Factor

The designer can learn the risk factor from the best response of the users towards a sample subsidy factor vector b given by her to the users. We can see that from the best response of the users given in (1), the risk factor of user i is obtained as,

$$r_i = \begin{cases} \frac{b_i}{2}, & \text{if } r_i \leq \frac{b_i}{2} \\ \frac{b_i}{1+g_i^*}, & \text{if } \frac{b_i}{2} \leq r_i \leq b_i \\ b_i, & \text{if } r_i \geq b_i. \end{cases} \tag{10}$$

for any benefit b_i given by the designer and the best granularity level response g_i^* taken by her. If the value of the risk factor calculated from best response is given in the range,

$$\frac{b_i}{2} < r_i < b_i,$$

then it is the true value. If $r_i = \frac{b_i}{2}$, then the designer needs to reduce the benefit b_i given to the user i until $r_i > \frac{b_i}{2}$. Similarly, if $r_i = b_i$ then it needs to increase b_i until $r_i < b_i$. If the shape of the benefit part of the cost function is a general concave function unknown to the designer, it can employ an online regression learning algorithm [14] or an iterative algorithm, to estimate the function in each step.

V. simulation results

The privacy mechanism is illustrated with a numerical example here. We considered 5 users and their risk factors are randomly generated between 1 and 2. The risk vector in one instance is

$$r = [0.18\, 0.45\, 0.89\, 0.98\, 1.1693].$$

We first plotted the variation of the best response granularity level of the users with the total budget of the company. For

illustration purpose let the weights given to the users in the global objective is

$$w = [1.78\, 0.945\, 0.99\, 1.098\, 0.869]$$

and the company has no control over these weights to manipulate them. The company(designer) learns the value of the risk factor of users by giving a sample value of subsidies to different users and observing their best response as explained in Section IV. We could observe from Figure 1 that the company can extract more and more granularity of information by increasing the total budget, as expected. The threshold level of granularity for all the users which is the minimum level required to provide the service is taken to be 0.2. The critical level of budget required for extracting more than this threshold level of granularity from all the users, can be obtained from this plot. Here for this instance the critical level of budget is given as 6.

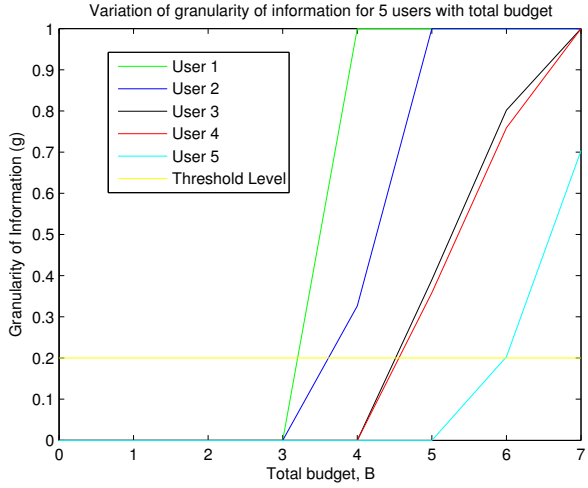

Fig. 1. Granularity of information of 5 users with the total budget.

Next, we consider the case where the company can adjust the weight given to different users in the global objective. In Figure 2, the setting remains as in the previous case except that the company varies the weight of the first user. From this plot, the weight required to get the desired level of granularity of information can be obtained. For user 1 the desired level of granularity of information is obtained with $w_1 = 0.21$.

As a future work, we plan to use the actual location data set to get the numerical results.

VI. Summary

This paper models and analyzes the interaction of a mobile commerce company with its users who obtain location based services, as a strategic game. A privacy mechanism is designed where the company motivates users to report their location information at a granularity level desired by the company. In return, the benefits obtained by a user depend on the weight the designer gives for her in the global objective. The users report their location with nonzero granularity of information when

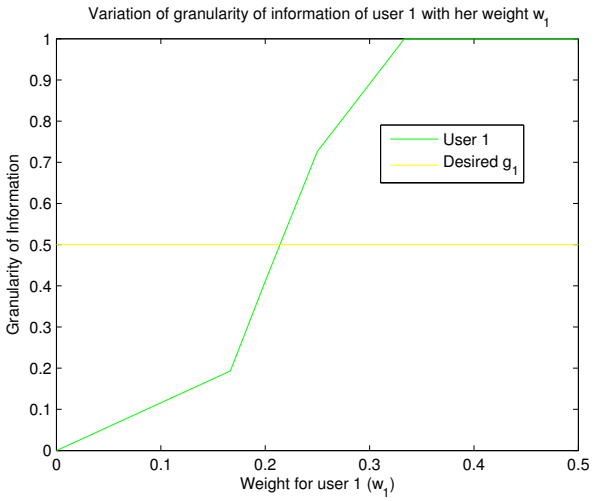

Fig. 2. Variation of granularity of information of user 1 with the weight in the global objective.

the subsidy by the company exceeds their perceived risk factor. The total budget required to obtain the desired minimum level of granularity of location information from all the users was obtained. As expected, the granularity of location information selected by the users decreases with the risk factor.

References

[1] V. Krishna, *Auction Theory*. Amazon: Academic Press 1st edition, 2002.

[2] R. Srikant, *The Mathematics of Internet Congestion Control*, ser. Systems & Control: Foundations & Applications. Boston, MA: Birkhauser, 2004.

[3] J. W. Bagby, Heng Xu, and T. R. Melonas, "Regulating privacy in wireless advertising messaging: FIPP compliance by policy vs. by design," in *The 9th Privacy Enhancing Technologies Symposium (PETS 2009)*, 2009, pp. 19–36.

[4] , "Privacy enhancing technologies symposium." [Online]. Available: http://petsymposium.org/

[5] , "Governments 'not ready' for new european privacy law," March 9 2011. [Online]. Available: http://www.bbc.co.uk/news/technology-12677534

[6] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. Philadelphia, PA: SIAM, 1999.

[7] L. Huang, H. Yamane, K. Matsuura, and K. Sezaki, "Towards modeling wireless location privacy," in *In Proceedings of Privacy Enhancing Technology(PET)*, 2005, pp. 59–77.

[8] M. Humbert, M. H. Manshaei, J. Freudiger, and J.-P. Hubaux, "Tracking games in mobile networks." in *GameSec'10*, 2010, pp. 38–57.

[9] P. Hui, T. Henderson, I. Brown, and H. Haddadi, *Targeted Advertising on the Handset: Privacy and Security Challenges*. Pervasive Advertising, Springer Human-Computer Interaction Series, 2011.

[10] M. Himmel, H. Rodriguez, N. Smith, and C. Spinac, "Method and system for schedule based advertising on a mobile phone," Patent, 2005.

[11] H. Liu, B. Krishnamachari, and M. Annavaram, "Game theoretic approach to location sharing with privacy in a community-based mobile safety application," in *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, ser. MSWiM '08. New York, NY, USA: ACM, 2008, pp. 229–238. [Online]. Available: http://doi.acm.org/10.1145/1454503.1454544

[12] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Proceedings of the 2nd international conference on Privacy enhancing technologies*, ser. PET'02. Berlin, Heidelberg: Springer-Verlag, 2003, pp. 41–53. [Online]. Available: http://portal.acm.org/citation.cfm?id=1765299.1765303

[13] T. Alpcan and L. Pavel, "Nash Equilibrium Design and Optimization," in *Proc. of Intl. Conf. on Game Theory for Networks (GameNets 2009)*, Istanbul, Turkey, May 2009.

[14] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/026218253X