Alexandra-Mihaela Olteanu*, Mathias Humbert, Kévin Huguenin, and Jean-Pierre Hubaux

# The (Co-)Location Sharing Game

**Abstract:** Most popular location-based social networks, such as Facebook and Foursquare, let their (mobile) users post location and co-location (involving other users) information. Such posts bring social benefits to the users who post them but also to their friends who view them. Yet, they also represent a severe threat to the users' privacy, as co-location information introduces interdependences between users. We propose the first game-theoretic framework for analyzing the strategic behaviors, in terms of information sharing, of users of OSNs. To design parametric utility functions that are representative of the users' actual preferences, we also conduct a survey of 250 Facebook users and use conjoint analysis to quantify the users' benefits of sharing vs. viewing (co)-location information and their preference for privacy vs. benefits. Our survey findings expose the fact that, among the users, there is a large variation, in terms of these preferences. We extensively evaluate our framework through data-driven numerical simulations. We study how users' individual preferences influence each other's decisions, we identify several factors that significantly affect these decisions (among which, the mobility data of the users), and we determine situations where dangerous patterns can emerge (e.g., a vicious circle of sharing, or an incentive to over-share) – even when the users share similar preferences.

**Keywords:** Location-based OSN; privacy; game theory

DOI Editor to enter DOI
Received ..; revised ..; accepted ...

## 1 Introduction

With the advent of mobile networking, mobile users can easily connect to the Internet and determine their actual locations with their smartphones, while on the go. Major online social network (OSN) providers, such as Facebook, understood early on the interest users have in sharing their location jointly, with their posts, pictures, etc. This location-sharing feature has gained even more momentum as users increasingly access their favorite OSNs from their smartphones (most Facebook check-ins and photos are made from mobile devices). Another popular feature, currently implemented in many mobile location-based social networks, is the ability to mention other users, such as friends, in posts or to tag them on pictures. Ilia et al. [22] perform a user study that demonstrates that 84.7% of posted pictures contain one or more face(s), whereas 87% contain one tag (users do not typically tag themselves) and 12.2% contain more than one tag. In many cases, such information indicates that the users mentioned in a post are co-located. As for location information, sharing co-location information – the fact that two users are together (the actual location might not be known) – brings social benefits (as also pointed out by Krasnova et al. [27]) to those sharing it but also to their friends who view it: Users enjoy knowing with whom their friends are and telling their friends with whom they are. Yet, these features also raise privacy concerns. Although it has been known for years that location information leads to severe privacy issues – this has been extensively studied in the literature (e.g., [13, 28, 39]; see also FindYou [51], a location privacy auditing tool, available at https://findyou.herokuapp.com/) – and location privacy risks have also been studied in the context of proximity detection (e.g., finding nearby friends in OSNs) [33, 43, 67, 68], it was only recently that the effect of co-location information on users' location privacy was studied [40]. A critical aspect of co-locations is that they relate to all the involved users (such information is co-owned by the involved users [17, 55]) and introduce interdependences between the users' location privacy, as the location information disclosed by users affects the privacy of their friends. As such, users lose partial control over their privacy and it becomes complex to evaluate the optimal sharing behavior. Such interdependent privacy risks are quite problematic if users have different, possibly opposite, views about sharing and privacy: it creates so-called multi-party privacy conflicts [55, 56]. Awareness about the interdependent nature of privacy is increasing, yet, due to its complexity, this is not explicitly addressed

***Corresponding Author: Alexandra-Mihaela Olteanu:** EPFL, E-mail: alexandramihaela.olteanu@epfl.ch
**Mathias Humbert:** Swiss Data Science Center, E-mail: mathias.humbert@epfl.ch
**Kévin Huguenin:** UNIL–HEC Lausanne, E-mail: kevin.huguenin@unil.ch
**Jean-Pierre Hubaux:** EPFL, E-mail: jean-pierre.hubaux@epfl.ch

BY-NC-ND

by current laws. Opinion 5/2009 on online social networking produced by the Working Party on Data Protection, which is an advisory board set up by the EU for the reform of the data protection laws, raises awareness about the case of users uploading data about others. Yet, even in the General Data Protection Regulation (GDPR) (Regulation EU 2016/679) which became enforceable on 25 May 2018, the case where individuals share data about individuals online is not directly mentioned, and the problem remains unsolved. Therefore, from a legal perspective, there are few regulations that apply to sharing on OSNs (except for the extreme case of sharing sexually explicit content, namely revenge porn) and this serious problem deserves further study.

We propose the first unified framework for modeling the direct and indirect benefits, and the privacy implications of location and co-location sharing, in addition to the resulting strategic behaviors of the users. Such a framework enables us to analyze the behavior of users regarding location and co-location sharing on OSNs. To this end, we build our framework by using two well-established modeling and analytical tools: game theory [16, 38, 58] and conjoint analysis [18]. Game theory enables us to model and formalize the users' sharing rationale and behavior. Such models include a number of parameters that, typically in the expression of the users' utility, characterize the users' behaviors. Conjoint analysis enables us to rigorously quantify, based on a personalized user survey, the relative benefits of sharing and viewing location and co-location information, and the associated relative costs in terms of location privacy. The values obtained through conjoint analysis are used to derive the different parameters of the game-theoretic model. Although several works [8, 44] have investigated interdependent privacy risks from a game-theoretic perspective (especially in the context of Facebook applications), this is the first work that investigates the strategic aspects of (co)-location sharing in the presence of interdependent privacy risks. Our framework could typically be used to gain insight into users' sharing behavior but also to design appropriate incentive mechanisms and location-sharing features. Our contributions are as follows. We identify the important problem of location sharing with interdependent privacy risks (introduced by co-location). And we propose the first game-theoretic framework to formalize it, namely the Sharing Game. Following a conjoint analysis approach, we design and conduct a user survey of Facebook users (N=250) to quantify users' preferences of (1) sharing or viewing posts, (2) location or co-location information, and (3) location privacy or sharing benefits. Our survey results indicate that, interestingly, there is no consensus regarding users' preferences; for instance, some users prefer sharing location information and others prefer sharing co-location information. We evaluate our analytical framework through simulations, in a number of key experimental setups and scenarios and on a real dataset, Geolife [70]. We use values of the parameters derived from the empirical data, avoiding the pitfalls of purely theoretical results, for a better understanding of realistic human behaviors. Our simulations unravel situations where users can be forced into a vicious circle of sharing their information or encouraged to over-share.

## 2 Related Work

Our work is related to two broad research areas.

**Information Sharing on OSNs.** Users share large amounts of information, including location, co-location and photos, with their friends on OSNs; this comes with privacy risks. Laufer et al. [31, 32] coined the term *privacy calculus*; it consists in a psychological framework formalizing users' decision making process through a cost-benefit analysis when sharing information. However, deciding whether to share information (and the precision at which the information is shared) is a complex process. It involves many factors including the users' contexts, the visibility of the shared information (i.e., who has access to it and the relationship between the user who shares the information and the users or the service providers who can access it [41, 49, 57, 62]), the shared information itself, and the benefits and privacy risks [60] associated with sharing. In some cases, the happiness of a user's friends also becomes part of the decision process; this is usually captured through a so-called *altruistic factor*, as introduced in [35] and experimentally measured using techniques based on conjoint analysis in [45, 46]. Conjoint analysis studies were also used to quantify the value that users attribute to their friends' information in the context of app adoption (e.g., in [47]). In practice, deciding whether to share information often comes down to finding a sweet spot between privacy and benefits [69]. The decision process can be automated by (1) maximizing privacy under benefits (service quality) constraints [53] (or conversely), (2) taking a game-theoretic approach for modeling the interplay between the users and the adversaries [52], or (3) by mimicking the users' sharing decisions using machine-learning techniques, after a training phase [9]. In our work, we model decision making as the optimization of a utility function that incorporates

both benefits and privacy. One of our contributions is to parametrize this function by applying conjoint analysis on user data collected through a targeted survey. Also, as users' decisions affect those of other users, we follow a game-theoretic approach for modeling the interplay between users and, ultimately, their decisions.

**Interdependent Privacy & Game Theory.** The notion of interdependent privacy, i.e., how actions performed by one user affect the privacy of another, was first formalized by Biczók and Chia [8]. Interdependent privacy raises the following concern: A user's privacy is no longer under her sole control. Numerous real-life examples of interdependent privacy risks were studied in the literature, including information about users' friends accessed by Facebook apps [8, 44], sensitive attributes inferred from those of a users' friends on OSNs [5, 14, 36], demographic information inferred from a user's interests [10], genomic data inferred from that of relatives [20, 21], location leaked from geotagged pictures that friends upload online [19], relationships inferred from pictures [54], and co-locations detected from the users' IP address at hotspots [59] or reported on OSNs [40]. From a social perspective, a large body of work has been devoted to the study of users' individual and collaborative coping mechanisms for multi-party privacy conflicts related to co-owned data (also referred to as regulation of interpersonal boundaries) [7, 11, 12, 23, 29, 55, 56, 63, 66]. These works focus mostly on the case of photo sharing on online social platforms and take an experimental and empirical approach to the problem,i.e., they rely on interviews and surveys. Misra et al. [37] propose a personal agent that recommends personalized access control decisions; Fogues et al. [15] propose a machine-learning-based policy recommender to predict the optimal sharing policy in multiuser scenarios. Game theory is a first class candidate tool for studying the interactions between users who are subject to interdependent privacy risks, as it enables the modeling of the effect of users' strategies on other users' utility, as well as the users' decision making process. It was successfully used to analyze users' application adoption behaviors [8, 44], the dynamics of individuals' privacy preferences regarding shared content [48], and privacy decision-making [2], such as sharing genomic data [21]. The study of interdependent privacy risks from an economic perspective follows the long line of research on interdependent security games surveyed in [30]. Our work is the first to study the interactions between OSN users in the case of (co-)location sharing, where shared co-locations create interdependent privacy risks. Unlike in the game-theoretic approaches surveyed above, in our framework we take into account the time dimension, future considerations, incomplete information, and an altruistic factor. In addition, we rely on a rigorous approach, based on user surveys, to determine realistic values of the different parameters of our model.

# 3 System Model & Formalization

We consider a mobile location-based online social network (OSN) with standard sharing features. Users are mobile and located within a given geographical region of interest (typically a city) and time is discrete. At some point in time, $t$, by checking-in at a given location, a user can post information about her location on her OSN profile. She can also post co-location information by tagging a close friend in a picture, or in a status update, thus making this information available to the OSN provider, all her friends and all her tagged friend's friends. In turn, a tagged user can "un-tag" herself from a post in which she is tagged, making this information unavailable to all users but not to the OSN provider. Sharing brings not only social benefits, but also *location privacy* implications, for both the user who shared the information and her tagged friend. At any time $t$, an adversary – either the service provider or the friends of one or both of these two users – has access to some of the previously reported locations and co-locations and can use this information to infer the users' locations at time $t$. We propose a framework in which, at any time, the decision to post (co-)location information, and the decision to allow a friend to post co-location information, is made strategically by both the users involved.

While users might act irrationally, especially when it comes to privacy-related decisions [4], this is still an active research topic. For instance, a recent result by Redmiles et al. [50] shows that users can actually make rational decisions in the context of adopting optional security behavior. We believe that privacy-protection demand will increase, notably because a growing number of people suffer the consequences of their (and others') carelessness. Furthermore, smartphones are increasingly involved in the sharing decisions users make, as demonstrated by the growing sophistication of the apps' permission systems. A tool run for this purpose *can* be 'rational' and strictly follow the parametrization provided by its user to aid him in decision making. It is therefore of interest to investigate what happens under the assumption of *rationality*.

## 3.1 User Model

We model the interactions between a user and one of her friends (also called players) as a game, called the *Sharing Game*, over a time window of interest ($\{1, \ldots, T\}$). The adversary, with respect to whom the users' privacy is evaluated (typically the service provider), is *not* a player of the game. We denote by $\mathbf{a}(t) = (a_i(t), a_j(t))$ the users' (denoted by $i$ and $j$) actual locations at time $t$. A potential strategy of user $i$ at time $t$ is denoted by $s_i(t)$ and $\mathbf{s}(t) \triangleq (s_i(t), s_j(t))$ denotes a strategy profile. $s_i(t)$ is chosen from the combinations of possibilities to share or not to share her own location and her possible co-location with her friend. We denote $s_i(t) \triangleq (sl_i(t), sc_i(t))$, where $sl_i(t)$ and $sc_i(t)$ are binary variables that represent whether user $i$ shares location and co-location, respectively. For alternate more compact notations, we use $\bar{L}$ for $sl_i(t) = 0$, $L$ for $sl_i(t) = 1$, $\bar{C}$ for $sc_i(t) = 0$ and $C$ for $sc_i(t) = 1$. When the two players are co-located, each of them can choose any combination of the four possible strategies: $\bar{L}\bar{C}$–sharing nothing, $\bar{L}C$–sharing only the co-location information, $L\bar{C}$–sharing only the location information or $LC$–sharing both. Yet, when the users are not co-located they can only choose whether to share their own location, choosing between two possible strategies: $\bar{L}\bar{C}$–sharing nothing and $L\bar{C}$–sharing location information. At each time in the windows of interest, both users choose their equilibria strategies–denoted by $\mathbf{s}^*(t) \triangleq (s_i^*(t), s_j^*(t))$. Information that the users share becomes available to an adversary: their actual locations at times at which they choose $L$ and the fact that they are co-located at times they choose $C$. For a time $t$, we denote by $\mathbf{o}(t-1)$ the information that the adversary observes up to time $t-1$ (we consider that the adversary observes information at $k \in \{0, \ldots, t-1\}$ time instants up to $t-1$). This set depends on both players' equilibria decisions up to time $t-1$ and is empty for $k = 0$ and $t = 1$. The information that the adversary observes at time $t$ depends on $\mathbf{s}(t)$.

User $i$'s social benefits that correspond to a strategy profile $\mathbf{s}(t)$ at time $t$ are denoted by $B_i(t, \mathbf{a}(t), \mathbf{s}(t))$. Note that the benefit function takes into account (i) the time $t$ to reflect the fact that check-ins at different times can have different meanings, (ii) both users' locations at time $t$ to reflect the fact that some locations can be more interesting to share or view than others (e.g., a hotel versus a park), (iii) who the other user is to reflect the fact that some co-locations can be more interesting to share than others, and (iv) both of the users' strategies, to reflect the benefit of sharing and that of viewing information shared by her friend. Her privacy at $t$, denoted by $P(i, t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j)$, is a function of (i) both users' actual location at $t$, (ii) the information observed by the adversary at the last $k$ time instants up to $t-1$–this depends on both users' strategies at those time instants, (iii) their strategy profile at $t$, and (iv) background user information (denoted by $\mathcal{B}_i, \mathcal{B}_j$), e.g., their mobility profiles. We emphasize that the privacy function takes into account previous time instants. In other words, a decision to disclose information at time $t$ has privacy implications at later time instants. Due to the dependency introduced by co-locations, the privacy function also takes into account decisions made by the other user, and the related background information.

Naturally, the information that a user has about the adversary's background knowledge of herself and of the other user, and her information about the other user's past or current locations, social benefits or privacy preferences, could be limited. These factors would influence her computation of her own privacy and of the other's privacy and social benefits. Some of these factors could be estimated (e.g., by completing surveys to compute their preference factors) and voluntarily shared among players (for instance through the service provider, in a private way). In the decision-making process, *players can be assisted by a tool for evaluating the privacy implications*, namely the value $P(\cdot)$, of each of the players' possible decisions regarding sharing. For instance, a Facebook client could compute this and suggest players' optimal decisions, using the information regarding users' locations and preferences in the computation of the game's equilibria. Another option is that such information about the friend is unknown and the players (i.e., their local tool) must estimate it or build probabilistic models of it. For the sake of keeping our model easy to understand, we consider the first option here; we also consider only immediate privacy implications in the users' estimation of the privacy (users were shown to often become "privacy myopic" and opt for *immediate gratification* in the context of their privacy decisions [1]); last, we assume players to be selfish. We present a model relaxing all of these assumptions in Section 6.

At any time instant $t$, a player's social benefits are computed as a normalized sum of the benefits of sharing information (i.e., location and co-location) and viewing information shared by her friend, specifically,

$$B_i(t, \mathbf{a}(t), \mathbf{s}(t)) = \frac{b^i_{sl}(\cdot) sl_i(t) + b^i_{sc}(\cdot) sc_i(t) + b^i_{vl}(\cdot) sl_j(t) + b^i_{vc}(\cdot) sc_j(t)}{b^i_{sl}(\cdot) + b^i_{sc}(\cdot) + b^i_{vl}(\cdot) + b^i_{vc}(\cdot)} \quad (1)$$

where $b^i_{sl}(\cdot)$ and $b^i_{sc}(\cdot)$ denote user $i$'s benefit of sharing location and co-location, and $b^i_{vl}(\cdot)$ and $b^i_{vc}(\cdot)$ her bene-

fit of viewing location and co-location. Note that these benefits take into account the parameters of $B_i(\cdot)$ and it is possible that they are correlated, e.g., if user $i$ has a large value of $b^i_{sl}(.)$, she might also have a large value of $b^i_{sc}(.)$. The utility of player $i$ for some strategy profile $\mathbf{s}(t)$ captures both her social benefits and her privacy.

$$U_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) = (1-\alpha_i)\cdot B_i\left(t, \mathbf{a}(t), \mathbf{s}(t)\right) + \alpha_i \cdot P\left(i, t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) \quad (2)$$

where $\alpha_i \in [0,1]$ denotes the weight with which user $i$ values her privacy over her social benefits. This formulation follows a privacy calculus approach [31, 32] under a pragmatic user model, as classified by Westin [61]. The choice of a linear model follows previous work (e.g., Acquisti [1]). The game is played successively, at time instants from 1 to $T$. At every time instant, we model the interactions as a perfect and complete information, non-cooperative extensive-form game. This type of game corresponds to the interactions in a typical OSN, where the players' actions at some instant are inherently sequential: The second player (or her application implementing the decision model) knows the choice of the first player and decides (or suggests to the player) her strategy accordingly. Therefore, we consider that the players' actions are ordered at every time instant. In reality, players would play such a game successively over time (reacting to each other's sharing actions). Note that this asymmetry between players can influence the outcome of the game.

We list our assumptions that model the existing OSNs' interfaces: (1) Location posts of a player are visible to all her friends *and* to the SP. (2) Co-location posts initiated by either of the players are visible to the SP and cannot be removed (even if the second player removes them, the service provider still has access to this information). (3) For a co-location post to be visible to friends of the two players, both of them have to agree to share it, in which case it is visible to the union of their friends. (4) If a player un-shares a co-location shared by the first player (by un-tagging or even asking it to be removed), the first player cannot share that co-location again. (5) Decisions made by the players are fixed. Once they strategically choose the best decisions at time $t$, they will not revisit them at later times.

## 3.2 Adversarial Models

Players' privacy always depends on the adversary: For the same strategy profile, different adversaries have access to all or only some of the shared information. We consider four possible adversaries, specifically the service provider and three different sets of users, essentially subsets of the players' friends. Note that these are all adversaries that our survey participants report being concerned about and we considered the adversaries and the information that is available to them for the typical default privacy settings for OSN posts.

**Service Provider Adversarial Model (SP).** The service provider adversary has access to all location and co-location posts made by the players. The specificity of this adversary is that, once either of the players shares information, this information is always known to him. In other words, the second player cannot un-share co-location information with respect to the service provider. We assume that the SP does not gather location information about its users through other channels, such as their IP address.

**Friends Adversarial Models (MF, FF, CF).** In these adversarial models, privacy is computed from the perspective of the players' friends. The common point of these models is that, unlike the SP model, the co-location information potentially shared by the first player can be removed by the second one (e.g., by un-tagging). We consider three different subsets of the friends, based on the information available to each of them, as illustrated in Figure 8: (i) "My other friends model" (MF) – this adversary has access to all the location posts made by the player and to co-location posts made by both players; (ii) "My friend's other friends model" (FF) – this adversary has access to all the location and co-location posts made by the other player and to co-location posts made by the player; and (iii) "Our friends in common model" (CF) – this adversary has access to all location and co-location posts made by both players. Note that the FF adversary can also be representative (with a possibly higher value for $\alpha_.$) for a public adversary–though this is not the default visibility for posts, users can post information with public visibility.

We emphasize that the action of un-tagging is not a strategy in our game. Yet, it is modelled in different ways: In the FF, CF and MF models, un-tagging is equivalent to the strategies that do not share co-location ($\bar{C}$) – these adversaries can no longer see the co-location; in the SP model, un-tagging has no effect – the SP has access to all the information shared by either player.

## 3.3 Analysis Methodology

At each time instant $t$, we use backward induction, a typical method for finding a subgame perfect Nash equilibrium (SPNE) that dictates the players' decisions. Observations made by the adversary at prior time instants,

stemming from the players' equilibria decisions, are used when computing the privacy of the players.

The first player, player $i$, anticipates the second player's (player $j$'s) best response, as a function of her possible strategies $s_i$, essentially

$$\forall s_i,\ s_j^*(s_i) = \arg\max_s\ U_j\left(t, \mathbf{a}(t), \mathbf{o}(t-1), (s_i, s), \mathcal{B}_i, \mathcal{B}_j\right)$$

This eliminates incredible outcomes that player $j$ would never rationally choose. Player $i$ chooses her best strategy out of the remaining outcomes, as follows

$$s_i^* = \arg\max_s\ U_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \left(s, s_j^*(s)\right), \mathcal{B}_i, \mathcal{B}_j\right)$$

The equilibrium decisions at time $t$ are given by

$$\mathbf{s}^*(t) = \left(s_i^*, s_j^*(s_i^*)\right)$$

We define social welfare, at time $t$, as the sum of the players' utilities, for any strategy profile, specifically

$$\begin{aligned} SW(t, \mathbf{s}^*(t)) \quad = \quad & U_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}^*(t), \mathcal{B}_i, \mathcal{B}_j\right) + \\ & U_j\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}^*(t), \mathcal{B}_i, \mathcal{B}_j\right) \quad (3) \end{aligned}$$

In the case of multiple equilibria at time $t$, the players coordinate and choose the one that maximizes their social welfare. The game is played in a similar way at successive time instants, each time taking into account the players' decisions from previous time instants.

We are interested in different properties for the players' equilibria decisions. Social optimality of the equilibrium at some time $t$ captures whether the decisions at equilibrium maximize the social welfare, an outcome that is desirable. A player's utility is maximized for the equilibrium decisions at $t$ if his utility at $t$ cannot be improved with a different strategy of either of the players. We consider both the proportion of time instants for which the equilibria decisions are socially optimal, and the proportion of time instants for which the equilibria decisions maximize each player's utility. Note that social optimality is defined only at time instants where both players play the game.

# 4 Survey

Our user model includes a number of parameters in the expression of the utility function that drives the users' strategic behaviors. As such, these parameters characterize the users' sharing behaviors; in practice, they vary from one user to another. In order to obtain realistic values for these, and to study the general trend and the variability across users, we conduct a survey of Fb users.

## 4.1 Methodology

Through the Amazon Mechanical Turk platform, we recruited participants with a Human Intelligence Task (HIT) approval rate of at least 95%, at least 100 past approved HITs and an active Facebook account. After the standard demographic questions (Part I), we polled the participants about their preferences regarding the posts they share or view on OSNs (Part II). The second part of the survey was composed of three questions to assess the participants' preferences regarding, respectively, (1) sharing vs. viewing posts with location information (i.e., check-in posts), (2) sharing posts with location information vs. sharing posts with co-location information, and (3) location privacy vs. benefits of sharing location information. We designed these three questions through a rigorous *full-profile conjoint analysis* approach [18] and making use of a dedicated tool ([64]). This approach enables us to quantify individual values for each of the participants' preferences factors.

**Sharing vs. Viewing ($f_{sv}$).** The participants were told that, for technical reasons, some of their two most-recent check-in posts and some of their friends' two most-recent check-in posts might be removed from Facebook. Then, the participants were asked to rank by preference a number of scenarios corresponding to different combinations of the numbers of posts kept (e.g., "two of *your* recent posts are kept and one of *your friend's* recent posts is kept", "none of your recent posts is kept and one of your friend's recent posts is kept"). The participants were asked to take into account only benefit considerations (i.e., not privacy). In order to limit the bias coming from the content of the posts, we explicitly mentioned that the posts to which we refer are posts they once shared and, hence, would like to keep, and we did not include the content of the participants' actual posts in the survey page. The initial ordering of these options was randomized. For this question, two attributes were used: the number of the participant's *own* kept check-in *posts* and the number of the participant's *friends'* kept check-in *posts*. Each attribute had three possible values (i.e., none, one or two). This yielded an optimal number of five options to rank (out of a total of nine). To detect sloppy answers, we included in the list of options a sixth option in which no posts are removed, and we explicitly stated in the text of the question that this should be the preferred option. The ranking provided by the users enabled us to compute their preference factors $0 \leq f_{sv} \leq 1$, from the importance values attributed to each attribute: $f_{sv}$ is the normalized importance value of the attribute *own posts*, whereas $1 - f_{sv}$ is the normalized importance value of the attribute *friends' posts*. A value greater than 0.5 denotes a preference for *sharing* over *viewing* information.

**Location vs. Co-location ($f_{lc}$).** This question was designed using the same methodology as for the first question: After a brief reminder about what a co-location post is (illustrated with screenshots), the participants were asked to order, according to their preferences, six options in which a number of their own recent posts with *location* information and a number of their own recent posts with *co-location* information would be removed (e.g., "two of your recent check-in posts are kept and one of your recent co-location posts is kept."). The ranking provided by the users enabled us to compute their preference factors $f_{lc}$, similarly to $f_{sv}$.

**Location Privacy vs. Sharing Benefits ($f_{pb}$).** After a brief reminder about location privacy, the participants were asked to order, according to preference, six options with different numbers of check-in posts and the corresponding levels of location-privacy, in terms of the average precision with which their location can be inferred during a day (e.g., "12 location posts for an average location privacy of 400 m"). These numbers were extracted from the experimental results presented in [40]. The ranking provided by the users enabled us to compute their preference factors $f_{pb}$, similarly to $f_{sv}$.

Finally (Part III), we polled the participants about their usage of Facebook, their privacy concerns, and about their knowledge of the privacy threats related to (co)-location information.

It took approximately ten minutes to complete the survey; the participants were paid $2. We ruled out the participants with inconsistent responses in Part II. More specifically, we considered as inconsistent a ranking that violates the natural order, i.e., considering that removing some of the existing posts is preferable to keeping them all. In the end, we obtained a sample of $N = 250$ valid participants; the sample was diverse and balanced in terms of the participants' demographics: 46% of the participants were female, the participants had various primary areas of employments, and their ages ranged from 19 to 68 years old, with an average of 33 and a standard deviation of 9.48. The participants were active Facebook users: 70% of the participants declared that they use Facebook multiple times per day (93% do so multiple times per week), 30% of them make at least one post with location information per week, and 37% of them make at least one post with co-location information (in statuses, in posts or in pictures) per week.

**Estimation of the Model's Parameters.** We estimate the parameters in our model ($\alpha$, $b_{sl}(t)$, $b_{sc}(t)$, $b_{vl}(t)$ and $b_{vc}(t)$) from the survey data. As we wanted to keep the number of questions low, we quantified only three preference factors $f_{pb}$, $f_{lc}$ and $f_{sv}$; to estimate the model's parameters from these, we make a few assumptions: We assume that (1) the users' preferences between sharing and viewing is the same for posts with location information as for posts with co-location information, (2) the users' preferences between posts with location information and posts with co-location information is the same for the users' own posts as for their friends' posts, (3) the users' benefits of sharing/viewing are the same over time. We derive the values of the model parameters as follows: $\alpha = f_{pb}$, $b_{sc}(t) = \frac{f_{sv}}{1-f_{sv}} b_{vc}(t)$, $b_{vl}(t) = \frac{f_{lc}}{1-f_{lc}} b_{vc}(t)$, $b_{sl}(t) = \frac{f_{sv}}{1-f_{sv}} \cdot \frac{f_{lc}}{1-f_{lc}} b_{vc}(t)$ where $b_{vc}(t)$ is a free variable (we set it to 1).

## 4.2 Results

We extracted the aforementioned three preference factors from the survey data by using XLSTAT. Note that, due to the fact that only a limited number of scenarios can be presented to the participants for ordering, the preference factors can take only a limited number of values. Table 2 in the Appendix presents relevant statistics (e.g., mean and standard deviation) and Figure 1 illustrates the CDFs of the derived preference factors. We observe that the average of the factors is close (yet slightly higher) than 0.5 (specifically, $.57 \pm .15$, $.56 \pm .15$ and $.60 \pm .39$ for $f_{sv}$, $f_{lc}$ and $f_{pb}$, respectively). This means that there is no strong consensus among the participants regarding their preferences. In fact, the distributions of the factor values are bi-modal: Users tend to have a clear preference for one of the two options (e.g., location vs. co-location). This phenomenon appears clearly for $f_{pb}$ (i.e., privacy vs. benefits) that has a high standard deviation (0.39). In the case of $f_{sv}$, for instance, the proportion of indifferent users (for whom $f_{sv} = 0.5$) is substantial (16.8%) and almost as large as the proportion of users who prefer viewing over sharing (23.2%). These results are in line with those of previous studies that showed that there exist multiple usage profiles on social networks: Some users connect to social networks mostly to share news with their friends whereas others do so mostly to view news about their friends [6, 42]. 54% of the users prefer location to co-location information ($f_{lc} > 0.5$) and 20% do not have a preference ($f_{lc} = 0.5$), whereas 63.2% favor privacy over social benefits ($f_{pb} > 0.5$).

As for the questions related to privacy issues on Facebook, 24.8% of the participants declared being "very concerned" about privacy, 50% declared being "moderately concerned" and 25.2% not concerned. When the participants report being co-located with a

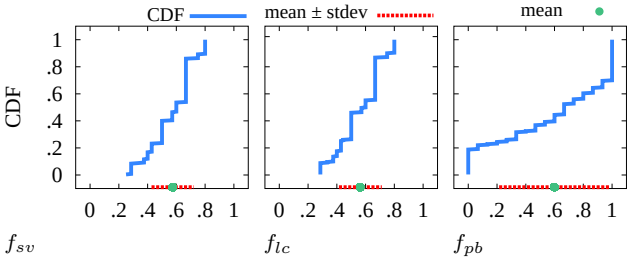


**Fig. 1.** CDFs of the preference factors of our survey participants.

friend (say Bob), their feared adversaries are Bob's friends who are not friends with the participant (i.e., the FF model, 44% of the participants), the common friends of Bob and the participant (i.e., the CF model, 24.4%), Facebook (i.e., the SP model, 24.4%) and the participants' friends who are not friends with Bob (i.e., the MF model, 21.2%); 26% of the participants reported not being concerned by any of these adversaries. 42.4% of the participants were not aware that their friends' posts that include location or co-location information can decrease their own location privacy. Only 50% of the participants declared being aware that their posts have privacy implications for themselves and for their friends, whereas 30.8% of the participants were not aware that their posts have any effect on privacy (as illustrated in Figure 7). Finally, we asked the participants whether the survey would affect their future sharing behavior on Facebook: A substantial fraction of the participants (around 35%) declared they would be more careful, especially for co-location information, for instance, by preventing their friends from tagging them in posts: e.g., "I may remove tags or ask friends not to tag me with locations in the future." (female, 35), "I may think twice before checking in, or at least consider the impact tagging others has on their privacy." (male, 31 y/o), "Yes because I was unaware of this issue and it now makes me a little scared." (male, 19 y/o). Of the participants who stated that their behavior would not change, 31% declared already being careful with their posts and tags. The full transcript of our survey, as well as an anonymized and sanitized version of the answers (part II and some of part III, password *FbS250*) are available at https://infoscience.epfl.ch/record/218755?&ln=en.

# 5 Evaluation

We evaluate our framework by simulating and analyzing the users' decisions in different experimental setups.

## 5.1 Quantification of the Users' Privacy

We quantify users' privacy ($P(\cdot)$) by relying on the inference framework proposed by Olteanu et al. [40]; we re-use the corresponding formalism and software library. There are, to the best of our knowledge, no reference scales for evaluating user perception of location privacy values expressed in meters. Note that our model is flexible enough to enable the use of other frameworks for inferring location privacy, for instance, that proposed by Xu et al. [65]. In short, we assume discrete locations (i.e., the geographical area of interest is partitioned into cells by using a regular square grid; when reporting their locations, users report the cells in which their actual locations fall; and the adversary has access to the users' mobility profiles in the form of transition probabilities between cells). Privacy is computed as the adversary's expected error when localizing users, using a junction tree exact-inference algorithm on the Bayesian network [25] that models the probabilistic dependencies between all the users' locations over the time period of interest. The location and co-location disclosures available to the adversary depend on the considered adversary, among those presented in Section 3.2, namely the SP, MF, FF, and CF models, and on the users' strategic decisions. For the sake of simplicity, we consider the same adversary for both users: For example, if the first user's location privacy is computed with respect to the OSN service provider, so is that of the other user. At time instances where a user's actual location is not known (sparse data), her privacy cannot be evaluated. At each other time instant $t$, the adversary considers *all* past location and co-location posts from the users when inferring their locations.

## 5.2 Scenarios

In order to evaluate our framework and to gain insight about the effects of the different parameters, we first consider the **canonical meeting scenario**, illustrated in Figure 9 of the appendix: Two users, Alice and Bob, coming from distinct locations ($t = 1$), meet for some time (one time unit, $t = 2$), and later separate in distinct directions ($t > 2$). We consider $T = 5$ time instants in total. At each time instant, both Alice and Bob can either report or hide their actual location. Additionally, at $t = 2$, either of them can choose to report being co-located with the other. Both users estimate the mobility profiles that an adversary would use in the inference process (i.e., $\mathcal{B}_i, \mathcal{B}_j$) by a very basic one: In one time unit, Alice/Bob either stays in the cell she/he is in (with probability .5) or moves to one of the neighboring cells

(with the remaining equal probabilities). We assume, for simplicity, a Manhattan-like model where users can move vertically or horizontally. Note that, other than from a velocity point of view (a user cannot move further than one cell in one time instant), this captures no real mobility information (all locations are equally probable). The rationale behind this choice is to understand the basics of the interplay between the users, independently from the specifics and the singularities of their individual data. The canonical scenario is, to some extent, also representative for an incomplete information game model, where users naively estimate the adversarial background information.

Additionally, we consider a **real-dataset scenario**, using the Geolife dataset [70] (collected in 2008) and the same co-location generation, user mobility profiles construction and space and time discretization as Olteanu et al. [40] (25 geographic regions covering the campus of Tsinghua University in Beijing and one-hour time subintervals splits of the continuous time interval). In this scenario, two users, Alice and Bob, follow their individual actual location traces (we consider sub-samples of $T = 300$ time instants from their full traces). At each time instant, both Alice and Bob can either report or hide their actual location (if this is known). Additionally, if co-located, they can also choose to report their co-location. In the privacy computation, mobility profiles constructed from real users' *full* location traces are used. Note that these are different and no longer uniform (among locations) and that they illustrate user-specific patterns of movement.

## 5.3 Experimental Results

In order to understand the effect of each of our model's parameters, we study, through simulations, the different strategic decisions players choose in several situations. We start with the canonical scenario, then move on to the real-dataset scenario.

### 5.3.1 The Effect of the Considered Privacy Adversary

In a first experiment, we consider a *homogeneous* canonical meeting scenario, where the parameters in both the users' utility are set using the average values of $f_{sv}$, $f_{lc}$ and $f_{pb}$ obtained in our survey, as presented in Table 2 in the Appendix. Figure 2 illustrates the different game outcomes, for the four adversarial models we presented in Section 3.2. A first observation is that the players' decisions are quite diverse, thus demonstrating that the adversarial model can influence what players share.

In the *SP* and *CF* models (Figures 2a and 2b), at $t = 1$ (when no co-location has yet been reported and thus there is no correlation between the users' locations or their privacy), the equilibrium decisions are that nothing be shared – the first blue rectangle and red circle pair. Note that for all time instants where users are not co-located ($t \neq 2$), the equilibrium decisions can only be "share nothing" or "share location". The equilibrium at $t = 1$ maximizes social welfare (there is a green triangle for $t = 1$), but either of the players would have a higher utility (both the blue rectangle and the red circle are empty) if the other one shared his own location (because, in the current absence of correlation, they would enjoy viewing where their friend is without any privacy cost to themselves). However, such an outcome is not an equilibrium because neither of them wants to share their location at this time (mainly due to the fact that the social benefit gained by sharing location would be less than their incurred privacy loss, weighted by $1 - \alpha$ and $\alpha$, respectively). At time $t = 2$, when the players are co-located, the additional benefit of sharing a co-location along with the benefit of sharing a location, overcomes the privacy loss; and the players' equilibrium decisions are that everything be shared $(LC, LC)$. This equilibrium not only maximizes social welfare, but also gives the best utility for both of the players at this time. Once these decisions to share have been made at $t = 2$, the privacy at $t = 3$ is already substantially compromised; hence the benefit of sharing location overcomes the (now) small relative privacy loss and both players choose to share everything, that is, their own locations. Similarly, the decision to share a location at $t = 3$ affects a player's privacy at $t = 4$ severely enough that they again decide to share their location (for the benefits) and this effect propagates at successive time instants.

In the *MF* model (Figure 2c), there is a different equilibrium at the time of co-location, $t = 2$. The outcome where both players share everything, $(LC, LC)$ is still the one that maximizes social welfare, but it is no longer an equilibrium because, to achieve better privacy, each of the players can now deviate from it by not sharing their own location. Hence utility (e.g., outcome $(LC, \bar{L}C)$ would be better for Bob than outcome $(LC, LC)$, because his adversary – his friends who are not Alice's friends – cannot see that Alice also shares her location). This was not the case in the *SP* model, where information shared by either player is automatically seen by the provider. In this case, the equilibrium is outcome $(\bar{L}C, \bar{L}C)$: Sharing only a co-location does come with a small privacy cost (privacy can decrease even when only co-location and no location information

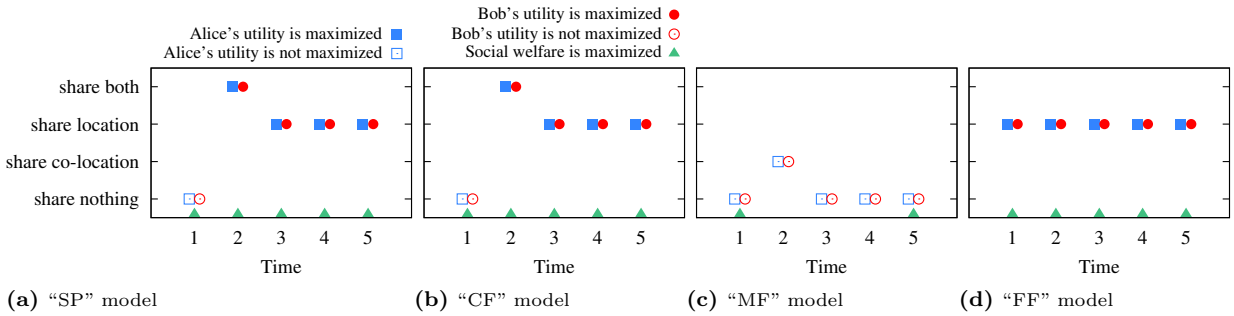


**Fig. 2.** Players' decisions at equilibrium, $\left(s^*_{Alice}(t), s^*_{Bob}(t)\right)$, for $f_{sv} = 0.57, f_{lc} = 0.56$, $f_{pb} = 0.60$ and different adversarial models: (a) Service Provider (SP), (b) Our friends in common (CF), (c) My other friends (MF), (d) My friend's other friends (FF) models. The $x$ axis shows the time window of interest. On the $y$ axis, for every time instant, Alice's decision is represented by a blue rectangle and Bob's decision by a red circle. A player's corresponding shape is full if its utility at equilibrium is maximized, and empty otherwise. Additionally, each time instant is marked by a green triangle, if the equilibrium decisions maximize social welfare.

is available due to the mobility profiles, as demonstrated in [40]), but this loss is smaller than the benefit gained by sharing. This equilibrium maximizes neither the social welfare nor a player's utility (either of them would have a better utility if the other would share their location, because they enjoy viewing where their friend is, at no privacy cost to themselves). At time $t = 3$, the players' privacy is higher than it was in the *SP* and *CF* models, for any strategy profile, because the decisions made at $t = 2$ provide the adversary with less information. Sharing the location is not justified because, in this case, the privacy cost this would bring is higher than the benefit gain, hence the equilibrium decisions are that nothing be shared. This equilibrium does not maximize players' utilities (each would still prefer to see the other's location at no privacy cost) or the social welfare. This effect is propagated over time, at successive time instants, and the equilibria decisions are the same: that nothing be shared. Furthermore, as the effect of the reported co-location at time $t = 2$ *fades away* over time, privacy increases, and at $t = 5$ the equilibrium also maximizes social welfare.

Finally, in the *FF* model (Figure 2d), the equilibrium at times when the players are not co-located is always $(L\bar{C}, L\bar{C})$: Sharing their own location brings them some social benefits without any privacy costs (this adversary cannot see if they share location). When players are co-located, the equilibrium is $(L\bar{C}, L\bar{C})$ and it maximizes both the social welfare and the players' utilities.

### 5.3.2 The Effect of Privacy vs. Benefits Preferences

We present a *heterogeneous* canonical meeting scenario, where players place different importance on privacy and social benefits. We consider the average values for $f_{sv}$ and $f_{lc}$ and vary $f_{pb}$ in $[0, 1]$. Figure 3 illustrates our results. We observe that, when players have different values for $f_{pb}$ (recall that $\alpha = f_{pb}$), their interests can be in conflict and their decisions at equilibrium might differ: When co-located ($t = 2$), one player might share only co-location, whereas the other shares both (e.g., in the $MF$ model when $\alpha_{Alice} = 0.6$ and $\alpha_{Bob} = 0.2$ Bob shares both, while Alice shares only co-location (recall that "share co-location" and "share both" decisions can only occur when the players are co-located, i.e., 20% of the times) or one shares his location, whereas the other shares nothing (e.g., in the $MF$ model when $\alpha_{Alice} = 1$ and $\alpha_{Bob} = 0.2$ Alice shares nothing, whereas Bob only shares his location).

An interesting observation is that, in the *SP* model, when the two players are co-located, the equilibria strategies are always in the form of $(\bar{LC}, \bar{LC})$, $(\bar{LC}, LC)$ or $(LC, LC)$. This stems from the fact that if *one* player wants to share the co-location information, as the service provider automatically has access to it, the privacy of the other player is already compromised and he is *forced into sharing* also but at least obtains the associated social benefits. This leads to equilibria in which one player's utility, or even the social welfare, is not maximized. Such outcomes can be avoided in the other models, where a player can undo the co-location shared by the other, and only equilibria with strategies where both players share or do not share the co-location information are permitted. An example can be observed in Figure 3, for $\alpha_{Alice} = 0.8$ and $\alpha_{Bob} = 0.2$: In the *SP* model, Alice is forced into sharing her location *and* co-location information at $t = 2$ because Bob, who places little importance on privacy, shares both, and the equilibrium is $(LC, LC)$; in the *CF* model, Alice does not al-

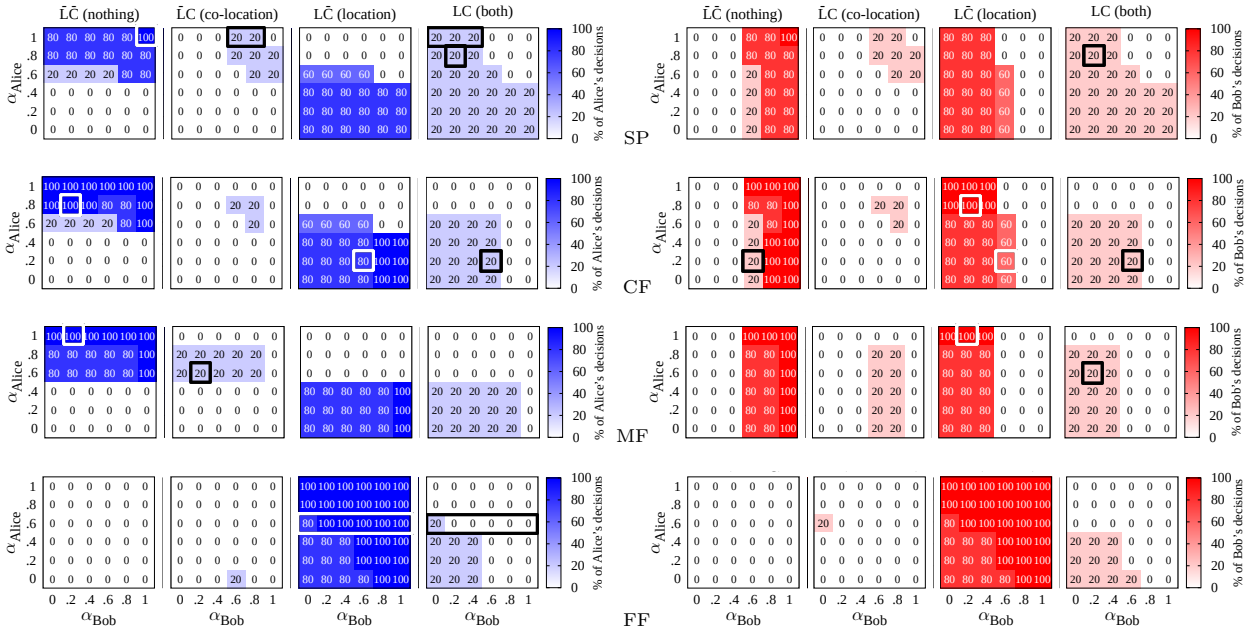


**Fig. 3.** Players' decisions at equilibrium, aggregated over time for $f_{sv} = 0.57, f_{lc} = 0.56$, and different adversarial models: Service Provider (SP)–first row, Our friends in common (CF)–second row, My other friends (MF)–third row, My friend's other friends (FF)–forth row models. For each adversarial model and each possible combination of values for $\alpha_{Alice}$ and $\alpha_{Bob}$, eight heatmaps (left four for Alice, right four for Bob) indicate the percentage of times, aggregated over the number of time instants, that a player made one of the four possible decisions: "share nothing", "share location", "share co-location" or "share both" (in all combinations $\alpha_{Alice}$-$\alpha_{Bob}$, the values of the four cells for a player sum to 100). We highlight with rectangles the cases that we discuss in Section 5.3.2.

low Bob to post co-location information about her and the equilibrium in this case becomes $(\bar{L}\bar{C}, L\bar{C})$: that Alice shares nothing while Bob only shares his location.

Another observation is that, **in all adversarial models, *both* players tend to share more as *one or both* their $\alpha$ decreases** (i.e., as one or both value privacy less). Notably, a player's strategy can change, even when only his friend's preferences change. For example, in the average case of $\alpha_{Alice} = 0.6$: As $\alpha_{Bob}$ decreases from 1 to 0, the amount of sharing Alice does increases (e.g., in the FF model, Alice only shares her location when $\alpha_{Bob} \in [0.2, 1]$, but she also shares the co-location when $\alpha_{Bob} = 0$). The same observation holds for the other values of $\alpha_{Alice}$. For the *SP* model, in particular, when Alice is very privacy conscious ($\alpha_{Alice} = 1$), her preferred outcome when co-located would be to share nothing, but she can only do this when $\alpha_{Bob} = 1$. She can gradually be forced into sharing her co-location with Bob (when $\alpha_{Bob} \in [0.6, 0.8]$) or even their co-location and her location (when $\alpha_{Bob} \leq 0.4$). Furthermore, the propagation of this effect can be observed not only at times where the players are co-located. Let us look, for example, at the case where $\alpha_{Alice} = 0.2$ and $\alpha_{Bob} = 0.6$: In the *CF* model, before his co-location with Alice (at $t = 1$ - a detail that is not directly readable form Figure 3, as it presents statistics aggregated over time instants), Bob decides to not share anything (20% of the times). Once co-located, Bob and Alice have enough incentive to share both their co-location and location (20% of the times). After their co-location, Alice still has incentive to share her location. Their previously reported co-location, as well as Alice's successive reports of her location, continue to damage Bob's privacy, and he counteracts these losses by also sharing his location for the benefits (60% of the times).

### 5.3.3 The Effects of Multiple Users' Preferences

We present a more realistic setup, based on the canonical meeting scenario. Each of the two players' parameters are assigned from the individual *preference profiles* of the survey participants. A preference profile represents the values of all preference factors $(f_{sv}, f_{lc}, f_{pb})$, for a specific participant; there are 250 such profiles. Analyzing the players' behaviors is substantially more complicated, due to the multiple influences present in such a complex setup. In order to find a meaningful interpretation, we alternatively split the 250 preference profiles into two subsets, based on the value of one of the preference factors.

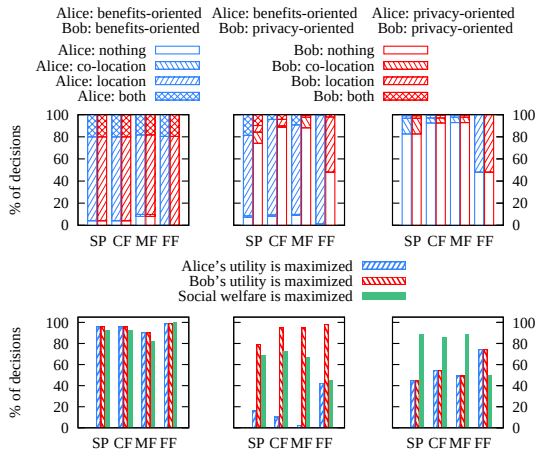


**Fig. 4.** Equilibria decisions (top row) and their properties (bottom row), when Alice and Bob have different preference profiles, corresponding to real survey data: 92 *benefits-oriented* profiles ($f_{pb} < 0.5$) and 158 *privacy-oriented* profiles ($f_{pb} > 0.5$). We present three scenarios: both Alice and Bob are *benefits-oriented* (left plots), Alice is *benefits-oriented* and Bob is *privacy-oriented* (middle plots) and both are *privacy-oriented* (right plots). Different adversarial models (SP, CF, MF, FF) are illustrated on the $x$ axis. In each of the top three plots, for each adversarial model, two bars (left blue for Alice and right red for Bob) indicate–on the $y$ axis–the proportion of times (aggregated over time instants and the number of preference profile pairs considered in that scenario) a player made one of the four decisions: share nothing (empty pattern), share only location (hash right pattern), share only co-location (hash left pattern) or share both (hash right-left pattern). Each of the three bottom plots show, on the $y$ axis, for each adversarial model, the proportion of times social welfare and individual utilities are maximized.

**The Case of *Benefits-Oriented* / *Privacy-Oriented* Players**. We present the case where the players have different values for the $f_{pb}$ factor. We select two subsets of preference profiles from our survey data: (1) the *privacy-oriented* (158 profiles), for which $f_{pb} > 0.5$; and (2) the *benefits-oriented* (92 profiles), for which $f_{pb} < 0.5$. We evaluate the outcome of the Sharing Game in three cases: (1) when Alice is *privacy-oriented* and Bob is *benefits-oriented*, (2) when both are *privacy-oriented* and (3) when both are *benefits-oriented*, for each possible pairs of preference profiles. Figure 4 illustrates our aggregated results (see caption for details). It is interesting that, when both players are *benefits-oriented*, the amount of shared co-location is substantial: It is *always shared* in the SP, MF and CF adversarial models (20% of all time instants), and is shared approximately 19.7% of all time instants in the FF model. To infer these numbers from Figure 4, we sum the values for "co-location" and "both". As discussed in Section 5.3.2, in any adversarial model, both players share the same amount of co-location. When one player is *benefits-oriented* and the other is *privacy-oriented*, **the amount of shared co-location varies significantly, with respect to the considered adversary**: It is always shared in the SP case, shared 5.4% of all time instants (27% of the time instants when the players are co-located) in the CF case, 10% of all time instants in the MF case and only 2% of all time instants in the FF case. One reason for this behavior is that the CF adversary has access to location information shared by both players, whereas the MF adversary has access only to location shared by one of them, so privacy losses stemming from shared co-locations are higher in the CF case, thus less co-location information is shared. It is interesting that this also causes ***both* players** to **share their location in the CF case *more frequently* than in the MF case** (in the CF case, it is enough that one player share his location after a shared co-location, for both players' privacy to be damaged, so the other player would be *forced* to also share his location for some benefit). When both players are *privacy-oriented*, location sharing is substantially reduced, but co-location is still shared 15% of all time instants in the SP case. The FF case illustrates a naturally emerging countermeasure: In all the cases, players find it most beneficial to report few co-locations (unlinking themselves from their friend makes the information unavailable to the FF adversary) and report their location most often (at no privacy cost). The equilibria decisions are frequently socially optimal: From 45% of the times (in the FF model, when Alice is *benefits-oriented* and Bob is *privacy-oriented*) to 99% of the times (in the FF model, when Alice and Bob are *benefits-oriented*). We notice that the case of players having opposite views regarding $f_{pb}$ is particularly problematic: Regardless of the considered adversary, this case presents the least amount of socially optimal equilibria decisions; furthermore, the utility of the *benefits-oriented* player is rarely maximized because his opponent would seldom share or allow sharing. Finally, misaligned preferences can lead to different decisions for the players as they only make the same decision 24% of the times in the SP model, 19.2% in the CF model and 11.6% in the MF model.

**The Case of *Sharer* / *Viewer* Players.** We study how the fact that the players have different values for the $f_{sv}$ preference factor affects their decisions and present these results in the appendix.

### 5.3.4 The Effects of Real Location Traces

In this section, we describe results on our **real dataset scenario**. We consider homogenous preference param-

eters in both the users' utility and set these using the average values of $f_{sv}$, $f_{lc}$ and $f_{pb}$ obtained in our survey, as presented in Table 2 in the Appendix. For the users' traces, we consider pairs of real traces (60 pairs of traces consisting of 300 time instants) sampled by Olteanu et al. [40] and the corresponding complex (and different) mobility profiles (we refer to Section 7.1 of the article for details). On these traces, users can report co-locations (i.e., meet), on average, 14.6% of the time instants (first quartile, median, third quartiles for the number of co-locations are 10.67%, 12.83%, and 16.67% of the time instants, respectively) and there are location samples at 37.55% of the time instants, on average. We simulate the game between Alice and Bob, for all the 60 pairs of traces, and twice for each of the pairs (s.t. each of the two traces in a pair is attributed to the first player).

**An Individual Snapshot.** We first present two randomly selected simulations of the game between Alice and Bob, illustrated in Figure 5. We choose a trace for Alice and two different traces for Bob; Alice's trace contains co-locations with that of Bob in 7% (Figure 5 top) and 10% (Figure 5 bottom) of the 300 time instants, respectively. A first observation is that a **vicious circle effect is noticeable**. After their first shared co-location ($t = 4$), users' behavior changes and they share more than they had done before that co-location: Alice shifts from not sharing anything at $t = 1, 2, 3$ to sharing both location and co-location at $t = 5$ and Bob from sharing only his location ($t = 1, 3$) to sharing both ($t = 5$). We repeatedly observe that, after a shared co-location, users continue to share. Specifically, in the first case (Figure 5 top), after a co-location was shared by either of the players, a player's subsequent decision involves sharing location (either only location, or both location and co-location) 95% and 90% of the times, for Alice and Bob, respectively. We consider only the time instants where location is available. This frequency is quite high, indeed higher than the frequency of deciding to share their locations (when available) over the entire traces: 68% and 81% of the time instants, respectively for Alice and for Bob. In the second case (Figure 5 bottom), the same effect is observed 91% and 88% of the times, respectively. Again, this is higher than the frequency of deciding to share their locations over the whole traces: 76% and 86% of the time instants, respectively. **The oversharing effect can, occasionally, be overcome at later time instants.** This can happen when users did not share for some time (either because they did not meet or because their location data is sparse) or when their location is particular (i.e., a rarely visited location, which would yield a higher error for the adversary, thus a higher privacy). Consequently, both Alice and Bob occasionally decide not to share anything (e.g., in Figure 5 top, Bob shares nothing at $t = 22, \cdots$; Alice shares nothing at $t = 65, 72, \cdots$; At $t = 74$, Bob shares nothing even though Alice shares her location and, *immediately after*, at $t = 75$, *the situation flips* and Alice shares nothing, whereas Bob shares his location; in Figure 5 bottom, at $t = 162$ both users only share co-location. Overall, of the times a player's location is available and he decides to share it, 33% and 21% in the first case (Figure 5 top), and 43% and 27% in the second case (Figure 5 bottom), respectively for Alice and Bob, happen right after a co-location was shared. To conclude, **even though Alice's location data, mobility profile, and preferences are constant, her behavior can change depending on the friend with whom she interacts (*even if he has the same preferences as she has*)**. Hence, **the specificities of the data** – the actual locations in the traces, the density of the location data, the meeting frequency (i.e., density of co-locations) and the patterns of the meetings, as well as the quality of the users' mobility profiles – **strongly affect the users' sharing decisions**, even when they have the same preferences.

**The Impact of Co-locations.** We now present the aggregated results of our simulations between all the pairs of traces, illustrated in Figure 6. We split and aggregate the results in two sets: those with traces that contain fewer co-locations than the median value for the entire dataset (12.83%) and those that contain more. We notice an **incentive to over-share co-locations and locations**: In the second case, where Alice and Bob have more co-locations, they *both share more locations and more co-locations, in all adversarial models.* For example, in the CF model, if in the first scenario Alice shares co-location and location 9.8% and 25.8% of the time instants, respectively, in the second one she shares co-location and location 18.8% (representing a relative increase of 87.8%) and 31.3% (representing a relative increase of 21.7%) of the times, respectively. Finally, on this dataset, the average users (in terms of their privacy preferences) would often choose to share some information. This can be attributed to the relatively high number of co-locations in the dataset, as well as to the quality of the mobility profiles.

## 5.4 Experimental Conclusions

In the *canonical meeting scenario*, which abstracts the specificities of the data, we exposed the fact that **the**

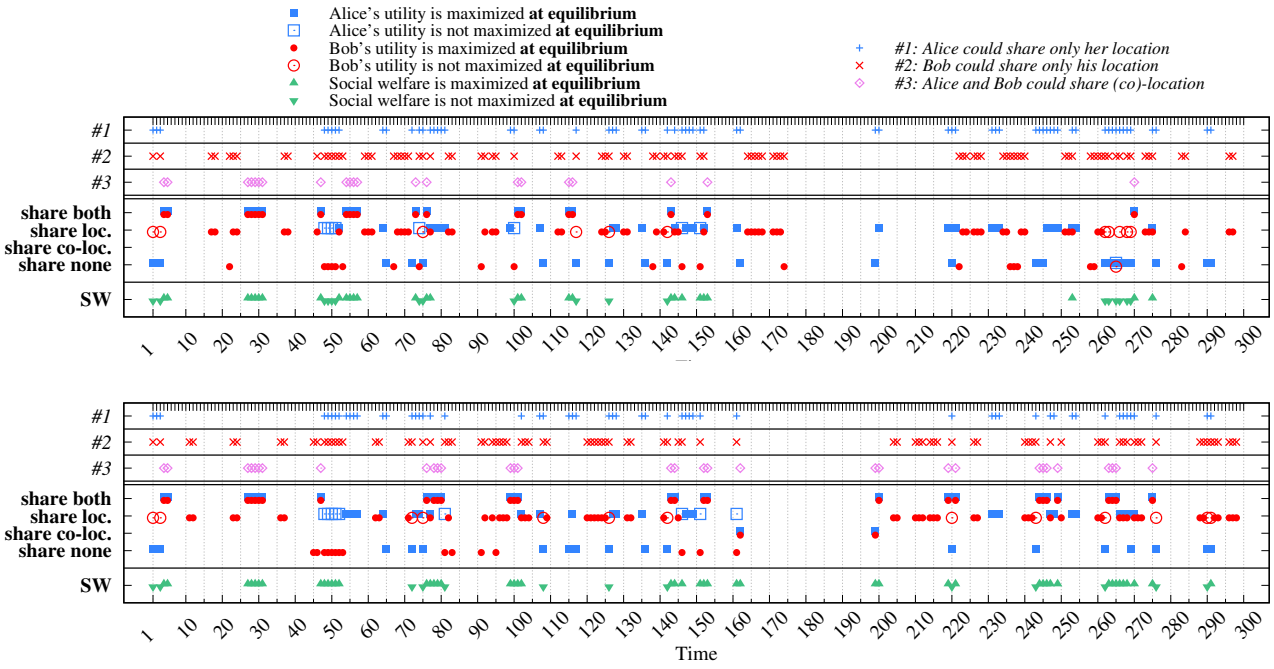


**Fig. 5.** Possible strategies and equilibria decisions (CF adversarial model) when Alice and Bob use two random traces from the Geolife dataset, for $f_{sv} = 0.57$, $f_{lc} = 0.56$, $f_{pb} = 0.60$. We illustrate the game outcome between Alice and Bob in two cases: %7 (top) and 10%(bottom) co-locations, respectively – the trace used by Alice is exactly the same in both cases. The $x$ axis shows the entire time window of interest. In the top section, on the $y$ axis, the *possible strategies* at each time instant of the game are illustrated: Alice can only share location (blue 'plus'), Bob can only share location (red 'cross'), Alice and Bob can both share location and co-location (violet 'rhombus'); For time instants where the game cannot be played (missing location data) there is no symbol. The middle and bottom sections of the $y$ axis illustrate the *equilibria decisions*: For every time instant, Alice's decision is represented by a blue rectangle and Bob's decision by a red circle; A player's corresponding shape is full if its utility at equilibrium is maximized, and empty otherwise. Additionally, each time instant at which both Alice and Bob can play is marked by an upward-pointing green triangle, if the equilibrium decisions maximize social welfare or by a downward-pointing green rectangle if they do not.

**users' different preferences factors lead to complexity in their interactions**. We noted in the *real dataset scenario* that **the users' equilibria decisions are also highly data dependent**: The users' actual traces, as well as the quality of their mobility profiles, can greatly influence users' sharing decisions, **even when they agree in terms of their preference factors**. It is not hard to imagine how much more complex the interactions between the users can become when different preference factors of the users are taken into account along with real (co-)location data. Understanding them is not a trivial problem and it is hard to draw generally applicable and quantifiable conclusions that take all these variations into account. We exposed, however, a few trends that we believe to be interesting. First, we identified the model parameters and the data specificities that have the strongest effect on the users' decisions: the frequency of co-locations, the quality of the users' mobility profiles, the considered adversary, the value of $\alpha$ ($f_{pb}$, the preference for privacy versus social benefits of *one* user); whereas other model parameters have a more moderate effect. Second, some interesting patterns of behavior emerge, for instance, the fact that a **vicious-circle effect** *can occur in the SP adversarial model*: When a player (say Alice) has a strong incentive to share, it is enough that she shares one co-location information and, with respect to the service provider, her friend (Bob), who might not be willing to share at all, will continue to have his privacy affected and be forced into sharing his location at several later times as well. This effect is propagated and stronger if Alice still wants to share her own location at other time instants, further damaging Bob's privacy. We also observed that the effect of a shared co-location can (sometimes) eventually fade away and that Bob, too, can influence Alice into sharing; these effects can quickly alternate, making the players' decisions vary over time. In other words, *the privacy effects propagate not only in space (influences from a friend) but also in time*. Third, we noticed that, in the FF model, a natural tendency is to share few co-locations but the users still share a significant amount of location information. Finally, we showed (e.g., in the

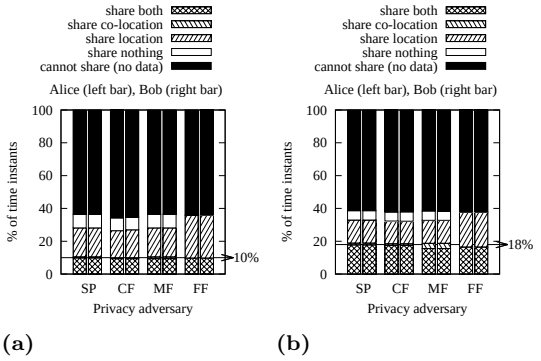


**Fig. 6.** Equilibria decisions when Alice and Bob use traces from the Geolife dataset, for $f_{sv} = 0.57, f_{lc} = 0.56, f_{pb} = 0.60$. We present two scenarios, depending on the number of co-locations they share: (a) # co-locations lower than the median and (b) # co-locations higher than the median. Different adversarial models (SP, CF, MF, FF) are illustrated on the $x$ axis. For each adversarial model, two bars (left for Alice and right for Bob) indicate–on the $y$ axis–the proportion of times (aggregated over the number of time instants and the number of runs considered in that scenario) a player made one of the four decisions: share nothing (empty pattern), share only location (hash right pattern), share only co-location (hash left pattern) or share both (hash right-left pattern); The fact that a player cannot share anything (i.e., missing location data) is illustrated by the solid pattern. The arrows represent, in each scenario, the average number of co-locations.

SP and CF models) that a (common) decision to share **co-location** creates the **incentive to *over-share locations at later times after the co-location***.

# 6 Extended Model

We now propose an extended model, relaxing the assumptions of selfishness, privacy-myopia and complete information for the players. We define the type of a player $i$ (denoted by $\theta_i(t)$) as the information that is private to her. This can include but is not limited to her actual location ($a_i(t)$), the background information that the adversary has on her ($\mathcal{B}_i$), her specific parameters that influence the computation of her social benefits and her privacy. Thus, $\theta_i(t) \triangleq (a_i(t), \mathcal{B}_i, f_{pb}, f_{lc}, f_{sv}, \dots)$. We assume that a player $i$ has information only about the possible domain and probability distributions of the other player's type and that she incorporates these into her utility function in the form of an expected value. Given $\boldsymbol{\theta}(t) = (\theta_i(t), \theta_j(t))$, the *individual utility* of player $i$ for some strategy profile $\mathbf{s}(t)$ at time $t$ captures both her social benefits and her privacy

$$\begin{aligned}\hat{u}_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j, \boldsymbol{\theta}(t)\right) \quad = \quad & (4)\\ (1-\alpha_i) \cdot B_i\left(t, \mathbf{a}(t), \mathbf{s}(t), \boldsymbol{\theta}(t)\right) + & \\ \alpha_i \cdot P\left(i, t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j, \boldsymbol{\theta}(t)\right) &\end{aligned}$$

Thus, the expected individual utility of a player can be computed as

$$\begin{aligned}\bar{u}_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) \quad = & \\ \mathbb{E}_{\boldsymbol{\theta}(t)}[\hat{u}_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j, \boldsymbol{\theta}(t)\right)] & \quad (5)\end{aligned}$$

As previous studies have shown (e.g., [35]), in the context of OSNs where users are friends, they might take into account their friends' individual utility hence choose their strategy based on their *perceived utility*, which we define for some strategy profile $\mathbf{s}(t)$ as follows

$$\begin{aligned}u_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) \quad = & \quad (6)\\ \bar{u}_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) + & \\ F_i \cdot \bar{u}_j\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) &\end{aligned}$$

where $F_i \in [0,1]$ denotes the altruistic factor of user $i$ for the other user. These factors can be experimentally measured using techniques based on conjoint analysis (e.g., [44–46]). Furthermore, as current decisions have privacy implications at future time instants, we consider the *cumulative utility* of user $i$ at time $t$ as the discounted sum of all perceived utilities from time $t$ (the present) until time $T$ (the future) as follows

$$\begin{aligned}U_i\left(t, \mathbf{a}(t), \mathbf{o}(t-1), \mathbf{s}(t), \mathcal{B}_i, \mathcal{B}_j\right) \quad = & \quad (7)\\ \mathbb{E}_{\mathbf{a}[t+1,\cdots,T]}\left[\sum_{t'=t}^{T} \delta^{t'-t} \cdot u_i\left(t', \mathbf{a}(t'), \mathbf{o}(t'-1), \mathbf{s}(t'), \mathcal{B}_i, \mathcal{B}_j\right)\right] &\end{aligned}$$

where $\delta$ is a discount factor, taking values in the interval $[0,1]$ and $\mathbf{a}[t+1, \cdots, T]$ denotes the vector of actual locations at times $t+1, \cdots, T$ for both $i$ and $j$. Note that, in computing her cumulated utility at time $t$, a user does not know any of the actual locations in the future ($t' > t$), hence the expectation value over $\mathbf{a}[t+1, \cdots, T]$. Given these locations, $\mathbf{s}(t')$ can be deterministically predicted.

At every time instant, we model the interactions as an incomplete information, extensive-form game (where the players observe each other's strategies). Solving such an extended game is not straightforward. Intuitively, players could choose the strategy that maximizes their expected utility with respect to the unknown information (i.e., the other player's type). However, the way in which players choose to estimate and use the unknown information (in the backward induction algorithm) strongly affects the complexity of the game. We plan to study the different possibilities and give a numerical solution in future work.

# 7 Discussion

This work represents the first step towards modeling the interplay between users in the context of (co-)location

sharing and the idea of combining a game-theoretic model with real-user parameters in this setting is also novel. For the first attempt to tackle the problem of understanding users' interaction in such a complex context, we focused on a number of specific scenarios and assumptions, that open several interesting directions for future work. We assumed that players do not report fake co-locations. Including *fake co-locations* into the model is straightforward and simply increases the number of possible strategies at times where players are not co-located. However, the benefit of sharing fake co-locations is likely different than that of sharing true co-locations. We included, in our evaluation, location privacy with respect to one adversary. As adversaries are not easily separable and a user is likely sensitive to more than one type of the adversaries that we mentioned, a *combination of* these *different adversaries* can be included in the utility function. We considered a game with two players. In doing so, we illustrated interdependence effects that work both in time (actions at previous times influence a user's future decisions for sharing) and in space (actions of a friend influence a user's decisions). To better illustrate the spacial effects, the framework can be extended to *more players and non-default visibility settings for posts*; intuitively, we expect that the cascading effect (one user's behavior affecting that of her friends', that of the friends of her friends, and so on and so forth) would occur, with an even greater impact with more players. Our evaluation focused on a few specific cases and maintained the number of free parameters low. We quantified only a limited number of preference factors (whose values could be specific to Facebook users) in order to avoid the questionnaire fatigue effect that would have decreased the quality of the participants' responses. *More preference factors can be evaluated through similar user surveys* (e.g., different values for $f_{pb}$ for the different adversarial models and their respective weights in the utility function; the benefit users gain for sharing fake co-locations). Furthermore, previous works (e.g., [3, 24]) have shown that user (reported) privacy attitudes do not always correspond to actual behaviors, and we have demonstrated that users decisions are highly data dependent. Hence our results should also be taken with a grain of salt: We believe that the trends that we have exposed are generic, but their magnitude will likely differ on other datasets or when considering heterogenous user privacy preferences. Finally, in future work, we plan to study the $N$-player $T$-time Sharing Game, by using multi-agent influence diagrams (MAIDs), which were introduced by Koller et al. [26] for efficiently solving complex games. Ultimately, our extensible model with quantifiable parameters can serve as the first building block to assist user decision-making in an informed manner. Specifically, we envision that a software tool (e.g., a Facebook client/extension) would use our framework to take these interactions into account and to assist users in improving their awareness and decision-making process.

# 8 Conclusion

It is well-known that the behavior of others affects our own privacy, in particular in the case of interdependent data. Yet, formalizing these complex interdependences and their implications is non-trivial, especially because human decisions play a dominant role. To address this issue, we focused on the (co-)location sharing features provided by major OSNs. We proposed a coarse-grained game-theoretic model and provided a first framework to study the interplay between two friends. A major challenge in such approaches is to assign meaningful values to the parameters that characterize user preferences. For this purpose, we carried out a survey of Facebook users, which also confirmed the anticipated high diversity of opinions in terms of social benefits and location privacy. We studied the resulting equilibria and their properties, in different settings. In particular, we showed how, because of conflicting preferences, one of the users can be forced into a situation that she does not desire (e.g., we exemplified on a mobility dataset how a vicious-circle effect emerges) and we demonstrated that sharing co-location information can additionally encourage users to over-share their locations. This is an interesting finding from a design perspective for the OSN service providers but a dangerous one for the end users: Advertising features that permit the sharing of co-location information could also encourage users to share their locations more often. Furthermore, we showed that user's decisions are strongly influenced by the adversary that they consider and dependent on the mobility data. We emphasized the need to develop appropriate warning mechanisms for the users, which we intend to develop in the future; these would help users better understand and anticipate the consequences of their (co-)location sharing decisions.

# Acknowledgments

The authors are very grateful to Elisa Celis and Italo Dacosta for their useful insights.

# References


[1] A. Acquisti. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce*, pages 21–29. ACM, 2004.

[2] A. Acquisti. Nudging privacy: The behavioral economics of personal information. *IEEE Security Privacy*, 2009.

[3] A. Acquisti and J. Grossklags. *Privacy Attitudes and Privacy Behavior*. 2004.

[4] A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security Privacy*, 3(1), 2005.

[5] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *WWW*, 2010.

[6] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *IMC*, 2009.

[7] A. Besmer and H. Richter Lipford. Moving beyond untagging: photo privacy in a tagged world. In *Proc. of SIGCHI*. ACM, 2010.

[8] G. Biczók and P. H. Chia. Interdependent Privacy: Let Me Share Your Data. In *FC*, 2013.

[9] I. Bilogrevic, K. Huguenin, B. Ağır, M. Jadliwala, M. Gazaki, and J.-P. Hubaux. A Machine-Learning Based Approach to Privacy-Aware Information-Sharing in Mobile Social Networks. *Pervasive and Mobile Computing (PMC)*, Nov. 2015.

[10] A. Chaabane, G. Acs, and M. Kaafar. You are what you like! information leakage through users' interests. In *NDSS*, 2012.

[11] J. Chen, J. W. Ping, Y. C. Xu, and B. C. Tan. Information privacy concern about peer disclosure in online social networks. *IEEE Transactions on Engineering Management*, 62(3):311–324, 2015.

[12] H. Cho and A. Filippova. Networked privacy management in Facebook: A mixed-methods and multinational study. In *Proc. of CSCW*. ACM, 2016.

[13] D. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proc. of PNAS*, 107, 2010.

[14] R. Dey, C. Tang, K. Ross, and N. Saxena. Estimating age privacy leakage in online social networks. In *INFOCOM*, 2012.

[15] R. L. Fogues, P. K. Murukannaiah, J. M. Such, and M. P. Singh. Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(1):5, 2017.

[16] D. Fudenberg and J. Tirole. *Game theory*. MIT press, 1991.

[17] L. González-Manzano, A. I. González-Tablas, J. M. de Fuentes, and A. Ribagorda. Cooped: Co-owned personal data management. *Computers & Security*, 47, 2014.

[18] P. E. Green and V. Srinivasan. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research*, 5(2), 1978.

[19] B. Henne, C. Szongott, and M. Smith. Snapme if you can: Privacy threats of other peoples' geo-tagged media and what we can do about it. In *WiSec*, 2013.

[20] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy. In *CCS'13: Proc. of the 20th ACM Conf. on Computer and Communications Security*, 2013.

[21] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti. On non-cooperative genomic privacy. In *FC*, 2015.

[22] P. Ilia, I. Polakis, E. Athanasopoulos, F. Maggi, and S. Ioannidis. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *CCS*, 2015.

[23] H. Jia and H. Xu. Autonomous and interdependent: Collaborative privacy management on social networking sites. In *Proc. of CHI*. ACM, 2016.

[24] S. Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers Security*, 2015.

[25] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. 2009.

[26] D. Koller and B. Milch. Multi-agent influence diagrams for representing and solving games. *Games and economic behavior*, 45(1), 2003.

[27] H. Krasnova, S. Spiekermann, K. Koroleva, and T. Hildebrand. Online social networks: Why we disclose. *Journal of information technology*, 25(2), 2010.

[28] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6), 2009.

[29] A. Lampinen, V. Lehtinen, A. Lehmuskallio, and S. Tamminen. We're in it together: interpersonal management of disclosure in social network services. In *Proc. of CHI*, 2011.

[30] A. Laszka, M. Felegyhazi, and L. Buttyan. A survey of interdependent information security games. *ACM Computing Surveys*, 2015.

[31] R. S. Laufer, H. M. Proshansky, and M. Wolfe. Some analytic dimensions of privacy. In *Proceedings of the Lund Conference on Architectural Psychology*. Lund, Sweden, 1973.

[32] R. S. Laufer and M. Wolfe. Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of social Issues*, 33(3):22–42, 1977.

[33] M. Li, H. Zhu, Z. Gao, S. Chen, L. Yu, S. Hu, and K. Ren. All your location are belong to us: Breaking mobile social networks for automated user location tracking. In *MobiHoc*, 2014.

[34] R. Mason, R. Gunst, and J. Hess. *Statistical design and analysis of experiments with applications to engineering and science*. J. Wiley, 2003.

[35] D. Meier, Y. A. Oswald, S. Schmid, and R. Wattenhofer. On the Windfall of Friendship: Inoculation Strategies on Social Networks. In *EC*, 2008.

[36] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You are who you know: Inferring user profiles in online social networks. In *WSDM*, 2010.

[37] G. Misra and J. M. Such. Pacman: Personal agent for access control in social media. *IEEE Internet Computing*, 21(6):18–26, 2017.

[38] R. B. Myerson. *Game theory*. Harvard university press, 2013.

[39] A. Noulas, M. Musolesi, M. Pontil, and C. Mascolo. Inferring interests from mobility and social interactions. In *Proc. of NIPS Workshops*, 2009.

[40] A.-M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, and J.-P. Hubaux. Quantifying interdependent privacy risks with

location data. In *IEEE Trans. Mobile Comput.*, 2016.
[41] Z. D. Ozdemir, H. J. Smith, and J. H. Benamati. Antecedents and outcomes of information privacy concerns in a peer context: An exploratory study. *European Journal of Information Systems*, 26(6):642–660, 2017.
[42] X. Page, B. P. Knijnenburg, and A. Kobsa. FYI: Communication Style Preferences Underlie Differences in Location-sharing Adoption and Usage. In *UbiComp*, 2013.
[43] I. Polakis, G. Argyros, T. Petsios, S. Sivakorn, and A. D. Keromytis. Where's wally?: Precise user discovery attacks in location proximity services. In *CCS*, pages 817–828, 2015.
[44] Y. Pu and J. Grossklags. An economic model and simulation results of app adoption decisions on networks with interdependent privacy consequences. In *GameSec*, 2014.
[45] Y. Pu and J. Grossklags. Towards a model on the factors influencing social app users' valuation of interdependent privacy. *PoPETS*, 2015.
[46] Y. Pu and J. Grossklags. Using conjoint analysis to investigate the value of interdependent privacy in social app adoption scenarios. In *ICIS*. Assoc. for Information Systems, 2015.
[47] Y. Pu and J. Grossklags. Valuating friends' privacy: Does anonymity of sharing personal data matter? In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, 2017.
[48] S. Rajtmajer, A. Squicciarini, J. M. Such, J. Semonsen, and A. Belmonte. An ultimatum game model for the evolution of privacy in jointly managed content. In *International Conference on Decision and Game Theory for Security*, pages 112–130. Springer, 2017.
[49] K. Raynes-Goldie. Aliases, creeping, and wall cleaning: Understanding privacy in the age of facebook. *First Monday*, 15(1), 2010.
[50] E. M. Redmiles, M. L. Mazurek, and J. P. Dickerson. Poster: Do users make rational security decisions? 2018.
[51] C. Riederer, D. Echickson, S. Huang, and A. Chaintreau. Findyou: A personal location privacy auditing tool. In *WWW*, 2016.
[52] R. Shokri. Privacy games: Optimal user-centric data obfuscation. *PoPETS*, 2015(2), 2015.
[53] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec. Protecting location privacy: optimal strategy against localization attacks. In *CCS'12: Proc. of the 19th ACM Conf. on Computer and Communications Security*, 2012.
[54] Y. Shoshitaishvili, C. Kruegel, and G. Vigna. Portrait of a privacy invasion; detecting relationships through large-scale photo analysis. *PoPETS*, 2015.
[55] J. M. Such and N. Criado. Resolving multi-party privacy conflicts in social media. *IEEE KDE*, 2016.
[56] J. M. Such, J. Porter, S. Preibusch, and A. Joinson. Photo privacy conflicts in social media: A large-scale empirical study. In *Proc. of CHI*, 2017.
[57] E. Toch, J. Cranshaw, P. H. Drielsma, J. Y. Tsai, P. G. Kelley, J. Springfield, L. Cranor, J. Hong, and N. Sadeh. Empirical models of privacy in location sharing. In *UbiComp*, 2010.
[58] J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Princeton university press, 2007.
[59] N. Vratonjic, K. Huguenin, V. Bindschaedler, and J.-P. Hubaux. A Location-Privacy Threat Stemming from the Use of Shared Public IP. *IEEE Trans. on Mobile Computing (TMC)*, 13(11), Nov. 2014.
[60] G. Wang, S. Y. Schoenebeck, H. Zheng, and B. Y. Zhao. "will check-in for badges": Understanding bias and misbehavior on location-based social networks. In *ICWSM*, 2016.
[61] A. F. Westin. Social and political dimensions of privacy. *Journal of social issues*, 59(2):431–453, 2003.
[62] J. Wiese, P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. Are you close with me? are you nearby?: Investigating social groups, closeness, and willingness to share. In *UbiComp*, 2011.
[63] P. Wisniewski, H. Lipford, and D. Wilson. Fighting for my space: Coping mechanisms for sns boundary regulation. In *Proc. of SIGCHI*. ACM, 2012.
[64] Xlstat statistical software for microsoft excel. https://www.xlstat.com/en/, 2016. last visited: Aug. 2016.
[65] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin. Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017.
[66] H. Xu. Reframing privacy 2.0 in online social network. *U. Pa. J. Const. L.*, 2011.
[67] M. Xue, C. Ballard, K. Liu, C. Nemelka, Y. Wu, K. Ross, and H. Qian. You can yak but you can't hide: Localizing anonymous social network users. In *Proceedings of the 2016 Internet Measurement Conference*, pages 25–31. ACM, 2016.
[68] M. Xue, Y. Liu, K. W. Ross, and H. Qian. I know where you are: thwarting privacy protection in location-based social discovery services. In *IEEE Conference on Computer Communications Workshops*, pages 179–184. IEEE, 2015.
[69] M. Yang, Y. Yu, A. K. Bandara, and B. Nuseibeh. Adaptive sharing for online social networks: A trade-off between privacy risk and social benefit. In *TrustCom*, 2014.
[70] Y. Zheng, L. Liu, L. Wang, and X. Xie. Learning transportation mode from raw GPS data for geographic applications on the web. In *WWW*, pages 247–256, 2008.

# 9 Appendix

## 9.1 Conjoint Analysis & Game Theory 101

Conjoint analysis [18] is an experimental approach used to detect the hidden rules users rely on to make decisions (involving trade-offs) between services. A service is viewed as a combination of attributes, each of which has different levels (values). Users are asked to rank multiple versions of the service (each being a different combination of attribute levels). The combination of attributes and levels can lead to a large number of versions to be ranked. In order to keep the complexity of this task manageable for the users, the number of proposed versions can be reduced, in an optimal way, to a reasonable yet meaningful number, through *fractional factorial de-*

| | |
|---|---|
| $s_i(t) = (sl_i(t), sc_i(t))$ | A possible strategy of user $i$ at time $t$ |
| $\bar{L}$ or $sl.(t) = 0$ (False) | Hide location |
| $L$ or $sl.(t) = 1$ (True) | Share location |
| $\bar{C}$ or $sc.(t) = 0$ (False) | Hide co-location |
| $C$ or $sc.(t) = 1$ (True) | Share co-location |
| $s^*(t) = \left(s_i^*(t), s_j^*(t)\right)$ | Equilibrium strategy profiles (decisions) at time $t$ |
| $\alpha_i$ | Weight with which user $i$ values privacy over benefits |
| $B_i\left(t, \mathrm{a(t)}, \mathrm{s}(t)\right)$ | User $i$'s benefits at time $t$ for strategy profile $\mathrm{s}(t)$ |
| $b_{sl}^i$ | User $i$'s benefit of sharing her actual location at $t$ |
| $b_{vl}^i$ | User $i$'s benefit of viewing her friend's location at $t$ |
| $b_{sc}^i$ | User $i$'s benefit of sharing co-location with a friend at $t$ |
| $b_{vc}^i$ | User $i$'s benefit of viewing co-loc. shared by a friend at $t$ |
| $f_{sv}^i$ | User $i$'s preference factor: sharing vs. viewing |
| $f_{lc}^i$ | User $i$'s preference factor: location vs. co-location |
| $f_{pb}^i$ | User $i$'s preference factor: privacy vs. benefits |
| $a_i(t)$ | User $i$'s actual location at time $t$ |
| $\mathrm{o}(t)$ | Information observed by the adversary in the time window up to $t$ |
| $P\left(i, t, \mathrm{a}(t), \mathrm{o}(t-1), \mathrm{s}(t), \mathcal{B}_i, \mathcal{B}_j\right)$ | User $i$'s privacy at time $t$ for some strategy profile $\mathrm{s}(t)$ |
| $\mathcal{B}_i$ | The adversary's background knowledge about user $i$ |
| $\theta_i(t)$ | User $i$'s type at time $t$ (includes actual location, benefits vs. privacy preferences, $\cdots$) |
| $\hat{u}_i\left(t, \mathrm{a}(t), \mathrm{o}(t-1), \mathrm{s}(t), \mathcal{B}_i, \mathcal{B}_j, \theta(t)\right)$ | User $i$'s *individual utility* at time $t$ for strategy profile $\mathrm{s}(t)$ and player types $\theta(t)$ |
| $\bar{u}_i\left(t, \mathrm{a}(t), \mathrm{o}(t-1), \mathrm{s}(t), \mathcal{B}_i, \mathcal{B}_j\right)$ | User $i$'s *expected individual utility* at time $t$ for strategy profile $\mathrm{s}(t)$ |
| $u_i\left(t, \mathrm{a}(t), \mathrm{o}(t-1), \mathrm{s}(t), \mathcal{B}_i, \mathcal{B}_j\right)$ | User $i$'s *perceived utility* at time $t$ for strategy profile $\mathrm{s}(t)$ (includes altruism) |
| $F_i$ | The altruistic factor of user $i$ for the other user |
| $U_i\left(t, \mathrm{a}(t), \mathrm{o}(t-1), \mathrm{s}(t), \mathcal{B}_i, \mathcal{B}_j\right)$ | User $i$'s utility/*cumulative utility* (includes future considerations) at time $t$ for strategy profile $\mathrm{s}(t)$ |
| $\delta$ | Discount factor for future considerations in the cumulated utility |
| $SW(t, s_i(t), s_j(t))$ | Social welfare at time $t$ for strategy profile $(s_i(t), s_j(t))$ |

**Table 1.** Table of notations.

| | $f_{sv}$ | $f_{lc}$ | $f_{pb}$ |
|---|---|---|---|
| **avg. $\pm$ stddev.** | **$.57 \pm .15$** | **$.56 \pm .15$** | **$.60 \pm .39$** |
| **proportion of users with $f_* > 0.5$ (prefer sharing/ location/ privacy)** | **60%** | **54%** | **63.2%** |
| **proportion of users with $f_* = 0.5$ (indifferent)** | **16.8%** | **20%** | **N/A** |
| **proportion of users with $f_* < 0.5$ (prefer viewing/ co-location/ benefits)** | **23.2%** | **26%** | **36.8%** |

**Table 2.** User preference factors extracted from the survey. $f_*$ denotes, depending on the column, $f_{sv}$, $f_{lc}$, or $f_{pb}$.

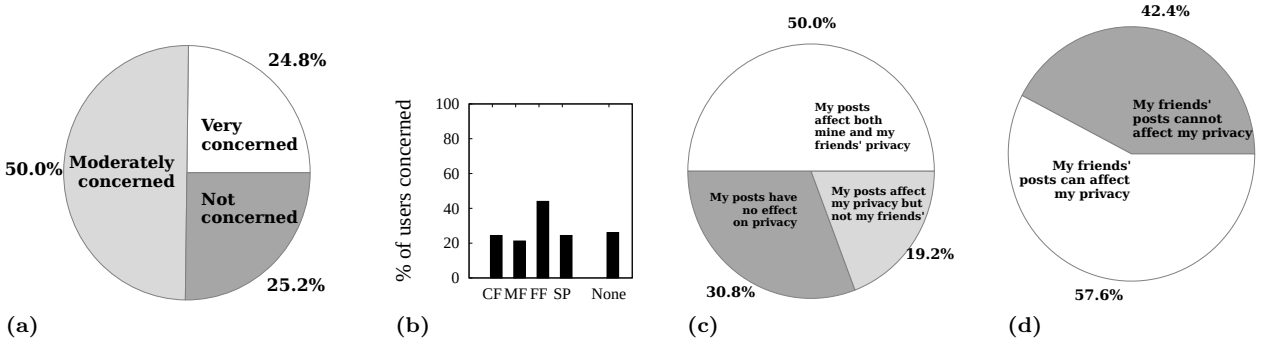


**Fig. 7.** (a) Users' concern about location privacy; (b) The adversaries that users are concerned about: Our friends in common (CF), My other friends (MF), My friend's other friends (FF), The service provider (SP). Users' awareness about (c) privacy risks stemming from their own posts; (d) own privacy risks stemming from friends' posts.

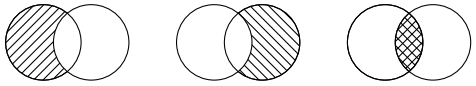


**Fig. 8.** Friends adversarial models (hashed area) for user $i$: (a) My other friends model (MF); (b) My friend's other friends model (FF); (c) Our friends in common model (CF). The social circle of user $i$ (resp. $j$) is represented by the left (resp. right) circle. The intersection represents the common friends of $i$ and $j$.

*sign* [34]. The hidden value users place on each of the attribute levels is then quantified through statistical analysis, as *part-worth utilities* and *importance values*. The importance values represent how much difference each attribute makes in the total utility of the service; these are represented as percentages for all the attributes.

Game theory is the study of the strategic interaction between multiple rational decision-makers who aim to maximize their own utility [16, 38, 58]. This mathematical theory enables us to derive more than the optimal strategy that a rational agent would adopt given various parameters: It enables the modeling and prediction of stable states, called *equilibria*, in which none of the agents can improve his utility given all other agents' utility functions and strategies. It has been notably used in economics, biology, political science, psychology, and computer science. It is especially relevant for our work as it enables us to model and analyze users' preferences and interactions, and to predict their resulting rational behaviors. A core concept of game theory is the Nash equilibrium (NE), which represents the stable state in which no agent (a so-called player), by taking into account other players' strategies (so-called opponents), has incentive to deviate from his strategy. A refinement of the NE is the subgame perfect Nash equilibrium (SPNE). This refers to an equilibrium derived by considering a smaller part of the whole game tree, by eliminating incredible threats (strategies that would not rationally be chosen). A common method for finding a SPNE is called backward induction; it first considers the last actions of the game and derives the best decision of the last player, given all other previous possible decisions in the game. Social welfare is defined as the sum of the utilities of all players. A strategy profile (set of players' strategies) is called *social optimum* if it maximizes the social welfare. A NE is not necessarily a social optimum, but that finding a socially-optimal NE is highly desirable from a mechanism design perspective. This process continues to the second to last actions, and so on until it reaches the first move of the game, i.e., the root of the game tree.

## 9.2 The Case of *Sharer* / *Viewer* Players

We study how the fact that the players have different values for the $f_{sv}$ preference factor affects their de-

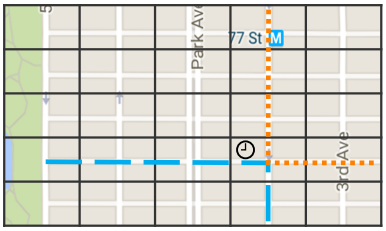


**Fig. 9.** Canonical meeting scenario: Two users, Alice (dotted) and Bob (dashed), coming from distinct directions, meet for some time, and later separate in distinct directions.

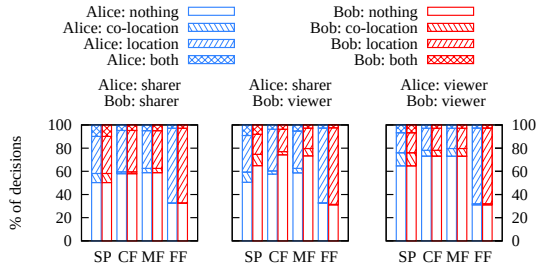


**Fig. 10.** Equilibria decisions (a) and their properties (b), when Alice and Bob have different preference profiles, corresponding to real survey data: 150 *sharers*' profiles ($f_{sv} > 0.5$) and 58 *viewers*' profiles ($f_{sv} < 0.5$). We present three scenarios: both Alice and Bob are *sharers* (left plots), Alice is a *sharer* and Bob is a *viewer* (middle plots) and both are *viewers* (right plots). Note that, due to the symmetry of the trajectories in the meeting scenario, the case where Alice is a *viewer* and Bob is a *sharer* is symmetric to the case where Alice is a *sharer* and Bob is a *viewer*. Different adversarial models (SP, CF, MF, FF) are illustrated on the $x$ axis. In each of the top three plots, for each adversarial model, two bars (blue on the left for Alice and red on the right for Bob) indicate–on the $y$ axis–the proportion of times (aggregated over time instants and the number of preference profile pairs considered in that scenario) a player made one of the four possible decisions: share nothing (empty pattern), share only location (hash right pattern), share only co-location (hash left pattern) or share both (hash right-left pattern). Each of the three bottom plots show, on the $y$ axis, for each adversarial model, the proportion of times social welfare and individual utilities are maximized.

cisions, on the canonical meeting scenario (illustrated in Figure 9). We select two subsets of preference profiles from our survey data: the *sharers* (150 profiles)–for which $f_{sv} > 0.5$–and the *viewers* (58 profiles)–for which $f_{sv} < 0.5$. We evaluate the outcome of the Sharing Game in three cases, for each possible pairs of preference profiles: when Alice has a *sharer*'s preference profile and Bob a *viewer*'s, when both have *sharers* profiles and when both have *viewers* profiles.

Figure 10 shows our aggregated results (see caption for details). We note that the interplay between the various parameters of the preference profiles (e.g., a *sharer* profile encourages sharing because $f_{sv} > 0.5$, but it could also discourage sharing if $f_{pb} > 0.5$) results in a large variety in the distribution of players' equilibria decisions. Despite this variability, a few trends are still distinguishable. First, in general, a *sharer* shares more information than a *viewer* and the most information is shared when both Alice and Bob are *sharers*, whereas the least information is shared when both are *viewers*. Second, regardless of the players' types (*sharer*/*viewer*), and due to the forcing effect, the largest amount of co-location is shared in the SP model (e.g., 17% of all time instants when both players are *sharers*); the smallest amount of co-location is shared in the FF model (e.g., 3.6% of all time instants when both players are *viewers*), when players find it most beneficial to report few co-locations and report their location most often (at no privacy cost). Furthermore, the equilibria decisions are frequently socially-optimal: From 52% of the times (in the FF model, when both Alice and Bob are *viewers*) to 85% of the times (in the CF model, when both Alice and Bob are *sharers*). Regardless of the adversary, the most socially-optimal equilibria are reached when both players are *sharers* and the least when both players are *viewers* (due to the fact that a *viewer* player shares less than a *sharer* player and, consequently, their opponent benefits less from their posts).