Understanding User Preferences and Awareness: Privacy Mechanisms in Location-Based Services

Thorben Burghardt, Erik Buchmann, Jens Müller, and Klemens Böhm

Universität Karlsruhe (TH), 76131 Karlsruhe, Germany
{burgthor,buchmann,muellerj,boehm}@ipd.uka.de

Abstract. Location based services (LBS) let people retrieve and share information related to their current position. Examples are Google Latitude or Panoramio. Since LBS share user-related content, location information etc., they put user privacy at risk. Literature has proposed various privacy mechanisms for LBS. However, it is unclear which mechanisms humans really find useful, and how they make use of them. We present a user study that addresses these issues. To obtain realistic results, we have implemented a geotagging application on the web and on GPS cellphones, and our study participants use this application in their daily lives. We test five privacy mechanisms that differ in the awareness, mental effort and degree of informedness required from the users. Among other findings, we have observed that in situations where a single simple mechanism does not meet all privacy needs, people want to use simple and sophisticated mechanisms in combination. Further, individuals are concerned about the privacy of others, even when they do not value privacy for themselves.

1 Introduction

Location based services (LBS) are important in many application areas, e.g., platforms for socializing like Google Latitude or geotagging applications like Panoramio. LBS require their users to disclose private information. Think of a person who is using a geotagging service with her GPS cellphone to annotate locations. This allows other users to find out (1) where the person has been at a certain time, (2) which data she finds interesting, and (3) her itinerary, by constructing coarse tracks from subsequent annotations. LBS that manage high-resolution tracks reveal further details, e.g., the fitness of a hiker. Almost any LBS application comprises similar privacy threats. Research has proposed various privacy enhancing technologies (PETs) for LBS. They vary in complexity and require different levels of awareness, mental effort and understanding from their users. However, as people adapt to the functionality available, it is unclear which PET is indeed suitable for LBS users. In this paper, we describe the results of a study that addresses the following research questions:

Q1: Which information do people disclose in the LBS context? To deploy PETs it is important to know (1) which data people disclose in LBS, and (2) what they generate for personal use only. We distinguish between information

R. Meersman, T. Dillon, P. Herrero (Eds.): OTM 2009, Part I, LNCS 5870, pp. 304–321, 2009.

on locations, tracks, content (e.g., annotations of a location) and metadata (e.g., creation date). We want to find out which, how much and what kind of data people reveal, and if the locations disclosed refer to hot spots of the daily life.

Q2: Which social groups are allowed to see private information? LBS are frequently used to share personal locations, e.g., to make appointments. Thus, we want to observe which information people disclose to the general public and which one to their social relationships.

Q3: How do people use privacy mechanisms? If people are unable to use a PET properly, it fails in practice. We find out how people cope with PETs ranging from simple to complex ones, and how they use mechanisms that require awareness and mental effort.

Q4: Which kind of privacy mechanisms do people prefer? PETs have to meet the expectations and desires of their users. We want people to rank PETs and PET combinations to find out which ones they would like to use.

These questions can be answered only by means of a user study. This is challenging: First, privacy preferences obtained from offline surveys or artificial lab experiments are not necessarily the natural ones [1]. Instead, people tend to overestimate their privacy needs. Second, as mobile devices with integrated GPS, cellphone and broadband Internet connection (XDAs) are just about to enter the mass market, people do not yet possess in-depth knowledge of privacy threats of LBS applications. Furthermore, we cannot assume that users know PETs proposed in the scientific literature [2,3,4,5]. Third, it requires high effort and financial expenses to equip study participants with equipment, to ensure that technical issues do not bias the study results and to keep participants motivated over several weeks. Fourth, many privacy threats in LBS originate from sharing information and from social interactions. Thus, meaningful results require a group of participants who are used to interact with each other.

We have conducted an extensive user study that considers these issues. Our primary interest is on privacy relationships between LBS users, i.e., we assume the service provider to be a trusted third party and leave aside privacy issues between service provider and consumers. We have developed a mobile geotagging application, which is a popular example of LBS that share user-generated content among many individuals. Thus, our study results will be relevant for other LBS as well. We are first to study LBS privacy in a real setup: We let individuals use our geotagging application in their everyday life for two weeks. Our participants access the geotagging application from XDAs and via a web application. Without revealing our interest in privacy right away, we ask the participants to assign tags, e.g., 'best coffee ever' or 'house of friend', to locations of interest. To raise privacy threats, we invite family members, teachers, friends, classmates and acquaintances to participate in our study as well. They are able to browse and search our LBS for any tags, tracks, metadata and time information, unless the participant who has provided this information has classified it as private. In order to observe privacy needs related to these threats and the usage of privacy mechanisms, we have implemented a number of well-known PETs, ranging from

a straightforward on/off switch to (de-)activate track recording as used in Google Latitude, to sophisticated mechanisms like anonymized requests for tags.

Our analysis shows that participants do not like to introduce differentiations between individuals with regard to privacy when it comes to making information available. For more than 83% of all information generated, they either make it visible to anybody or to nobody. From the opposite perspective, participants wanted to use fine-grained privacy mechanisms for 17% of the information generated, i.e., our study participants are privacy-aware. We also find that they want to combine PETs of different complexity. Finally, people are concerned about the privacy of their friends and acquaintances, even if they do not mind disclosing information on themselves.

Paper structure: We discuss related work in Section 2, privacy threats and PETs in Section 3 and our study methodology in Section 4. In Section 5 we present our results, Section 6 concludes.

2 Related Work

Today virtually everybody is aware of the potential of LBS as well as of their privacy threats. The EU directive (2002/58/EC) [6] requests the consent of a user before her exact location may be processed. However, following the well known definitions of privacy of Alan Westin [7] – "Privacy is the claim of an individual to determine what information about himself or herself should be known to others." – such undifferentiated, binary decisions to always disclose everything when having given consent once are inappropriate. Research offers PETs to protect privacy in a fine grained way. However, there is a well-known gap between the social requirements and the technical mechanisms [8], and it is important to consider how people perceive and use PETs. In the following we give an overview of relevant PETs and of existing user studies.

PETs for LBS range from simple switches to (de-)activate GPS [9] over mechanisms where the user can differ between social groups and vary the accuracy for each piece of information disclosed [10] to mechanisms that work without any user interaction. Automatic mechanisms include [2] where the authors adapt the idea of k-anonymity to LBS, i.e., a request is indistinguishable from k-1 requests of other users. [3] introduces algorithms to create false dummies to conceal the real location. CliqueCloak [4] allows for a variable k and combines spatial and temporal aspects. Casper [5] provides a location anonymizer extending k-anonymity with a minimal area and offers a privacy aware query processor. [11,12] describe PET tailored to continuous LBS and trajectories. We adapt the idea of k-anonymity in one PET evaluated in our study.

Studies on LBS: Studies, e.g., [9], examine the privacy implications of several imaginary location-based services. They asked 16 participants questions referring to usefulness, how often they would use the service, and which privacy implications they see. The main result concerning privacy is that continuous location tracking causes more concerns than services that simply are location-aware. [13]

presents a questionnaire survey with 130 participants. Participants stated their privacy preferences for different situations (e.g., work lunch and leisure activity) and for location requests from different persons. They found that the requester is paid more attention to than the situation. Two further studies address a similar scenario: In [10], 16 participants had to give the names of persons from different social relationships. Over two weeks, participants received randomly generated hypothetical requests from these persons and replied with their current occupation, and what they would disclose about their position to the requester. The results are that the main factors to disclose information is the requesting person, the reason for the request, and if the information is deemed useful for the requester. However, requests have been hypothetical, and participants knew this. In our study, requests are real, and decisions to disclose information take immediate effect. In [14], the authors used a real application (running on a mobile phone) involving location requests and automatic notifications that are triggered when participants approached places defined previously. The location is the cell tower reachable. The study was performed with 8 participants (developers of the application and their spouses) over a period of five days. As a result, most of the requests granted came from spouses, but the small volume of data did not allow for any further conclusions. The application used is based on SMS communication. In our study, participants are permanently connected via broadband Internet, exchange their location and, in addition, content. We offer much higher accuracy by using GPS, i.e., threats are more obvious, and the areas defined are more precise.

To our knowledge, no study has yet explored the usage of different PETs under real life conditions, with a real application and realistic privacy threats.

3 Privacy and PETs in LBS

In this section, we describe our study scenario, respective privacy threats and the PETs we consider.

3.1 Mobile Geotagging

A popular variant of LBS is mobile geotagging where people assign various kinds of content to locations. For example, Panoramio.com allows to browse pictures, tags, comments, location, camera information and the photographer on a Google Maps mashup. Panoramio shows that the locations managed by LBS are related to places where the user has been or currently is, and the content provided is informative regarding user interests and attitudes.

To observe how people deal with privacy threats in an LBS context, we have implemented a mobile geotagging application based on the Streamspin framework [15] and a spatial database. With our application, a user could assign the position of his house the tags 'my home'. Our LBS stores the real name of the person tagging, a user pseudonym[1], creation times, tags, geo-coordinates and

[1] We distinguish between real name and pseudonym, because we want to observe if individuals differentiate between these when disclosing personal identifiers.

tracks of consecutive locations. The users of our LBS can browse and search for all tags and the corresponding metadata if the person who has generated the information has made it visible. Our application can be used via a XDA (a HTC Trinity), or with a web application. The XDA allows to assign tags to the current position of its user, i.e., it connects to the LBS, and it stores tracks and tuples <*user, timestamp, location, tag list*> in a database. Furthermore, the XDA allows to search for tags assigned to locations next to the current position. The XDA is updated in real time, if new tags appear in the vicinity. The web application displays the tags, positions and tracks with a Google Maps mashup. It lets each user view and edit all information provided. To fill the database with examples, we imported 200.000 geo-referenced keywords from Wikipedia.

3.2 Privacy Threats

When considering information accessible to other LBS users, privacy problems can arise from four sources:

Content. Content is the information people generate explicitly [16]. For example, the content of a geotagging application is the tags or photos provided, and the content of a people-finder service can be the social network of the participants. Clearly, content can be sensitive information.

Location. Mobile LBS obtain the current position of the user whenever she provides content or issues a query [5]. Other LBS might also manage locations the user has visited before. This is problematic, as location information can be private, e.g., the house of a friend, a hospital or a religious building.

Tracks. Tracks containing user itineraries can be the result of a LBS that records the movement of an XDA. Alternatively, LBS users can generate tracks by sorting locations publicly visible chronologically. Tracks can reveal where a person lives, buildings visited, relationships, daily routine etc [17].

Metadata. In LBS, metadata relate to the content, locations and tracks provided. Examples are the user name or the creation date. Metadata of content depend on the application area. For example, photos could contain EXIF information on the camera model. This information may be specific to individuals, and its dissemination may be a privacy leak.

3.3 Privacy Enhancing Technologies for LBS

A huge number of PETs is available for LBS. They range from simple, intuitive mechanisms to sophisticated ones that require a fundamental understanding of privacy threats and the technical background. To investigate a wide range of approaches, we have decided to implement and make available five PETs as representatives of PET classes from literature and real applications. Since we want to observe which social groups are allowed to see which personal information, our mechanisms include approaches where a user can say when information is disclosed to certain groups. More specifically, the study participants had to

name persons belonging to each of the social groups *teacher, parents, classmates, friends* and *acquaintances*.

$PET_{checkbox}$ *(Public/Private Checkbox):* To observe how our study participants make use of simple and intuitive privacy mechanisms, we have implemented a straightforward checkbox on the XDA that allows or prohibits disclosing content, locations and metadata to anybody. The default is 'private'.

PET_{fine} *(Fine-grained Control):* This PET allows to specify privacy preferences for each piece of information. It is part of our web application, i.e., a user can define her privacy settings after she has generated the tag. Checkboxes let the participants define for each location, tag, real name, user pseudonym and time which social group is allowed to see the information. For example, a participant can specify 'Make the tags *my school* public for my friends, but hide my real name and the creation time.'. We ensured that it requires exactly the same effort to make a piece of information public or private.

In order to observe the privacy preferences and the privacy mechanisms preferred for tracks, we have implemented three mechanisms. Furthermore, we have adapted PET_{fine} so that it can be used on fragments of a track.

PET_{areas} *(Private Areas):* This PET lets the user specify areas where a track will not be disclosed to others. Thus, when a user enters a private area, the part of her track visible for others ends at the border of the area. The users can configure this for each social group by using an intuitive Google Maps mashup, integrated in our web application. The mashup lets the participants draw closed polygons, each representing a private area. Examples of such areas span from small rectangles around the user's home to polygons that cover the entire world. Private areas are specified in advance, and PET_{areas} takes effect at the moment of data acquisition on the XDA.

PET_{switch} *(GPS Switch):* This mechanism is motivated by [9] that request a simple option to turn off location tracking. It is similar to the PET implemented in Google Latitude. PET_{switch} lets the participants (de-)activate the GPS receiver, i.e., it provides immediate control over the disclosure of track information. However, it also requires permanent attention to deactivate the GPS whenever entering a sensible region.

PET_{anon} *(Anonymizer):* This PET, motivated by [5], deploys the notion of k-anonymity so that queries for tags at a location nearby cannot be distinguished from queries of k-1 other users. The anonymizer works as follows: Before issuing a query, the user specifies a k, i.e., a number of users from which she wants to be indistinguishable. Now suppose a user issues a query for tags in her vicinity. The anonymizer extends the query region so that at least k other persons could have sent the same query. The LBS answers with the tags for the extended area, which the XDA filters for the original query region. It is known that this algorithm does not prevent the construction of precise tracks for (i) frequent (continuous) requests (therefore participants had to refresh tags shown manually) and (ii) in populated regions where the extended areas are small. Note that

literature, e.g., [5], provides a minimum region size and an adaptive location optimizer to avoid these problems. Others group 'similar' tracks. However, we only consider k to not burden the participants with too many parameters. Using this PET properly requires a fundamental understanding of its operation principles and of the privacy threats of LBS.

4 Study Design

In the following, we describe our key design decisions and how we have structured and conducted our user study.

4.1 Key Design Decisions

The design of the study is based on five key decisions:

Intensive Study. We have decided to run a study requiring a high degree of participant involvement, under real conditions, with a relatively small number of individuals. The rationale is as follows: (1) Competence regarding the technology is required to obtain meaningful study results [18]. Thus, we have to *intensively train* our participants on GPS, the XDAs, and our applications. (2) Participants of offline surveys tend to overestimate their privacy needs [1]. We expect more realistic results if our participants use an application in their *daily lives* in a natural way; and we observe their behavior.

Two Phases of Evaluation. We have tested PETs in two phases: (i) A tagging phase where we focus on locations, content and metadata disclosed, and (ii) a track-recording phase where we analyze the behavior of the participants for each PET in isolation. We do so because some PETs interfere with others, and we do not want to overstrain the participants with several complex PETs active at the same time.

Web Application. We have designed and implemented parts of our setup as a web application, for two reasons. First, some PETs need a detailed parameterization. To avoid that unintuitive user interfaces bias our results, we have implemented the preference dialogs of PET_{fine} and PET_{areas} as a web application. Second, we have invited persons from the social environment of our participants to observe the web application in order to generate real privacy threats.

Participants. Typically, tagging services are used to facilitate retrieval, for opinion expression, sharing, play and competition etc. [19]. Thus, our study requires technology-affine individuals who are part of the same social network, i.e., who live in a nearby area to see the tags of friends and interact with each other frequently. Furthermore, our study participants should be a relevant target group for LBS providers. We decided for a German upper grade high-school class of 25 students aged between 16 and 17 years. We equipped 10 of them (eight female and two male) with our XDAs. In order to come up with real

privacy threats, we provided the parents, teachers and other pupils who did not receive a XDA with access to our web application. We have equipped parents and teachers with accounts. Thus, many persons from the social environment of our participants could see content, metadata, locations and tracks which participants have made public. For contacts with individual accounts in turn, participants could differentiate in more detail what to reveal.

Incentives. To encourage participants to use our LBS, we let them send one SMS for free for any 5 different locations tagged and for every 30 minutes the XDA was connected to the LBS. The payment was independent from the position or the kind of tag provided. Our participants have used the XDA and our LBS very frequently.

4.2 Study Procedure

We have divided our study into the four phases *Introduction*, *Tagging*, *Track Recording* and *Completion*. Each phase starts and ends with a meeting where we instruct the participants and ask control questions to ensure proper usage of the mechanisms. In the following, we will describe these phases. During the entire experiment, we record any keystroke and any GPS information from the XDA. Thus, our logs contain complete track information even if a participant has required that some tracks shall remain private[2]. More details on our study, e.g., the questionnaires used and screenshots of our web application, can be found on our complementary web page[3].

Introduction. We start with a questionnaire (Q.1) to obtain information on (i) demographic data, (ii) the usage habits regarding cellphones and Internet services, and (iii) general privacy attitudes. In order to provide a plausible motivation for our application without revealing our interest in privacy, we describe popular tagging applications and LBS, e.g., del.icio.us and mobile city guides. Furthermore, we introduce the XDA with the mobile location tagging application. We explain our application and guide participants to tag training objects at our university campus. After that, we demonstrate how those objects appear in our web application. Finally, we tell our participants that mobile tagging applications usually contain features to decide who may access which content. Therefore, we explain how to use the privacy mechanisms for tags ($PET_{checkbox}$, PET_{fine}). To this end, we let our participants name the members of their social groups and configure their buddy lists accordingly.

Tagging. The Tagging Phase lasts one week. Without yet revealing our interest in privacy, we encourage our participants to use the XDA in their everyday life

[2] Note that this has been the only chance to really measure the gap between actual privacy preferences and stated ones. We have informed their teacher on the privacy focus of our study in advance, and we did not make any track accessible to others before activating the PETs. Further, all parents gave their consent that their children participate in the location tagging experiment.

[3] http://privacy.ipd.uka.de

and to tag different objects. We give examples like sights, meeting points, own house and house of friends. We secretly record all movements of the XDAs, i.e., we log participant tracks. At the end of the week, we reveal our research interest in location privacy to the study participants. First, we hand out a questionnaire (Q2.A) on the usability of our tagging application, technical problems that might have biased results up to this point and on the incentives. Afterwards, we explain that privacy threats do not only arise from the tags and locations provided, but also from tracks that a mobile tagging application can record.

Track Recording. At the beginning of this phase we hand out a questionnaire (Q2.B) asking for all situations and locations where our participants do not want their tracks to be visible for others. We then activate the privacy mechanisms that allow the participants to hide their tracks. Note that we have trained our participants on the PETs but have not introduced them to different privacy threats regarding tracks. As some of the PETs are not orthogonal, e.g., defining private areas (PET_{areas}) and the GPS switch (PET_{switch}), we decided to evaluate each individual PET in isolation.

We start with PET_{areas} and ask the participants to specify private areas where their track may not be disclosed. Afterwards, to analyze if users have defined areas that meet their preferences, they are allowed to look at the tracks we have recorded during the tagging phase, and we let them refine their private areas if needed. In the following two days, our participants use PET_{areas} in their everyday life to hide track information that is supposed to be private. For another two days, we provide the GPS switch that activates and stops track recording (PET_{switch}). For the last two days, we let the participants use the anonymizer (PET_{anon}).

At the end of both the Tagging and the Track Recording phase, participants can use the fine grained control (PET_{fine}). With PET_{fine} users can specify who can see what for tracks recorded during the Tagging phase, and tracks that are not hidden by one of the other mechanisms during the Track Recording phase.

Completion. We finish our study with a questionnaire (Q.3) on the study and on the privacy mechanisms offered.

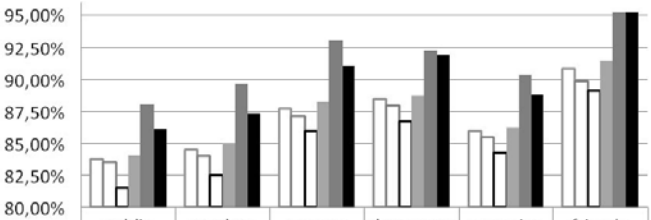
5 Study

In this section we will answer the research questions form the introduction.

5.1 Location, Content and Metadata

Our participants assigned 1042 tags to 442 different locations, i.e., they have provided 442 <*location, tag list, metadata*>-tuples that make locations of their everyday life explicit. Most locations have been tagged with 1 to 3 tags. This is in line with other tagging applications [20]. 41% of all locations were tagged with two or more tags.

Which Information Do People Disclose? (Q1). To find out which information people would like to share with others, we evaluate which tags and metadata our participants have made publicly visible by $PET_{checkbox}$ and PET_{fine}. Our expectation is that our participants want to disclose most of the information.



	public	teacher	parents	classmates	acquaint	friends
date	83,74%	84,48%	87,68%	88,42%	85,96%	90,89%
time	83,50%	83,99%	87,19%	87,93%	85,47%	89,90%
full name	81,53%	82,51%	85,96%	86,70%	84,24%	89,16%
pseudonym	83,99%	84,98%	88,18%	88,67%	86,21%	91,38%
tags	88,08%	89,67%	93,05%	92,25%	90,37%	95,23%
locations	86,13%	87,28%	91,04%	91,91%	88,73%	95,23%

Fig. 1. Content and metadata disclosed

The Column 'public' in Figure 1 shows how often information in the various categories has been disclosed. The figure distinguishes between the date and time tags have been created, the name and the pseudonym, the tag per location, and the location itself. The figure shows that our participants wanted 88% of all tags generated and 86% of all locations tagged to be visible for anybody. 81% of all <*location, tag list, metadata*>-tuples are marked public, and 2% are specified private in their entirety.

The participants have made fine distinctions regarding the disclosure of information from the remaining tuples, i.e., they disclosed only some attributes from a <*location, tag list, metadata*> tuple. In 10% of all tuples, only some tags from the tag list were made public to anybody. In 4% of the tuples, only the location and the tags are disclosed, but all metadata is kept private. 3% of the tuples reveal only some of the metadata. Creation date, creation time and user pseudonym have been marked public nearly as often as the tags. Our participants disclosed their real name least often. None of our participants has considered that linking data from multiple tags with different privacy settings might allow to infer data marked private only for some tuples. For example, if a participant has ever revealed her pseudonym together with her real name, the real name can be reconstructed for tuples where only the pseudonym was public.

Leaking information on important locations of everyday life can be a severe privacy threat. In order to find out if participants let others know their centers of life, we have measured the time a person stays in the vicinity of a location she has made public to someone else. Figure 2 shows the logarithmic cumulative

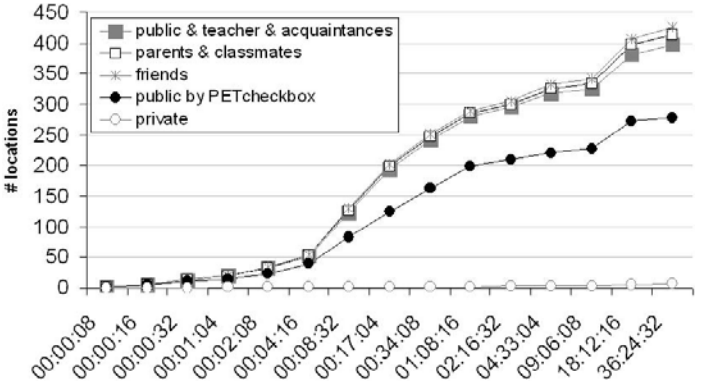


Fig. 2. Time users spent in the vicinity of their tags – time.log(hh:mm:ss)

distribution function of how long participants stayed in a distance of less than 200m from a location they have tagged. 20% of all tuples refer to locations where the user has stayed for at least 20 hours during 14 days of our study. Such locations include the school building or the home of the participant. Remember that our participants made 2% of all tuples private. 50% of these private tuples refer to locations where its creator spent at least 20 hours.

These findings lead to the following conclusions: (1) Instead of disclosing a minimum of information, people tend to disclose everything to everybody. Only very sensitive information is kept private. (2) People do not think that joining data might lead to privacy leaks. This calls for automated PETs which warn the user if she is about to disclose privacy-threatening combinations of information. (3) Even if many tuples were disclosed in an all-or-nothing manner, there is a need for PETs that specify in a fine-grained way which particular information is publicly visible. (4) The time spent in the vicinity of tags disclosed indicates that people frequently disclose important locations of their everyday life.

Which Social Groups Are Allowed to See Private Information? (Q2). PET_{fine} lets the participants distinguish not only between kinds of data, but also between social groups, namely teacher, parents, classmates, acquaintances and friends. Figure 1 shows which groups are frequently allowed to see which information. Remember that the participants made 81% of all data visible to anybody and kept 2% entirely private. Thus, each social group was allowed to see at least 81% of all tuples (public). Our participants had disclosed most to the group friends and least to the group teacher. When comparing our initial questionnaire with Figure 1 we found some inconsistencies. In the questionnaire, 6 of our 10 participants did not want to reveal their whereabouts to their parents, 2 were undecided, and the remaining 2 were inclined to tell their locations. However, our study results show that parents are allowed to see between 86% and 93% of all information, depending on the kind of data. For all social groups the correlation coefficient between disclosed metadata and disclosed locations is 99%, for tags 97%.

We have provided our participants with a list of example locations which they could tag, including their home and houses of friends. As expected, all participants have tagged their home but even four of them have made this tag public. 8 of 10 participants tagged at least one house of a friend. However, only one participant made these tuples public, and one disclosed it to friends. This is in line with our preceding questionnaire, where 8 of the participants stated to care about the privacy of friends.

Figure 2 shows the time the participants have spent close to the locations they have tagged, and which social groups were allowed to see this location. Thus, the figure tells who might observe important locations of the daily life. To avoid clutter in the figure, there is one curve representing several groups with very similar values. 16% of all <*location, tag list, metadata*>-tuples describe where the creator of the tuple stayed for at least 20 hours during our study. Friends are allowed to see 95% of all locations. 19% of these locations are close to places where the issuer of the tuple stayed at least 20 hours.

Summing up, (1) our participants differentiate between social groups for 17% of all information provided, and (2) behave similarly regarding the different social groups and metadata disclosed. Furthermore, (3) they care for privacy of certain others more than for their own privacy.

How Do People Use PETs, and What Kind of PETs Do They Prefer? (Q3, Q4). Our participants had three options to keep their privacy: (i) They could use a checkbox on the device that makes an entire tuple public ($PET_{checkbox}$). (ii) They could use our web application to specify in detail which group is allowed to see which information from a tuple (PET_{fine}). (iii) Our participants could abstain from providing information, if they deem it too cumbersome to use the PETs provided. We have observed that the participants have made 63% of all tuples publicly visible using $PET_{checkbox}$. The privacy settings of 37% of the tuples were specified with PET_{fine}. The participants used PET_{fine} to make an additional 23% of all locations public. Thus, though our participants prefer simple mechanisms, they readily use sophisticated PETs. Furthermore, in many cases, e.g., for their homes or for houses of friends, they spent effort in providing detailed privacy settings, instead of simply not providing information. Thus, there clearly is a need for sophisticated PETs.

5.2 Location Tracking

In the following, we will compare the usage of PET_{areas}, PET_{switch} and PET_{anon}, followed by an analysis of the tracks that have been made visible.

How Do People Use PETs, and What Do They Disclose? (Q1, Q3). To find out how our participants use the PETs, we look at the track information we have recorded.

PET_{areas} (Private Areas). This PET lets the participants define areas where tracks are not shown to others, e.g., not to 'anybody'. We have obtained private area definitions at the beginning and at the end of the track recording phase.

At the beginning, we handed out a questionnaire where the participants specified private areas in plain text. Furthermore, our participants had to draw these areas in a Google Maps mashup. Our participants defined 26 private areas, which are related to leisure activities (11), home (6), school (5), (boy)friends (2), work (1) and relatives (1). Table 1 shows how many areas mask a track, the average area size and its standard deviation. In order not to bias our statistics, we have excluded two outlier areas that cover the country and the city. Note that our participants have removed those areas in the real usage phase. To measure if the areas cover hot spots of the daily lives, we have also calculated for how long our participants stayed in one of the private areas. The initial areas cover regions where the participants stay for 25% of the time they have used our system.

We obtained refined areas at the end of the track recording phase, i.e., after the participants have seen the tracks recorded in the tagging phase, and after two days of using PET_{areas}. The refinements include 6 new areas that prevent the disclosure of tracks to anybody. On average, our participants more than doubled the size of their initial areas. One area was increased by factor 16 after a participant realized that an area of the size of a house allows to see when entering the building. Overall, we found that the area specification includes 29 areas which the participants actually pass from time to time. When comparing the time the participants have spent in private areas to the time spent in the initial areas (for the entire experiment time recorded), it increased from 24.5% to 33.2%.

Table 1. PET_{areas}: private areas initially and refined

Initial Areas	anybody	teacher	parents	friends	acquaint.	classmates
Number of Areas	12	23	16	16	18	19
Avg. size (km^2)	0.712	1.121	0.578	1.576	0.504	1.339
Stddv. size	1.191	3.496	1.051	4.146	1.007	3.826
Time in areas (%)	24.5%	26.9%	24.8%	26.9%	25.0%	26.5%
Refined Areas	anybody	teacher	parents	friends	acquaint.	classmates
Number of Areas	18	28	23	20	22	25
Avg. size (km^2)	0.984	0.720	0.823	0.966	0.832	0.806
Stddv. size	1.494	1.258	1.356	1.418	1.387	1.307
Time in areas (%)	33.2%	40.2%	33.9%	34.7%	34.0%	34.3%

Our results show that people wish to adapt their privacy settings frequently. Furthermore, people cannot imagine location based privacy threats without seeing their tracks. The fact that our participants significantly increased their areas after we confronted them with track recordings supports this.

PET_{switch} (GPS Switch). The second mechanism we evaluated is the manual on/off switch for the GPS receiver. We have logged when and where the GPS has been turned off, i.e., when the participants wanted to keep their movements

private. We expected our participants to use the switch when entering one of the private areas they have defined for PET_{areas}. But the GPS switch was rarely used. During two days, the participants used the switch 6 times on average, and each of them used it less than 12 times.

To evaluate PET_{switch}, we assume that the area definitions from PET_{areas} are an exhaustive specification of the private areas. As Table 2 shows, our participants have made 44% of their tracks visible on average while beeing in a private area, i.e., when they wanted to be unobserved.This indicates that PET_{switch} fails in practice. Thus, PETs that require continuous attention might lead to unintended disclosure of personal information, even if they are as simple as a switch for a GPS receiver. Our final questionnaire supports this finding.

Table 2. Comparing PET_{switch} to PET_{areas}

GPS / Position	in private area	out of area	sum
active (%)	43.56%	31.71%	75.27%
inactive (%)	16.45%	8.27 %	24.73%
sum (%)	60.01%	39.99%	**100.00**%

PET_{anon} (Anonymizer). This PET adapts k-anonymity to avoid that sequences of positions can be combined to a track. But the choice of an appropriate k requires technical understanding. Thus, we have explained PET_{anon} in detail. Note that the participants expected that others would use the application, i.e., values greater than 10 have been possible.

Table 3. PET_{anon}: Average distance to the k-nearest neighbor

k	2	3	4	5	6
distance (km)	1.06	1.67	2.00	2.35	2.72
k	7	8	9	10	–
distance (km)	3.17	3.83	5.79	13.43	–

Our participants have chosen a k that varies between 1, i.e., no anonymity, and 30. The average k was 7.33 with standard deviation 8.64. In order to find out which value of k is appropriate, we calculate the average distance between each of the 14,015 positions recorded and the k nearest neighbor. For example, Table 3 shows that the average distance to the 3rd neighbor is 1.67km. Thus, a position anonymized with $k = 3$ means that the user can be located anywhere in a circular area[4] of $8.87km^2$ on average. The average k corresponds to a circle of $38.5km^2$, which is much more than the average area size of $0.7km^2$ specified

[4] We have used a circular area just to provide an intuition. [5] uses rectangular grid cells interleaved to a pyramid scheme.

with PET_{areas}. Thus, we conclude that our participants are unable to use the anonymizer in line with their preferences.

Which Social Groups Are Allowed to See Private Track Information? (Q2). PET_{fine} and PET_{areas} allow to restrict the visibility of parts of a track for certain social groups. At the end of the study, 5 of our participants have used PET_{fine} to assign privacy preferences to 270 tracks. The other half has not disclosed any track information. Only 28% of all tracks have been made visible for others. Table 4 shows which social groups these 28% of the tracks have been disclosed to. Only 2% of all tracks have been made visible for anybody, 4% for teachers; parents could see 21%, friends 51%.

Table 4. Tracks disclosed using PET_{fine}

Phase	pub	teacher	parents	friends	acqu.	classm.
Tagging (%)	2%	4%	13%	28%	7%	10%
Track Rec. (%)	0%	0%	8%	23%	2%	4%
Sum (%)	2%	4%	21%	51%	9%	14%

PET_{areas} lets the participants configure who is not allowed to see a track recorded in a private area. Table 1 compares the number of areas defined, their size and the time the users have spent in it, for each social group. Our participants have defined most areas to hide from teachers (28) and fewest to hide from friends (20). This is in line with our findings from PET_{fine}. When considering the time spent in a private area, our participants do not want to be seen from their teachers for 40% of the time. For all other social groups, the participants want not to be seen for less than 35% of the time.

Our evaluation supports two findings. First, people seem to be more concerned about revealing tracks than disclosing locations, content and metadata. While the participants make 81% of the latter public, they disclose hardly any tracks to anybody. Second, there is a difference in the track information disclosed with PET_{fine} and PET_{areas}. As the disclosure rate is larger for the PET that lets the users see the tracks they are about to reveal, we conclude that people are not aware of track based privacy threats. This is in line with our findings on the usage of the GPS switch.

Which Kind of Privacy Mechanisms Do People Prefer? (Q4). To answer this question, we evaluate a final questionnaire. We ask our participants to assign a score of 1 to 5 in categories safety, ease of usage, complexity to understand, and effort to use to PET_{areas}, PET_{switch} and PET_{anon}. The average scores are displayed in Figure 3. Furthermore, we ask our participants to rank PET or PET combinations they would like to use in the future. Figure 4 shows the average rank.

Among all PETs in isolation, our participants prefer the GPS switch. They assigned it the highest grades in all categories. PET_{areas} obtained the worst

grades in all categories. Since our participants noticed that the GPS switch needs permanent attention, they prefer combining it with other PETs. Here, our participants preferred the GPS switch together with the anonymizer. We found this surprising, because our participants neither used the GPS switch nor the anonymizer in line with their privacy preferences. A possible explanation is that the participants wanted to combine the most intuitive PET with a PET that needs only one parameter.

5.3 Discussion

It has been our design decision to execute an elaborate study under real conditions with a limited number of well-prepared participants. Nevertheless, we deem our study results representative, for three reasons:

Relevant study group. Our study included 25 individuals, their parents and one teacher. We equipped 10 of them with mobile devices. Our participants were interested in new technologies, all are used to cellphones and Internet applications, and they are mobile in the sense of visiting many different locations instead of having a fixed schedule each day. Furthermore, our participants are a relevant target group for commercial mobile applications.

Real application. Our results have been obtained with a geotagging application similar to a wide class of services. This includes all applications where users share location-based content related to a position that has been visited before or is visited currently. The study participants displayed usage patterns comparable to those described in the literature [19]. At the end of our study, some participants asked if they could continue using our application.

Real privacy threats. The participants integrated our application into their everyday life. The study period included working days, holidays and weekends. The average usage time of our application was 6 hours per day, i.e., the usage intensity was very high. During the study, the XDAs logged spatial resolutions of the GPS receiver that were better than 100m. This is sufficient to distinguish the individual buildings the participants have entered.

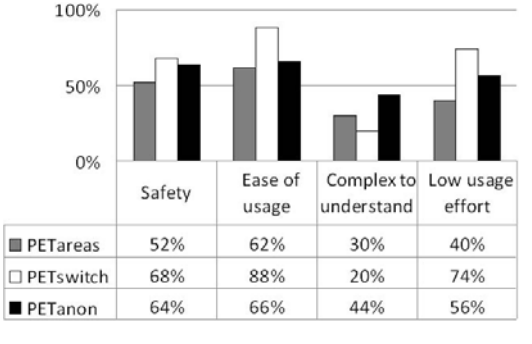


Fig. 3. Comparison of Mechanisms

PET Ranking	∅
GPS switch & Anonymizer	2,9
All mechanisms combined	3,1
Private Areas & GPS switch	3,4
Private Areas & Anonymizer	3,9
GPS switch	4,2
Anonymizer	4,8
Private Areas	5,4

Fig. 4. PET Ranking

We summarize the lessons learned from our study as follows:

Social groups. Participants frequently wanted to disclose information to certain groups, e.g., to friends or classmates. Thus, we recommend PETs that allow a differentiation between social groups.

Different information. LBS do not only process locations and tracks, but other kinds of data as well, e.g., content or metadata. PETs for LBS have to consider such information, because privacy threats arise from any data, and people frequently want to disclose a subset of information, e.g., a position without the date of creation.

Complex PETs. For transparency reasons, it is often assumed that users prefer PETs which are easy to understand. However, we have shown that people favor complex, sophisticated PETs over simple ones – if the sophisticated PETs reduce the effort of the users to have their privacy preferences enforced.

6 Conclusions

Location based services (LBS) are an important recent development. However, LBS put user privacy at risk. In this paper, we have investigated which privacy mechanisms individuals want to use, and how. To this end, we have implemented a fully operational geotagging service, and we let our study participants use this application in their daily lives.

Among other results we found that people tend to use mechanisms that are easy to understand. However, when mechanisms require constant awareness, they fail in practice, i.e., cannot serve all privacy needs. Thus, our participants wanted to combine them with automated approaches that allow a fine-grained control over the data disclosed. Although people tend to disclose most information, we have observed for 17% of the data provided that our participants wanted to specify who is allowed to see it in a very detailed manner. Finally, we found that people care for the privacy of their friends even if they disclose everything about themselves.

Acknowledgements. This work was partly funded by DFG BO2129/8-1. We thank Ursula Kotzur and the authors of [15] for supporting our study.

References

1. Spiekermann, S., Grossklags, J., Berendt, B.: E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In: EC. ACM, New York (2001)
2. Gruteser, M., Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking. In: MobiSys. ACM, New York (2003)
3. Kido, H., Yanagisawa, Y., Satoh, T.: Protection of location privacy using dummies for location-based services. In: ICDEW. IEEE, Los Alamitos (2005)
4. Gedik, B., Liu, L.: A customizable k-anonymity model for protecting location privacy. Technical report, Georgia Institute of Technology (2004)

5. Mokbel, M.F., Chow, C.Y., Aref, W.G.: The new casper: query processing for location services without compromising privacy. In: VLDB (2006)
6. European Parliament: Directive 2002/58/ec (2002)
7. Westin, A.F.: Privacy and Freedom. Bodley Head (April 1970)
8. Ackerman, M.S.: The intellectual challenge of cscw: The gap between social requirements and technical feasibility. In: HCI. ACM, New York (2001)
9. Barkhuus, L., Dey, A.: Location-based services for mobile telephony: A study of users' privacy concerns. In: CHI INTERACT (2003)
10. Consolvo, S., Smith, I.E., et al.: Location disclosure to social relations: Why, when, & what people want to share. In: SIGCHI. ACM, New York (2005)
11. Xu, T., Cai, Y.: Location anonymity in continuous location-based services. In: GIS, pp. 1–8. ACM, New York (2007)
12. Terrovitis, M., Mamoulis, N.: Privacy preservation in the publication of trajectories. In: MDM. IEEE, Los Alamitos (2008)
13. Lederer, S., Mankoff, J., Dey, A.K.: Who wants to know what when? privacy preference determinants in ubiquitous computing. In: CHI. ACM, New York (2003)
14. Smith, I., Consolvo, S., et al.: Social disclosure of place: From location technology to communication practice. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 134–151. Springer, Heidelberg (2005)
15. Wind, R., et al.: A testbed for the exploration of novel concepts in mobile service delivery. In: MDM (2007)
16. Ames, M., Naaman, M.: Why we tag: motivations for annotation in mobile and online media. In: SIGCHI. ACM, New York (2007)
17. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature 453 (2008)
18. Babbie, E.R.: The Practice of Social Research, 10th edn. Academic Internet Publ. (2007)
19. Marlow, C., et al.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In: HYPERTEXT. ACM, New York (2006)
20. Millen, D.R., Feinberg, J., Kerr, B.: Dogear: Social bookmarking in the enterprise. In: SIGCHI. ACM, New York (2006)