Pervasive and Mobile Computing 25 (2016) 125–142

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

A machine-learning based approach to privacy-aware information-sharing in mobile social networks☆

Igor Bilogrevic[a,*,1], Kévin Huguenin[c,1], Berker Agir[b], Murtuza Jadliwala[d], Maria Gazaki[b], Jean-Pierre Hubaux[b]

[a] Google, 8002 Zurich, Switzerland
[b] School of Computer and Communication Systems, EPFL, 1015 Lausanne, Switzerland
[c] LAAS-CNRS, 31400 Toulouse, France
[d] Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, KS 67260, USA

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form 18 December 2014
Accepted 23 January 2015
Available online 31 January 2015

Keywords:
Information-sharing
Decision-making
Machine learning
User study
Privacy

ABSTRACT

Contextual information about users is increasingly shared on mobile social networks. Examples of such information include users' locations, events, activities, and the co-presence of others in proximity. When disclosing personal information, users take into account several factors to balance privacy, utility and convenience — they want to share the "right" amount and type of information at each time, thus revealing a selective sharing behavior depending on the context, with a minimum amount of user interaction. In this article, we present SPISM, a novel information-sharing system that decides (semi-)automatically, based on personal and contextual features, whether to share information with others and at what granularity, whenever it is requested. SPISM makes use of (active) machine-learning techniques, including cost-sensitive multi-class classifiers based on support vector machines. SPISM provides both ease of use and privacy features: It adapts to each user's behavior and predicts the level of detail for each sharing decision. Based on a personalized survey about information sharing, which involves 70 participants, our results provide insight into the most influential features behind a sharing decision, the reasons users share different types of information and their confidence in such decisions. We show that SPISM outperforms other kinds of policies; it achieves a median proportion of correct sharing decisions of 72% (after only 40 manual decisions). We also show that SPISM can be optimized to gracefully balance utility and privacy, but at the cost of a slight decrease in accuracy. Finally, we assess the potential of a one-size-fits-all version of SPISM.

1. Introduction

Mobile devices are becoming the gatekeepers of people's digital selves. More and more personal and private information is stored, shared and managed on-the-go. Having access to people's personal data and physical contexts (through an

☆ This article is a revised and extended version of a paper that appears in the Proceedings of the 15th ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013) Bilogrevic et al. (2013).

* Corresponding author.
E-mail addresses: igor.bilogrevic@gmail.com (I. Bilogrevic), kevin.huguenin@laas.fr (K. Huguenin), berker.agir@epfl.ch (B. Agir), murtuza.jadliwala@wichita.edu (M. Jadliwala), maria.gazaki@epfl.ch (M. Gazaki), jean-pierre.hubaux@epfl.ch (J.-P. Hubaux).

1 Parts of this work were carried out while the author was with EPFL.

http://dx.doi.org/10.1016/j.pmcj.2015.01.006
1574-1192/

increasing number of embedded sensors), mobile devices represent a simple means to quickly share information with others via mobile social networks, such as Facebook, WhatsApp or Google, without the need for manually typing their current locations; location and photos are just two examples of data that can be easily shared. In addition to the user-triggered sharing decisions, applications such as Foursquare and the now-closed Gowalla enable users to configure their smartphones to share their location and co-presence automatically, in a push-based fashion. With a small set of default information-sharing policies, users have the possibility to adjust the settings in order to match their sharing behaviors with their privacy concerns.

Usually, there are several behavioral and contextual factors that influence users when they share their personal information, as extensively shown in [1–5]. By analyzing people's sharing behaviors in different contexts, it is shown in these works that it is possible to determine the features that most influence users' sharing decisions, such as the identity of the person that is requesting the information and the current location [2]. For instance, tools such as the location-sharing systems Locaccino [6] and PeopleFinder [4] have been used to gain significant insight into the benefits of providing users with the ability to set personal sharing policies. Two recurrent findings in studies related to information sharing are that (i) users are not particularly good at effectively articulating their information-sharing policies (compared to their actual behavior) [4] and (ii) that sharing policies evolve over time [6,4].

In order to overcome these two issues, machine-learning techniques have been applied to improve to some extent the decision-making process [7,8,4,9]. The advantage of such systems is that they can decide, in a semi-automatic fashion, whether or not to share information. Most existing schemes, however, enable users to share only a specific kind of information (e.g., location). Moreover, they only make binary decisions on whether to share the requested information. This last issue in particular, is often mentioned as a crucial catalyst for overcoming concerns related to privacy [10,11] and to a more open, sharing behavior.

In our work, we perform a comprehensive study of information-sharing in mobile social networks, by tackling, all at once, the issues related to context, user-burden and privacy trade-offs. We introduce SPISM (for Smart Privacy-aware Information Sharing Mechanism), a novel *pull-based* information-sharing system (i.e., users explicitly request information from their friends) implemented on Android; it decides in a semi-automatic fashion, whether or not to share information and the level of detail of the information to be shared with other users or services, based on personal and contextual features and past behavior. SPISM makes use of a cost-sensitive multi-class classifier, typically based on naive Bayes (NB) or non-linear Support Vector Machines (SVM) with Gaussian or polynomial kernels, fed with different contextual features including the time of day, the current location of the user and the identity of the requester. The cost-sensitive aspects enable SPISM to find an appropriate balance between over-sharing (i.e., unduly sharing the requested information) and under-sharing (i.e., unduly retaining the requested information). The multi-class aspects of the classifier enable SPISM to decide the correct level of detail of the shared information (e.g., whether to share the name of the street or the city the user is in). The decision-making core is supported by an active learning method that enables SPISM to either decide automatically – whenever the confidence in the decision is high enough (based on the entropy of the distribution, computed over the possible decisions of the classifier) – or to rely on the user's input otherwise. As such, SPISM continuously learns from the user and, over time, it requires less and less user-input. Note that, in this paper, we do not develop new machine learning techniques. Instead, we leverage on appropriate existing machine learning techniques to provide useful features for protecting the users' privacy while reducing their burden. As such, the main contribution of this paper is the design and implementation of a feature-rich decision-making mechanism and the evaluation of its performance, in addition to the analysis of users' sharing behaviors, based on a novel and rich dataset. SPISM can work with any existing (mobile) social networks and could even be used transparently by users, as it can operate at the operating-system level, filtering all requests for personal information from mobile apps and websites and replying according to the user's behavior.

The contributions of this work are as follows. First, we develop a novel information-sharing system (SPISM) for (semi-) automatic decision-making in mobile social networks: It enables users to share different types of information (location, activity and co-presence of other people) with other users or services in a privacy-aware fashion. Second, we conduct a personalized online study which involves 70 participants where, in addition to collecting data about their sharing behaviors, we provide insights into two other crucial factors in studies related to information sharing [12]: The *reason* behind a decision to share and the confidence that the user has in her decision. Third, we evaluate SPISM with respect to the amount of training data (provided by the user) and its performance, and we compare it against two policy-based mechanisms. Our results show that SPISM significantly outperforms individual privacy policies specified by users, and it achieves a median proportion of correct binary sharing decisions (i.e., whether to share the requested information) of 72% when trained on only 40 manual decisions. Moreover, SPISM is able to infer the level of details at which the information should be shared (e.g., street-level accuracy vs. city-level accuracy for location information) with a median proportion of correct decision of 60%. We also demonstrate the advantages of active learning techniques in our setting. Fourth, in this substantially extended article (as compared to [13]), we enrich SPISM with a new functionality that significantly enhances the privacy of the users while incurring a negligible decrease of performance and no side-effects on the usability. We also assess the potential of a one-size-fits-all model for decision making.

The rest of the paper is organized as follows. We survey the related work in Section 2, and we present the SPISM information-sharing platform in Section 3. We describe the survey methodology and discuss the participants in Section 4. We present the performance results in Section 5, and we conclude this article in Section 6.

2. Related work

A substantial research effort has been made on the topic of privacy and information sharing in mobile social networks, notably with respect to the attitudes of people when sharing static and contextual data with other users. The studies that are most related to our work can be grouped, from a high-level perspective, into two categories: (i) contextual information sharing and privacy [1–3] and (ii) machine learning for information sharing [7,8,14–16,4].

2.1. Contextual information sharing and privacy

Smith et al. [1] provide an early investigation on technologies that enable people to share their contextual information, such as location, in mobile social networks. In addition, to enable users to manually decide when to share their location with others, the authors implemented a system called *Reno* that can automate the process based on a set of pre-defined regions. By allowing Reno to automatically send notifications whenever the user enters or exits such regions, the authors show that there are both a value and a cost associated with automatic information disclosure. In particular, they show that static rules for location sharing in pre-defined regions are ineffective in accurately expressing the users' actual behavior when other contextual elements change, such as the time of the day or the day of the week. By taking into account such limitations in our work, we consider a wide set of contextual features (discussed in the "SPISM Information-Sharing Platform" section) in order to increase the flexibility of the decision-making process.

More recently, Toch et al. [2] have studied the effect of the type of locations visited by the users on their willingness to share them with others. By considering simple statistical models that take into account factors other than the geographic location, the authors show that the semantic category of the location being shared (such as a shopping center or a hospital) and the social group of the person asking for the location are significant factors in deciding whether to share the location. These results support earlier efforts [17,18,4] in providing a set of contextual features that have a statistically significant impact on the location-sharing behavior of mobile users. We use these results for our application when defining initial universal sharing policies, and we will describe them in the "Evaluation" section.

In an attempt to capture the cost of mistakenly revealing a location due to ineffective sharing policies, in addition to sharing preferences, Benisch et al. [3] compare simple access control policies (white lists) to more sophisticated ones (based on time, day and location). They found out that (i) the accuracy of the sharing policies increases with their complexity (or flexibility), and that (ii) the accuracy benefits are the greatest for the highly sensitive information. This suggests that the cost of mistakenly revealing information to unauthorized parties (in particular contexts) is an important factor in designing and optimizing automated information-sharing mechanisms. As users have varying degrees of sensitivity to privacy in online services [19], our evaluation integrates the notion of cost of unwanted release of private information to other users or services.

Wiese et al. [20] investigate the effect of physical and perceived social closeness on people's willingness to share information with others. Among the main results of the study, social closeness and the frequency of communication are shown to be better predictors of sharing than physical proximity. Moreover, these two factors were also shown to have a capacity to predict sharing better than the social groups of the people asking for the information. Therefore, the authors suggest that automatic methods for inferring social closeness could be suited for accurate information-sharing decisions more than physical co-location, in the case of automated mechanisms (such as in [21–24]) are envisaged.

2.2. Machine learning and information sharing

Whereas studies on information-sharing attitudes and privacy shed light on the behavior of people and the factors that influence their decisions, they focus mostly on the causes and effects of such behavior. Meanwhile, there has been a substantial effort in devising methods that help and nudge the users to make information-sharing decisions, or that even make decisions on their behalf. We present some of these methods, including both supervised and unsupervised approaches for decision-making.

In [4], Sadeh et al. compare the accuracy of user-defined sharing policies with an automated mechanism (case-based reasoner) and a machine-learning based approach (random forests), showing that these approaches have an accuracy better than the user-defined policies. Owing in part to the greater flexibility of the supervised machine-learning approaches compared to the more coarse-grained user-defined policies, the automated methods also benefit from the fact this users appear to not be able to create sharing rules consistent with their own choices. The feedback provided by the users to the machine-learning methods did however appear to be consistent with their actual sharing behavior, which helped the automated methods to achieve better accuracy results. We include the user feedback in our learning mechanism and use it to adapt the automated decisions to the user behavior that can change over time.

Unsupervised or semi-supervised methods, which reduce the initial setup burden of the default sharing policies for each user, are investigated in [7,8]. For instance, Danezis [7] proposes a method for automatically extracting privacy settings for online social networks; the method is based on the notion of a limited proliferation of information outside of a given social context. Their method, which determines cohesive groups of users where users belonging to a group have stronger ties to the users outside of the group, shows promising results on a limited set of evaluation samples. This study also shows that the

social groups, and especially methods for their automated extraction, are a key factor to sharing private information in social networks. Our work uses both the Facebook social graph and our system's contacts list to automatically extract social groups or communities; it uses them to relieve the user from the burden of manually assigning people to different social groups.

Fang and Lefevre [8] propose a novel approach to the inference and definition of access control policies for personal information on online social networks. They enable the supervised-learning mechanism to learn the sharing preferences of a user by asking her a limited number of questions about her sharing behavior with some of her friends; these specific friends are the most "informative", i.e., those for which the classifier is most uncertain about. The authors show that their approach of iteratively asking questions about the most uncertain case (active learning with uncertainty sampling) reduces the effort required by the users and maintains a high accuracy compared to the ground truth (based on a 45-user study on Facebook). Active learning is a feature that we exploit in our application as well. Moreover, we enable users to update their sharing decision *a posteriori*, which means that users are able to change their decisions after they have been made; the application then learns from these new decisions and takes them into account the next time the same context appears again.

Bigwood et al. [25] evaluate different machine-learning algorithms for information sharing in terms of information over-exposure and correct decisions. Although their work focuses exclusively on binary (yes/no) location-sharing, the authors provide a machine-learning-based determination of the most influential features for the sharing decisions; moreover, they take into account cost-sensitive classifiers to reduce over-exposure.

Schlegel et al. [26] propose a mechanism to show smartphone users summaries of the requests for their location information made by their contacts. The proposed mechanism displays a pair of eyes on the background of the home screen for each user who requested that user's location; the size of the eyes increases with the number of location requests sent by the requester. The mechanism also proposes to the users to adjust their privacy policies in an intuitive way, based on the reported information (typically, deny access to users who made too many requests). Their approach is motivated by the fact that the users' privacy concerns increase with the frequency at which their contacts access their locations, according to their study. Such a concern cannot be taken into account by static policies that do not rely on the frequency of location requests (which is the case of the policies specified on most location-based social networks); therefore, results from [26] suggest that the frequency of information requests should be used as a feature in a machine-learning-based approach.

Recently, Xie et al. [9] evaluate the effect of different contextual (incl. the semantics of the location) and personal features on location-sharing behaviors, and they propose a recommendation system for privacy preferences. Their work focuses on determining the audience of a given piece of information to share, whereas our work predicts, given a certain audience (i.e., the person or service who requests the information), the decision to share and the granularity of the shared information (if shared).

Liu et al. [27] address the problem of decision making for accessing private data (such as location) in the context of Android permissions. To do so, they build a decision profile for each user, cluster the users based on the similarities between their decision profiles, then build a single predictive model for each cluster of users. Unlike our work, the authors disregard both the social aspects (the requesters are applications, not users) and the contextual aspects (the decisions are static). Moreover, they consider only the basic case where users can either grant or deny permissions, thus ignoring the granularity of the shared information.

3. The SPISM information-sharing platform

In this section, we describe the functionality, the operating principle, the architecture and the design of our Smart Privacy-aware Information Sharing Mechanism (SPISM). In order to better understand the following, we need to distinguish between two different kinds of subscribers to SPISM: (i) the *requester*, who wants to know something about other subscribers by sending information requests, and (ii) the *target*, who receives requests for information. The SPISM platform is composed of the SPISM *application* that runs on mobile devices (as for now it is implemented only for the Android platform) and the SPISM *Information Sharing Directory (ISD)* that runs on a dedicated server.

3.1. Overview

The SPISM application enables subscribers, who can be users, third-party online services or mobile apps, to request information about other users. The information that can be requested includes contextual data (the geographic location and the wireless identifiers of physically co-located devices) and the time-schedule availability. The geographic location is determined by processing data obtained from the embedded GPS sensor (if available) or by WiFi tri-lateration (which relies on the Google localization service). The list of devices that are physically co-located with the target users is obtained through periodic scans of the Bluetooth and WiFi interfaces. If a MAC address in the vicinity of the target is a known MAC address (there exist an entry associated with a user in the contact list of the target), the name of the contact is displayed. Finally, the schedule availability is obtained from the user's calendar (accessed through the on-device calendar application). Subscribers can specify a level of detail for the requested information: low, medium or high. The information sent by the target user is provided with a level of detail lower or equal to the requested level. For the location, the coordinates are truncated; for the neighboring devices, the presence (i.e., some devices/no devices), the number, or the identifiers of the devices are provided; for the schedule availability, the availability (i.e., busy/available), the title or the detailed record of the calendar activity is provided. Fig. 1 shows the main application windows, where users can log in and register, request the

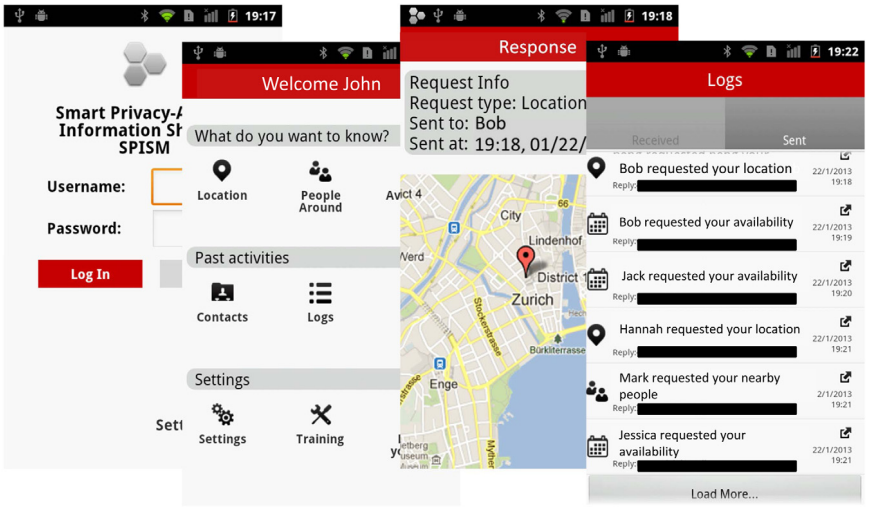


Fig. 1. SPISM mobile application interfaces. From left to right, the different windows enable users to register and log in, check other users' current location, the other devices around them, and their availability. The users can access other features such as the record of past activity and their contacts' lists.

location, the co-located devices and the availability of their contact, and enjoy additional features such as visualizing the past activity and their contacts' list.

3.2. System model

The SPISM platform is composed of the Information Sharing Directory (ISD) and the subscribers of the service, who can be either users or third-party online services. The roles of the ISD and of the subscribers are as follows:

- **ISD**: Its main purpose is to enable users to discover the current IP addresses of their contacts when they want to send them information requests (which are sent directly to the target user, without passing through the ISD; note that some existing software/protocols rely on such directories to obtain the addresses of the users' contacts and to subsequently establish direct connections, e.g., file transfers in Pidgin). The ISD stores the list of registered SPISM subscribers, their credentials, their contact lists and the MAC addresses of the Bluetooth interfaces of each user's mobile devices. The subscribers interact with the ISD in the registration phase (once per user), during the log-in phase (once per application start), when downloading the contacts lists, when periodically reporting their IP and updating their online status, and when sending information requests to one of their contacts.
- **Subscribers**: A subscriber, either an online service or a mobile user, can be a requester (when she sends queries to another subscriber) or a target (when she receives queries from other subscribers). In order to inform the ISD of her online status, each subscriber connected to the ISD sends periodic keep-alive messages. At any time, requesters can see the list of online and offline contacts, and they can choose to send queries to the online subscribers in their contacts list, in order to know their location, the devices around them and their availability. The requests that target subscribers receive and process are based on several features of their current physical and social contexts, including their current location, the time of the day and the people that are currently close by.

In order for SPISM to be more generic hence easily adopted, it can be implemented over an existing instant messaging protocol such as XMPP [28] (used by GoogleTalk[3] and Facebook Messenger[4]). We are currently working on an Android implementation of SPISM that uses XMPP and interacts transparently with existing XMPP-compatible instant messaging applications. Specifically, for users that do not use a SPISM-compatible messaging client, the received SPISM messages would either be hidden or displayed in plain text, for example, "Where are you?". For the users who use a SPISM-compatible client, the location request messages are handled and displayed properly, as explained below.

3.3. Operating principle

SPISM works as follows. A user first logs into the ISD with her username and password. She can subsequently report her online status and obtain the online status (and IP addresses) of her contacts from the ISD. In a typical scenario, the user requests some information from one of her (connected) contacts. To do so, the user first chooses the type of information she

3 https://www.google.fr/hangouts/.
4 https://www.facebook.com/mobile/messenger.

Table 1
Features used by the SPISM machine-learning framework to decide whether or not to share information and with what accuracy.

	Feature	Type
Who? Person	Familiarity	Float
	Social tie	Cat.
	User ID	Cat.
Who? Service	Service category	Cat.
	Service ID	Cat.
What?	Info type	Cat.
	Details	Cat.
Where?	Latitude	Float
	Longitude	Float
	Semantic	Cat.

	Feature	Type
When?	Time	Int.
	Weekday	Cat.
	Daytime	Cat.
	Activity	Cat.
With whom?	Neighbors	Int.
	Neighbors Type	Cat.
Last request	Time since last request	Float
	Details of last request	Cat.

wants to request, by selecting the corresponding icon in the main window (see Fig. 1), then she selects the target user from the list of her connected contacts. Finally, the user specifies the level of detail for the requested information, then the request is prepared and sent directly to the target user's device. If the reply is received within a fixed amount of time (typically a few seconds), it is automatically shown to the user, together with the requested information if shared by the targeted requester (see Fig. 1); otherwise, the user is redirected to the main window and she will be notified when the reply is received. At the targeted subscriber's device, the request is processed automatically when it is received: (1) The requested information is stored and (2) the information linked to the request (i.e., the time, the type of information requested and the requester) is combined with various contextual features (periodically collected in the background by SPISM from the various data sources and sensors available on the device) and fed to the decision core that we describe in detail in the next section. If SPISM can make the decision with enough confidence, based on the target user's past decisions, the request is processed automatically. Otherwise, the target user is notified and asked to decide; her decision is then stored (note that the target user can postpone her decision). Once a decision is made, it is sent to the requester together with the requested information if the decision is positive. Before being sent, the requested information is processed to match the level of detail specified by the decision. All the sent and received requests are stored and can be accessed by the user by selecting the corresponding icon in the main window.

SPISM includes a number of key features to further improve the personalization and the usability of information sharing. In particular, users can audit automatic decisions and fix, a posteriori, those they disagree with (to avoid similar errors in the future). In addition, a user can set the confidence threshold of the decision making core, i.e., the confidence level under which SPISM asks her to manually make the decision. By choosing high values of the threshold, the users push SPISM toward a more cautious behavior in terms of information sharing. The users can also set a parameter to push SPISM more toward sharing or toward retaining information (this is achieved by means of cost-sensitive classifier as explained below). Finally, users can set a parameter to adjust the relative influences of old versus recent manual decisions in the automatic decision making process to adapt to changes in sharing behaviors.

3.4. Decision making

The SPISM information-sharing decision-making core processes each incoming information request. In order to make the decision, several contextual features are taken into account by the target device. Features such as the identity of and the social ties with the requester, the current location and the activity of the target, the people around the target and the time of the day were extensively studied in the past; several independent pieces of work show (with statistical significance) that they are strongly correlated with the information-sharing behavior of mobile users [2,17,29,1,30]. With these findings, we list 18 such features that could be incorporated in the SPISM decision-making core (Table 1). Note that this list is not exhaustive: More sophisticated features from social networks, such as the frequency of interaction or the number of friends in common, could be used as well. Due to the different natures of the features, some of them are defined as categorical (they are in a finite and pre-defined set of values, such as the social ties with the requester) or numerical (floating or integer values for the time and location coordinates).

Some of these 18 features can be extracted from the request itself or the target mobile device, such as the time, the current schedule availability or the requester ID, whereas other features require more information, e.g., the social ties with the requester and the semantics of the current location of the target user. To obtain such information, SPISM can leverage on the existing social networks, such as Facebook, and other data available on the phone (e.g., call logs). In addition, other third-party services (such as Google Maps, OpenStreetMap and the Android application store, i.e., Google Play) are used to obtain more information about the location and type of application (in the case where the requester is a mobile application). In some cases, the extraction of the features requires access to the sensors embedded on the device; GPS and Bluetooth scans usually require a non-negligible amount of time and resources [31], and a per-request access to such sensors can drain the battery. For this reason, some time- or energy-consuming features (such as the GPS coordinates and Bluetooth MAC addresses of the nearby devices) are obtained periodically and cached, so that they can be polled by the device at any time

instant without incurring resource-consuming operations. Note that the location, the list of nearby devices and the schedule availability are all used to make the decision and to be shared.

After all 18 features have been extracted from the request or determined from the context, they are aggregated into a feature vector and fed to a classifier. In the basic version of SPISM (evaluated in Section 5.3.1), the output space of the classifier comprises two different classes (i.e., we use a binary classifier) that encode whether the information is shared. In the advanced version of SPISM (evaluated in Section 5.3.2), the output space comprises four classes (i.e., multi-class classifier) that encode whether the information is shared and, if yes, the corresponding level of detail, specifically "No", "Yes (low)", "Yes (medium)" and "Yes (high)". The decisions made by the users are used as ground-truth data for the learning, i.e., training the classifier: the training consists in choosing the optimal values for the different parameters of the classifier (e.g., the coordinates of the border hyperplanes for linear SVM). For instance, the training sets the parameters of the classifier in such a way that the number of misclassifications on the ground-truth data is minimized.

By using cost-sensitive classifiers, SPISM can assign different penalties to different misclassifications. For instance, to capture the gravity of the misclassifications one can set the penalty of (mis)classifying a "Yes (low)" decision as "Yes (high)" to be twice as large as the penalty of (mis)classifying a "Yes (low)" decision as "Yes (medium)". With cost-sensitive classifiers, the *total penalty* of the misclassifications is minimized upon training (instead of just the total number of misclassifications). Optimizing classification performance by means of penalties or costs is a quite common practice for limiting specific types of errors.

Cost-sensitive classifiers also enable SPISM to find an appropriate balance between the cases where too much information is shared and the cases where too little information is shared. Setting the penalty of (mis)classifying a "No" decision as "Yes" to be twice as large as the penalty of (mis)classifying a "Yes" decision as "No" pushes SPISM toward not sharing information. Such a configuration is beneficial to privacy, as fewer requests are unduly answered, but it is also detrimental to the utility of the service, as more information requests are unduly left unanswered.

Finally, in order to give a greater importance to the recent decisions, SPISM assigns different weights to the users' past decisions (i.e., the ground-truth data) during the training, thus dynamically adapting to the changes in the users' sharing attitudes. SPISM makes use of the classifiers implemented in the WEKA[5] Android library, mainly the naive Bayes classifier, the SVM classifier with Gaussian or polynomial kernels, and the logistic regression classifier.

4. Study and data collection

In order to better understand how users share information and to evaluate the efficacy of the SPISM framework with respect to sharing decisions, we ran a user study in early 2013. The study consists of an online survey that puts the participants in realistic, personalized, and contextual sharing scenarios where they are asked to answer a set of questions regarding their willingness to share private information, the confidence in and reason for their decisions.

4.1. Participants and remuneration

We recruited people directly from four large university campuses (in the US, Canada and Europe), and indirectly via the Amazon Mechanical Turk platform (MTurk).[6] The latter allowed us to draw participants from a pool of non-student population, in order to limit the bias toward academic and student behaviors (note that our user sample might not be representative of the entire population, due to some bias introduced by MTurk [32]). To advertise our study, we used dedicated mailing-lists and we ran a media campaign through Facebook, LinkedIn, Google+ and official university websites, coordinated by our academic media office. We screened participants according to the following prerequisites: (i) aged between 18 and 80 years, (ii) with an active Facebook account with at least 50 friends and (iii) uses a smartphone. Such criteria were selected so as to sample people that are active in social networks and are aware of the information-sharing possibilities linked to the use of smartphones. Furthermore, we screened the MTurk workers who could access our survey based on their past Human Intelligence Task (HIT) approval rate ($>95\%$) and the number of past approved HITs (>100). This was only a preliminary step for preventing non-serious and inexperienced MTurk workers from accessing our survey.

The survey requires access to the participants' private information (such as names of their friends on Facebook[7]) We obtained the list of the Facebook friends of all the participants (those recruited from both MTurk and social platforms). This was achieved through the Facebook API and the participants had to grant our survey engine access to their Facebook friend lists. We further asked them to specify the social ties they have with certain of their Facebook friends, as explained below. The survey demands a significant amount of time (40–60 min). To provide incentives for the completion of the survey, we implemented two separate reward schemes: (i) the chance for one participant to win a tablet and (ii) a fixed amount of

5 http://www.cs.waikato.ac.nz/ml/weka/.

6 https://www.mturk.com/mturk/welcome.

7 Before beginning the survey, the participants were informed that they would need to reveal the names of their Facebook friends for the purpose of this study. They approve a data retention and processing agreement, informing them that all data collected in our study is used solely for the purpose of our academic research project, and that we will not disclose or use it in any other way than what explicitly mentioned. Once the survey is completed, the name of the Facebook friends are replaced with anonymous identifiers.

money (US$4.5/HIT [33]). The first option was proposed to the participants recruited at the universities and through the academic media, whereas the second option was offered to the workers of the Amazon Mechanical Turk. We chose not to offer the second option to the academic participants due to our experience gained from previous on-campus studies: The motivation for financial rewards was lower than for the possibility of winning a popular gadget.

4.2. Online survey

We structured our survey in five parts: With a total of 94 questions, the first 19 were fixed (the same for each participant) and the last 75 were personalized (based on each participant's Facebook friends). In the very first part, the participants were required to log in to their Facebook account and grant survey engine access to their friend list.

In the first 15 questions, the participants were asked about their demographics, technology usage and privacy attitudes, in particular with respect to online social networks.

In the next question (16), the participants were asked to assign some of their Facebook friends to five different social categories (based on [20]): (1) school colleagues, (2) friends, (3) family members, (4) work colleagues and (5) acquaintances. Each participant could assign one Facebook contact to at most one category. It is possible, however, that one such contact is a member of several categories (a school colleague that she works with currently). In this case, the participants were instructed to assign the contact to the most appropriate category. This categorization defines the social ties between a user and her contacts and it is used by SPISM in the decision making process.

In question 17 through 19, the participants were asked to enter a set of information-sharing rules in free-text. The sharing rules were entered as a set of logical expressions based on the following *features*: (1) the participant's current location, (2) people nearby, (3) social group of the requester, (4) time of the day and (5) weekday/weekend. They were allowed to put *conditions* on these features (such as $=$, $<$, $>$, $\neq$, $\in$ or categorical values). For example, a location-sharing rule could be defined as:

> "I am at a *friend's place* **and** with *acquaintance* **and** the requester is a *work colleague*: do not share"

In the remaining 75 questions, the participants were presented with sharing scenarios and they were asked to decide whether they want to share the specific information in the given context, their confidence in the decision and the level of detail. A typical scenario is "Would you share your *location* with *John Smith* on *Saturday* at *11PM*, assuming you are with *friends*?" (where the requester name is chosen from the participant's Facebook friends and the other features are chosen at random). Fig. 2 shows a screenshot of our survey webpage for a sample sharing scenario. Note that we consider fewer features than specified in Table 1, because (1) it would be cumbersome for the participants to read and analyze the complete description of the context, and (2) the resulting data would be too sparse (in terms of the number of occurrences of each possible value of each feature) to efficiently train a classifier.

Depending on their answers ("Yes", "No" and "Uncertain") to the questions in this part, participants were presented with sub-questions. More specifically, "Yes" and "No" answers were followed by a set of additional questions asking the participants about the confidence in their decisions (i.e., "not so confident", "confident", "very confident") and the features that influenced the most their decision (i.e., "requester", "time", "location"). For "Yes" answers, the participants were also asked about the level of detail of the shared information ("low", "medium" or "high"). Similarly, "Uncertain" answers were followed by sub-questions regarding the reasons for being uncertain, such as a conflict between some features (in this case, the participant can specify the features that motivates her the most to share and to not share, and then specify in free text the reason they conflict) or simply a lack of information (in this case the participant can specify which information would have helped her reach a decision).

In order to detect sloppy answers (e.g., random answers or bots), we included a number of "dummy" questions that require human understanding to be correctly answered [33,34]. These are questions such as simple computations (e.g., "3+4") or general-knowledge questions (e.g., "How many days are there in one week?"). Based on the answers to these questions and on the survey timing data (explained below), we ruled out dishonest participants from the dataset.

4.3. General statistics and validation

A total of 194 participants took part in our survey. 78 (40%) of them did not complete it, leaving 116 (60%) questionnaires completed. Out of these, 56 (48%) came from the university advertisement campaign (Univ.) and 60 (52%) were recruited via MTurk. The average age of all the respondents is 27y $\pm$ 7 (MTurk avg. 31y $\pm$ 6, Univ. avg. 25y $\pm$ 6), and 74% of them are male. 42% of all participants are students, 25% work in the IT industry and 8% in the education sector. It took 44 $\pm$ 15 min on average to complete the survey (MTurk avg. 42 min, Univ. avg. 47 min). We observed a sharp contrast, with respect to privacy concerns, between the two groups of participants: Most MTurk participants were not, or only slightly, concerned about their privacy, whereas most Univ. participants were concerned about it (see Fig. 3): For most of the information types, university participants reported significantly higher concerns as compared to the MTurk participants. In particular, for the three types of information considered in SPISM ("location", "activity" and "people nearby"), the proportion of Univ. participants not concerned at all is much lower than that of MTurk participants.

Based on internal survey tests and detailed timing statistics, only the questionnaires that meet the following four validation criteria were retained:

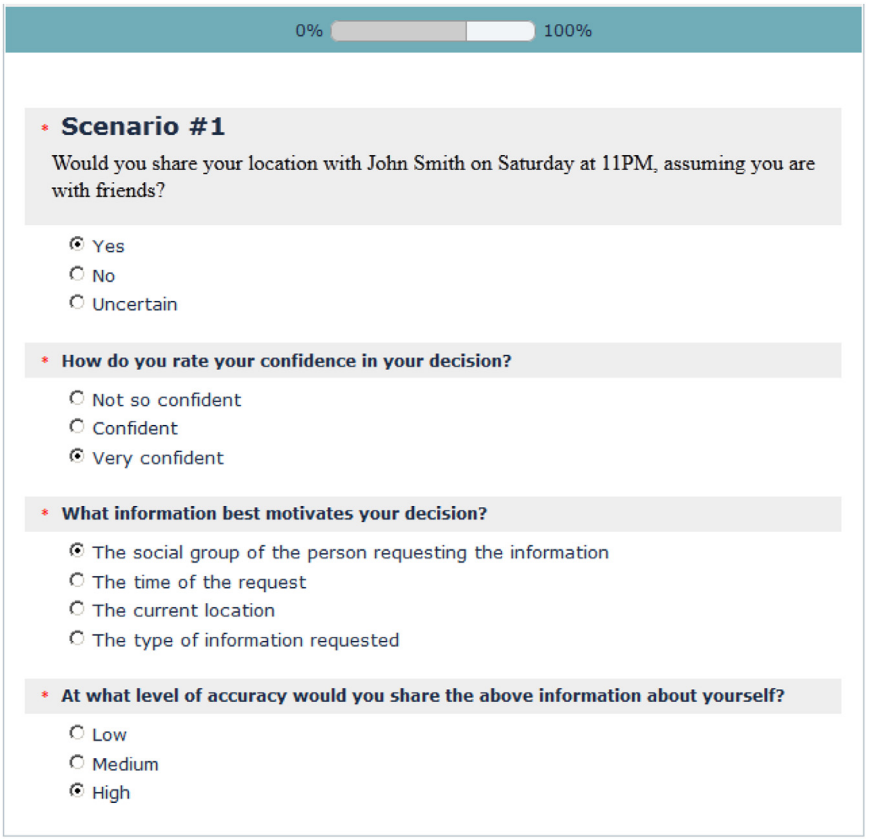


Fig. 2. Screenshot of our survey webpage showing a sample sharing scenario and the corresponding questions.

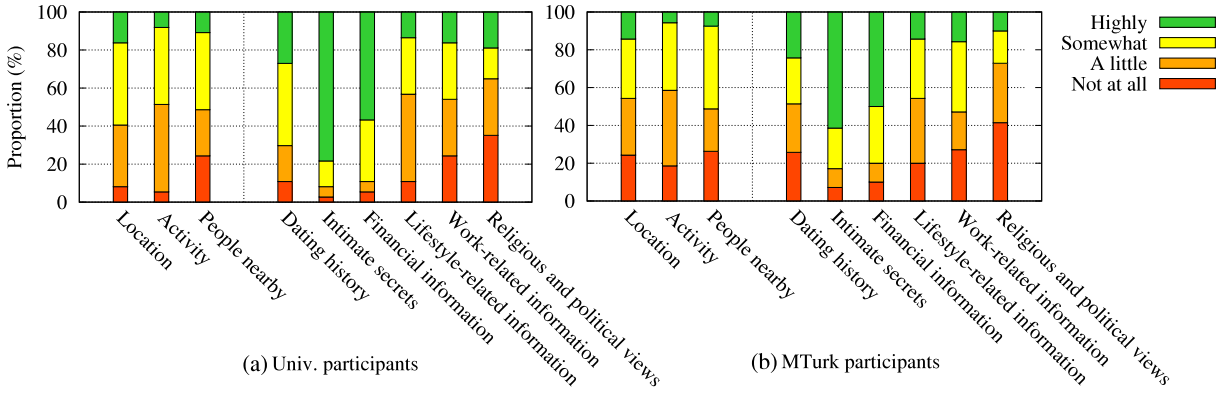


Fig. 3. Privacy concerns reported by the survey participants (Univ. and MTurk), for various types of information, including the three considered in SPISM ("location", "activity" and "people nearby").

- All answers to the dummy questions are correct.
- At least one different Facebook friend is assigned to each of the 5 social groups.
- The survey completion time is greater than 30 min.
- At least three of the following four timing conditions are met[8]: (1) Facebook friends assignment to groups time >5 min, (2) location sharing scenarios time >4 min, (3) activity sharing scenarios time >4 min, (4) nearby people sharing scenarios time >4 min.

8 These timing conditions were determined based on the observed timing distributions among all participants and on sample executions performed by test users.

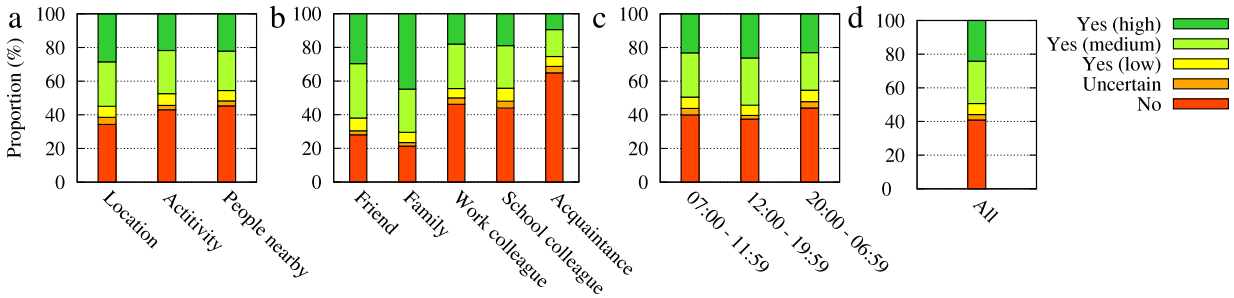


Fig. 4. Histograms of the information-sharing decisions distinguished by (a) information type, (b) social group of the requester, (c) the time of the day or (d) without (feature) distinctions.

All participants correctly answered the dummy questions. Based on timings, 46 (40%) of them were ruled out and 70 (60%) were kept for the analysis (33 MTurk and 37 Univ.). Note that the relatively high number of participants who were ruled out can be explained by the "respondent fatigue" phenomenon [35], as the last part of our survey (i.e., the what-if information sharing scenarios) was particularly long and repetitive. The demographics remained mostly unaltered.

5. Analysis and evaluation

In this section, we present three sets of results. First, using descriptive statistics of the survey questionnaire, we discuss the effect of different contextual features (the requester, the information type) on the sharing decisions and the main reasons behind the decisions. Second, we compare the performance of the SPISM automated decision-making process against that of the users' own policies. Third, we discuss the effects of the increase of user-involvement on the performance of SPISM, by using active learning with different confidence thresholds. Finally, in addition to the results outlined in [13], we present (1) an optimization, based on cost-sensitive classifiers, that significantly increases users' privacy with minimal effect on the overall performance of SPISM, (2) a detailed analysis of the multi-class case in which the classifier predicts not only the decision to share (or not) the requested information but also the granularity of the shared information (if shared), and (3) an evaluation of a universal one-size-fits-all model built from all the users' data for decision making.

5.1. Survey statistics

Based on the survey data (composed of 75 decisions for each of the 70 survey participants, totaling 5250 data points), we computed the proportion of "Yes/No/Uncertain" decisions for the different values of each contextual feature we considered, such as the participant's current location, the social group of the requester, the time of day, day of week, and the type of information requested. We found that the two that have the largest effect on the decision are the social group of the requester and the type of information that is being requested.

Regarding the type of information being asked, Fig. 4a shows that users disclose their location in 64% of the cases (the sum of the "Yes (low)", "Yes (medium)" and "Yes (high)" bars, aggregated over the 70 participants and for all the 25 location-sharing questions – out of the 75 questions – that is a total of 1750 answers), and only 8% of the time at a coarse granularity ("Yes (low)"). The information about activities and people nearby is disclosed 50% of the time. People tend to be slightly more willing to share their location than to share other information[9]: Location, contrary to the activity and the co-presence of other people, is widely shared information in most mobile social networks. This was confirmed by the statistics on the self-reported privacy concerns about information sharing on OSNs (Fig. 3).

Fig. 4(b) shows the percentage of disclosure of information based on the social ties with the requester. We can see that, in accordance with previous studies related to information sharing, there are substantial differences (see footnote 7) between the close ties ("family" and "friend") and the distant ones ("acquaintance" and "colleague"). For instance, the close ties are granted access to any type of information (70%–80%) more than twice the times compared to the distant ones (30%). Moreover, the level of detail of the shared information is much higher for the close ties (up to 45% of "Yes (high)") compared to the distant ones (down to 8%). In fact, the proportion of "Yes (low)" and "Yes (medium)" does not vary significantly. Hence, the results indicate that users tend to exhibit a more tailored sharing behavior depending on the type of information, the social ties and closeness with the requester [20]. As illustrated in Fig. 4c, the time at which the request is sent does not substantially influence the decision: users are slightly less willing to share in the evening but exhibit the same behavior in

[9] With statistical significance, based on global and pair-wise χ^2 homogeneity tests with $p < 0.01$. The χ^2 homogeneity test is the standard way to determine whether the differences observed (for a categorical variable, e.g., the decision) between different groups (e.g., different types of requested information) reflect real behavioral differences or sampling errors. We perform both pairwise tests, between every pair of groups (e.g., location vs activity in Fig. 4), and a global test that considers all the groups together. The p-value captures the likelihood that the observed differences are caused by sampling errors; p-values below 0.01 are commonly considered to denote a real difference between groups.

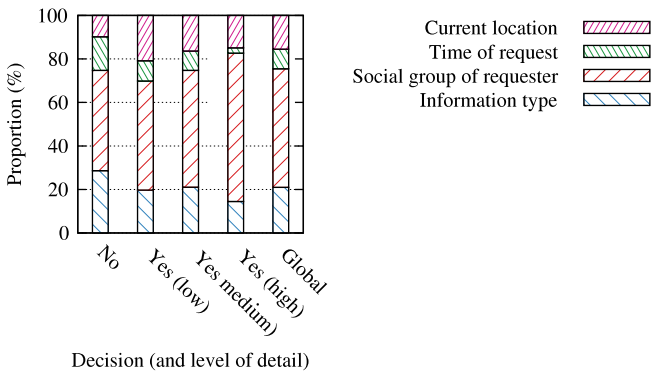


Fig. 5. Histograms of the main reasons for (not) sharing as a function of the decision and level of detail of the shared information. The "Social group of requester" and the "Information type" consistently appears as the first and second reasons for (not) sharing information.

the morning as in the afternoon (see footnote 7). Our findings are aligned with those obtained in [25], where the time of day and the location do not appear to be influential factors when personal information is shared about location, as opposed to the type of social ties with the requester.

We further investigated the "Uncertain" decisions. The total number of "Uncertain" decisions was 170 out of 5250 decisions (75 scenarios for each of the 70 participants), which is a proportion of 3.2%. Out of the 170 "Uncertain" decisions, 57 (34%) were due to a conflict between the different pieces of contextual information, and the remaining 113 (66%) were due to the lack of contextual information. The participants who were unable to make a decision because of a conflict between the different pieces of information presented to them specified the conflicting information for 22 of the 57 such scenarios. More specifically, they specified which piece of information pushed them the most toward sharing and which piece of information pushed them the most toward *not* sharing. Among the 22 scenarios, the distribution was as follows (by decreasing frequencies): "(+) Requester/(−) Information type" (10), "(+) Requester/(−) Time of request" (5), "(+) Time of request/(−) Requester" (3), "(+) Time of request/(−) Information type" (2), "(+) Information type/(−) Requester" (1), "(+) Information type/(−) Time of request" (1).[10] Interestingly, location was never mentioned as conflicting information. A possible explanation is that location information is not a very important factor in the decision anyways, as shown by the results presented in the next paragraph. The participants who were unable to make a decision because of the lack of information entered comments for 29 of the 113 such scenarios. The two most frequent comments were that the name of the people nearby and specific information about the current activity would have helped them to reach a decision.

We also looked at the reasons for (not) disclosing information and at the users' confidence in their decisions. First we observe that the social tie with the requester is by far the most frequent reason for sharing (or not) information (45%–67%), followed by the type of information (15%–28%) and the current location (11%–21%). Second, we see again that the higher the level of detail of the shared information is (see Fig. 5, where the lowest level is "No" and the highest is "Yes (high)"), the more important the social ties with the requester are (on average). Unsurprisingly, the confidence that the participants have in their decision (Fig. 6) is lower for the intermediate decisions (i.e., "Yes (low)" and "Yes (medium)"): It can be observed that the proportion of "Very confident" is significantly lower for "low" and "medium" levels of detail than for "No" and "Yes (high)". In addition, the proportion of "Not so confident" is more than doubled for the most borderline decision, i.e., "Yes (low)". This could be explained by the fact that users try to minimize the risk by limiting the level of detail when their confidence is low.

5.2. Static policies

We compared the performance of our SPISM decision framework with a policy-based approach. For the following comparison, we used 10-fold cross validation and a logistic regression binary classifier (SVM and naive Bayes classifiers give similar results). In order to be consistent with the policy-based approach, we only compare the binary ("Yes/No") decisions here as the participants were instructed to only specify share/not share policies in the survey. The policy-based approach is derived from the individual policies that each participant specified in free text in the survey. We selected a random sample of 20 participants (among those who specified meaningful policies) and we manually transposed their free-text policies to a format suitable for evaluation against their own decisions. The participants specified between 1 and 15 policies (avg. 6.9).

The results of the comparison are shown in Fig. 7 where the results are sorted in descending order of performance (for the participant's individual policies). We can observe that the SPISM machine-learning approach consistently outperforms the static policy approach. The SPISM performance rate is between 53% and 100%, with an average of 71% (for the 20

[10] We denote with a "+" the piece of information that pushed the participants the most toward sharing and with a "-" the piece of information that pushed them the most toward *not* sharing (as reported by the participants).

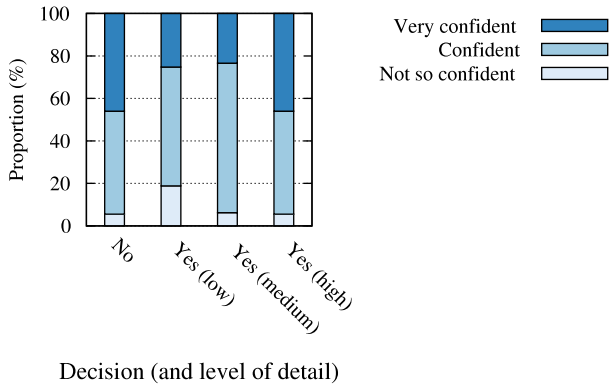


Fig. 6. Histograms of the users' confidence in their decisions as a function of the decision and level of detail of the shared information. People have lower confidence in their decisions when they share information at an intermediate level of detail (low or medium).

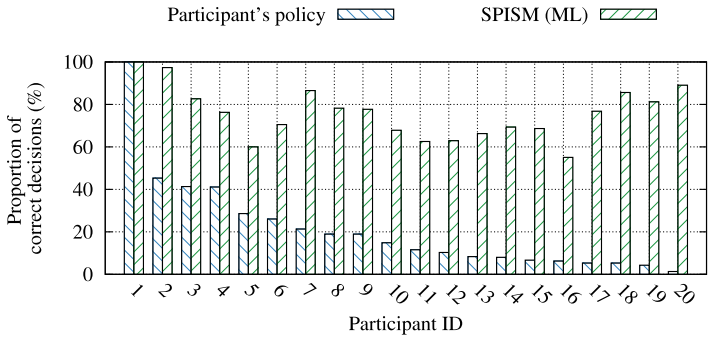


Fig. 7. Histograms of the proportion of correct sharing decisions: For each participant, the left bar shows the performance of SPISM and the right bar shows the results of the performance of the individual policies. The SPISM approach uses a logistic classifier (10-fold cross validation) and the participants' individual policies are derived from their free text answer in the survey.

users considered in this experiment). Compared to the participant's policy (avg. 21%), SPISM is significantly better as it automatically learns the user's sharing behavior.

For the individual policies, we also observed the correctness of the decisions as a function of the number of policies, and we found that a small number of policies (1–5) achieved up to 41% of correct decisions, followed by a slightly better performance for the number of policies between 6 and 9 (up to 45%), but then a much worse performance (up to 28% of correct decisions) for the highest number of policies (10–15). This suggests that there is an advantage in having a moderate number of sharing policies (up to 9) but not higher; with a larger number of policies, the risk of having overlapping but contradicting policies is higher, which could result in a worse overall performance.

5.3. Machine learning

In order to assess the potential of (semi-)automatic information-sharing decision making, which constitutes the core of SPISM, we evaluate the performance of a classifier in predicting the users' sharing decisions. To do so, we use the survey data comprised of 75 scenarios for each of the 70 participants: Each scenario corresponds to a feature vector and the decision made by the participant constitutes the ground truth. We evaluate the performance of the classifier in terms of the proportion of correct predictions (i.e., that match the user's decision), the proportion of cases where the information is shared although the user would have not shared it (namely *over-share*), thus compromising the user's privacy, and the proportion of cases where the information is not shared although the user would have shared it (namely *under-share*), thus reducing the utility of the system.

We describe the results of the machine-learning based approach in the following order:

1. Binary sharing decisions ("Yes"/"No")
   (a) Correct/incorrect performance, where we describe the behavior of the classifier after training on a variable number of sharing scenarios.
   (b) Under-/over-share performance, where we discuss the incorrect decisions and the tendency to over-share information.
   (c) Active learning, where we show how to balance the performance and the user burden by asking the users to manually make a decision when the confidence of the classifier is low.

(d) Cost-sensitivity, where we tune the classifier toward under-sharing by means of a cost-sensitive classifier fed with an error-penalty matrix.
(e) One-size-fits-all model, where a single classifier is trained on the data of all users and is used to predict the sharing decisions for each of them.
(f) Confidence in the sharing decision, where we compare the users' confidence against the classifier's confidence in the sharing decisions.

2. Granular sharing decisions ("Yes (High/Medium/Low)"/"No"), where we discuss the correct/incorrect and under/over-share performances for the multi-class cost-sensitive classifier.

In our evaluation, we used the following classifiers from the WEKA library: the naive Bayes classifier, the SVM classifier with polynomial kernels, and the logistic regression classifier. We used the default parameters from WEKA to avoid over-fitting.

5.3.1. Binary case ("Yes"/"No")

We begin by looking at the "binary case", where the classifier outputs either "Yes" or "No", without taking into account the level of details.

Training. First, we consider the case where the users first manually make n decisions to train the classifier,[11] and then the classifier makes the remaining decisions automatically (we later discuss the case of active learning where the system asks the users to manually make the decisions when the confidence of the classifier is too low). For several values of n, and for each participant, we compute the average proportions of correct and incorrect decisions following a repeated random sub-sampling validation approach. For each value of n, we obtain one data point (i.e., a proportion of "correct", "under-share", and "over-share" decisions) for each user, that is 1260 data points. We represent the results across the different users and folds by showing the median, the first and third quartiles, and the 5 and 95-percentiles,

(a) *Correct/Incorrect sharing.* Fig. 8(a) shows that the median proportion of correct decisions increases from 59% and reaches 69% for a training set of only 30% of the data, which correspond to ∼25 scenarios. The proportion of correct decisions then quickly stabilizes around 72% after approximately 40 decisions (i.e., ∼50% of the data) and goes up to 76%. The third quartile and the 95-percentile show that for more than 25% of the users, the proportion of correct decisions goes up to 80% and for some of them, it is consistently higher than 96%. For comparison, a simple classifier that always returns the most frequent class in the training set (namely a ZeroR classifier), achieves up to a median proportion of correct decisions of 64%. Naive Bayes and SVM classifiers both achieve up to a median proportion of correct decisions of 76%.

(b) *Under- and over-sharing.* The proportion of incorrect decisions is evenly distributed between "over-share" and "under-share", although slightly biased toward "over-share". Without penalties and active learning, "over-share" happens in 10%–20% of the cases, in line with the results reported in [25] and obtained with different classifiers. Note that the size of the training set (represented on the x-axis) represents the burden of the user as she has to manually make the corresponding decisions.

We further investigate the reasons behind the under- and over-sharing cases. To do so, we consider each feature independently and, for each possible value of the considered feature, we compute the proportions of under- and over-share (among the misclassifications) and we compare them to the total proportions of under- and over-share. We do not observe a significant deviation from the total proportions, except for the social group of the requester. For the social group of the requester, we observe a proportion of over-share of 46% for acquaintances, 56% for friends and 57% for family members. This can be explained by the fact that the general trend is to share with friends and family members (as shown in Fig. 4) but to not share with acquaintances. Therefore, in a specific context where a user does not want to share information with one of her friends, the classifier tends to (over-)share the requested information (e.g., if a user always shares with her friends, except on Sunday mornings). In our evaluation, such specific contexts might not be captured during the training phase because of the sparsity of the data. This sparsity is due to the limited number of scenarios, compared to the number of features and the number of possible values for each of them.

(c) *Active learning.* We now consider the case of active learning, where SPISM asks the users to manually make the decisions for which the confidence of the classifier does not meet a confidence threshold (which can be set by the users). The classifier outputs a distribution over the possible decisions; we define the confidence as the normalized entropy of this distribution. The classifier is first initialized with 10% of the data. For each user, we run the active learning-based classifier for several values of the confidence threshold, under which the user is asked to make the decision. Each experiment gives one data point comprised of (1) the proportion of decisions (including the first 10%) the user has to manually make and (2) the proportions of correct and incorrect decisions among the decisions that are made automatically. In order to represent the data in a form that is comparable to that of Fig. 8(a), we group the data points in bins of size 5% (on the x-axis as represented in the figure) based on the proportion of manual decisions. Note that the number of data points varies across the different bins. Within each bin, we compute the median and the relevant percentiles. The result are depicted in Fig. 8(b). It can be observed that

[11] These n scenarios can be either real requests from the users' contacts or virtual what-if-scenarios, like those considered in our survey. In the latter case, SPISM could use uncertainty sampling for choosing the most informative scenarios about which to poll the users.

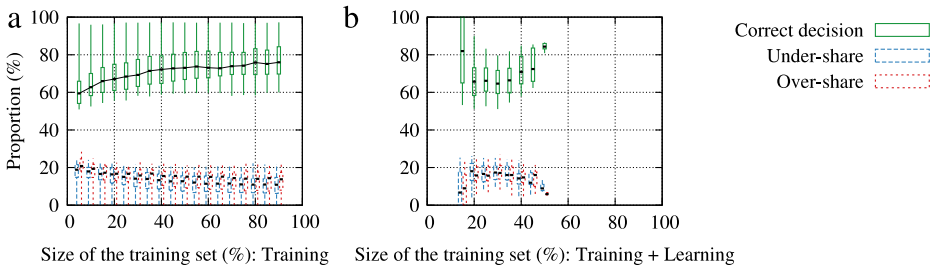


Fig. 8. Performance of the machine-learning-based decision making algorithm in the binary case ("No"/"Yes") with (a) training and (b) active learning (logistic classifier).

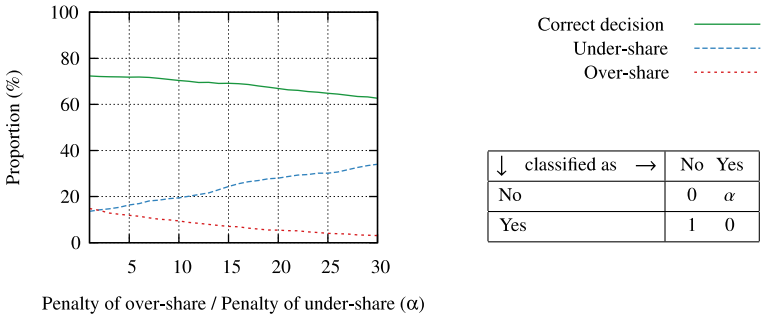


Fig. 9. Effect of error penalties on the median proportions of "correct", "under-share", and "over-share" decisions (training set size of 50%). The error-penalty matrix is shown in the bottom-right corner (cost-sensitive SVM classifier).

active learning outperforms training-only learning in most cases (i.e., for a given number of manual decisions, it provides a higher proportion of correct decisions). The proportion of manual decisions remains lower than 50% which shows that the classifier can make the decision with very high-confidence for at least half of the scenarios. For some users, the proportion of manual decisions remains low (∼20%), regardless of the confidence threshold, and the proportion of correct decisions is high (∼80%). This corresponds to the users whose decisions are highly predictable. With active learning, we observe a significantly improved performance in terms of over-sharing compared to the absence of active learning. We posit that, coupled with cost-sensitive classifiers, active learning can be used to improve the correctness of the sharing decisions and maintains a significantly lower over-sharing rate.

(d) *Cost-sensitivity*. We now evaluate the extent to which incorrect decisions can be biased toward the "under-share" cases (which have a negative effect on utility) instead of "over-share" cases (which have a negative effect on privacy), when a user favors her privacy over the utility of the system. To do so, we make use of error penalties in the cost-sensitive version of the classifier (only the SVM classifier has a cost-sensitive version, so we used a SVM classifier in this experiment). This functionality enables us to specify different error-penalties for the different cases of incorrect decisions (here "Yes" instead of "No" and "No" instead of "Yes"), and the classifier minimizes the total error-penalty instead of the number of incorrect classification during the training. We assign a penalty 1 to "under-share" and a penalty α to "over-share" ($\alpha \geq 1$). This means that, during the training, we minimize the following quantity: $\alpha \times$ "number of over-share cases" $+ 1 \times$ "number of under-share cases", on the training (ground-truth) data. We plot the proportions of "correct", "under-share", and "over-share" decisions for a training set size of 50%. The results are shown in Fig. 9, together with the error-penalty matrix. It can be observed that the proportion of "over-share" decisions can be significantly reduced at the cost of a slight decrease in correctness. For instance, for $\alpha = 15$, the proportion of "over-share" decisions decreases from 15% to 7% (compared to the case without error penalties), and the proportion of correct decisions only drops from 72% to 70%. This is a significant result, as it shows that a fail-safe approach for information sharing, where information is not shared rather than mistakenly shared, works very well with only a minimal effect on the overall performance: SPISM is able to gracefully balance utility and privacy by means of cost-sensitive classifiers.

(e) *One-size-fits-all classifier*. We also evaluate the potential of a one-size-fits-all model where a single classifier, trained on the data of *all* the users, is employed by the users. For the sake of fairness, we exclude the data of a user (i.e., we train the classifier on the data of all users, except the considered user) when testing the one-size-fits-all classifier on her data. We also remove the ID of the requester from the list of features. The results, with a logistic classifier, are depicted in Fig. 10. It can be observed that the one-size-fits-all classifier performs relatively well, with a median of 67% of correct decisions. For comparison, a simple ZeroR classifier that always returns "Yes" (i.e., the most frequent decision across all users) achieves a median accuracy of 58% (i.e., the proportion of "Yes" decisions, as shown in Fig. 4(d)). Such a classifier has a zero proportion of "under-share" decisions (as it always returns "Yes") and thus a high proportion of "over-share" decisions (42%). Naive Bayes

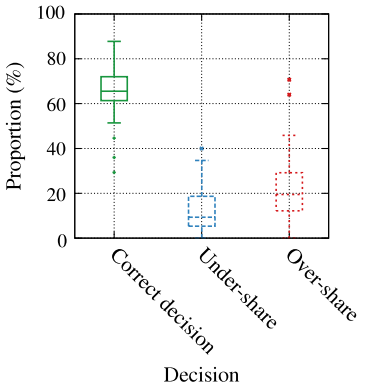


Fig. 10. Performance of a *one-size-fits-all* classifier. Such a classifier has similar performance as the personal classifiers trained on 15% of the data (logistic classifier).

and SVM classifiers achieve a median accuracy of 65% and 67%, respectively, with a higher proportion of "over-share" (27%) for SVM. Even though there is only an 11% increase on average, the one-size-fits-all model does not require any prior training for a new user; instead, by relying on the consistency of users' sharing behaviors, such a model could be used, for example, to suggest a "default" sharing option for new users, until the personalized classifier has enough data to outperform it. For comparison, a personalized classifier (i.e., trained on a user's data) requires a training set size of 15% (after ∼11 manual decisions) to achieve such a performance (see Fig. 8(a)). Note that the proportion of "over-share" is higher than that of "under-share": A possible explanation is that some users in our dataset tend to always share the requested information, thus biasing the one-size-fits-all model toward sharing; this induces a larger proportion of "over-share" (compared to the personal classifiers).

(f) *Confidence in the sharing decisions*. To better understand the decision process, we study the relationship between the confidence and the influence of the different features in the automatic (i.e., the classifier) and manual decision processes (the users).

In Fig. 11, we plot the self-reported user confidence levels ("Very confident", "Confident", "Not so confident", "N/A"[12]) versus the confidence of the classifier (when trained on the rest of the data), captured by the probability assigned (by the classifier) to the chosen class (i.e., "Yes" or "No"). More specifically, for each scenario we train a classifier on all the other decisions (from the same user) and we classify this scenario. We aggregate the data for all users and all decisions and we represent the first, second and third quartiles, as well as the 5th and 95th percentiles. We evaluate the confidence of the classifier, based on the probabilities it assigns to "Yes" and to "No". The users' self-reported confidence levels are obtained from the survey data, as described in Section 4. Surprisingly, except for the case where the user is "Very confident" in which the classifier's confidence is significantly higher, there are no significant differences between the different level of self-reported user confidence (in terms of the classifier's confidence).This shows that although machine-learning techniques succeed at predicting the users' behaviors, the automatic decision process differs from that of the users. A possible explanation is that the users, unlike the machine-learning algorithms, are aware of other important information that could help the decision making process (66% of the "Uncertain" decisions were due to the lack of contextual information). In addition, the sparsity of the data can be a possible explanation: Because specific scenarios that would illustrate the subtleties of the decision making process (e.g., the differences of sharing behaviors toward two different friends) are not present in the dataset, the classifier cannot detect that such a decision is in fact tricky.

We consider all the data of all the users and we extract, by using a specific functionality of WEKA (namely, the Correlation Feature Selection subset evaluation method, which is independent from the classification), the influence of each feature in the decision. In order for our results to be comparable to those reported by the users (shown in Fig. 5), we consider only the following four features: "Information type", "Social group of the requester", "Time of request" and "Current location". We obtained the following ranking (ordered by decreasing influence): (1) "Social group of the requester", (2) "Information type", (3) "Time of request" and (4) "Current location". These results are consistent with the information reported by the users; except that "Time of request" and "Current location" features are swapped in the ordering (note however that this is not the case for the "No" decisions).

5.3.2. Granular sharing decisions ("Yes (High/Medium/Low)"/"No")

We now look at the multi-class case, where the classifier outputs also the level of detail at which the information is shared. We define and use four different classes: "Yes (high)", "Yes (medium)", "Yes (low)" and "No". Note that the binary

[12] The "N/A" label corresponds to the case where the user could not reach a decision, i.e., the "Uncertain" decision, which constitutes the lowest level of confidence.

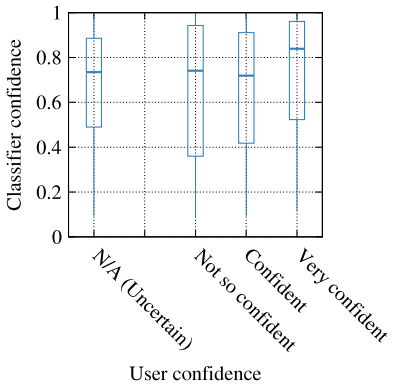


Fig. 11. Confidence of the classifier vs. self-reported confidence of the users.

Table 2
Error-penalty matrix for the multi-class case. The rationale behind the error-penalty matrix is as follows. The basic error-penalty is 1 (e.g., for classifying "Yes (medium)" as "Yes (low)"); penalties are additive (i.e., the penalty for classifying "Yes (high)" as "Yes (low)" is the sum of the penalty for classifying "Yes (high)" as "Yes (medium)" and "Yes (medium)" as "Yes (low)"); over-sharing (i.e., moving to the right in the error-penalty matrix) is α times worse than under-sharing (i.e., moving to the left in the error-penalty matrix); deciding "No" instead of "Yes" has a fixed penalty of β. By doing so, we reduce the number of parameters in the matrix from 12 (i.e., the non-diagonal terms) to 2 (i.e., α and β).

Classified as	No	Yes (low)	Yes (medium)	Yes (high)
No	0	$\alpha\beta$	$\alpha\beta + \alpha$	$\alpha\beta + 2\alpha$
Yes (low)	β	0	α	2α
Yes (medium)	$1 + \beta$	1	0	α
Yes (high)	$2 + \beta$	2	1	0

case corresponds to the case where the last three classes are merged. In order to handle multiple classes, we use a one-versus-all strategy for the classification; note that we obtained similar results with a one-versus-one strategy. The output of the classifier is a probability density function over the four possible classes. We decide on the class with the highest probability[13]. We use a cost-sensitive classifier, and we define penalties in order to encode the facts that (1) over-sharing is more problematic than under-sharing and (2) the different cases of under- and over-sharing have different levels of gravity (e.g., when a users wants to share information at a low level of detail, it is more problematic if it is shared at a high level rather than at a medium level). In practice, the coefficients of the error-penalty matrix could be set based on the users' sensitivity to each type of misclassification.

Fig. 12 shows the results in the multi-class case, with the penalty-errors given in Table 2. For the sake of clarity, we show only the proportions of correct, "under-share" and "over-share" decisions. Therefore the case where a "Yes (low)" decision is classified as "Yes (medium)" cannot be distinguished from the case where a "Yes (low)" decision is classified as "Yes (high)" (they are both counted as "over-share"). It can be observed that the performance of the classifier is good: the median proportion of correct decisions goes up to 60% with a median proportion of "over-share" decisions as low as 11%. For comparison, a simple ZeroR classifier that always returns the most frequent decision achieves up to a median proportion of correct decisions of 49% and a median proportion of "over-share" decisions of 0%. This is because, for a majority of users, the most frequent decision is "No" (as in the multi-class case, the "Yes" decision is split into three sub-decisions); therefore, for these users, the ZeroR classifier always returns "No", thus never making "over-share" decisions. Note that, in the multi-class case, a correct decision means that not only the classifier correctly decides whether to share the information or not, but it also correctly determines the level of detail at which the information is shared. It shows that, by using appropriate and optimized machine-learning techniques, SPISM can make correct and accurate decisions in an automated fashion. A classifier that is not cost-sensitive would achieve better raw performance in terms of the median proportion of correct decisions, but it would make more "over-share" decisions and, in general, more serious incorrect decisions (e.g., "Yes (high)" instead of "No").

As in the binary case, the classifier can be further biased toward "under-share" decisions by optimizing the error penalties (i.e., by increasing α). We evaluate the performance of a cost-sensitive classifier with an error-penalty matrix that assigns 1 for all under-share misclassifications and α for all over-share misclassifications (regardless of the granularity of the decision). Fig. 13 shows the results for various values of α. It can be observed that, as in the binary case, the proportion of over-share can be significantly reduced with a very limited effect on the proportion of correct decisions. The effect on the proportions

[13] Note that more complex classification mechanisms could be used to take into account the fact that the levels of granularities are ordered. For instance, if the output of the classifier is: "No": 0.0, "Yes (low)": 0.51, "Yes (medium)": 0.0, and "Yes (high)": 0.49, a wise choice would be the intermediate granularity level "Yes (medium)" instead of "Yes (low)". For the sake of simplicity, in our evaluation, we only consider the class with the highest probability.

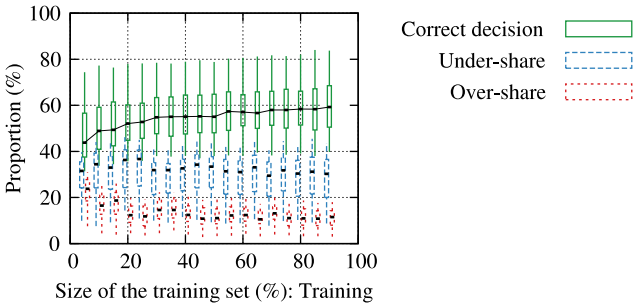
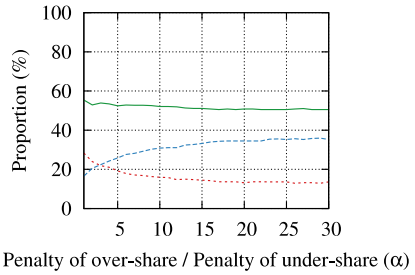


Fig. 12. Performance of the machine-learning-based decision-making algorithm in the multi-class case ("No", "Yes (low)", "Yes (medium)" and "Yes (high)"), using the error-cost matrix from Table 2 with $\alpha = 5$ and $\beta = 2$ (cost-sensitive SVM classifier).



↓ classified as →	No	Yes (low)	Yes (medium)	Yes (high)
No	0	α	α	α
Yes (low)	1	0	α	α
Yes (medium)	1	1	0	α
Yes (high)	1	1	1	0

Fig. 13. Effect of error penalties on the median proportions of "correct", "under-share", and "over-share" decisions (training set size of 50%). The error-penalty matrix is shown on the right-hand side of the figure (cost-sensitive SVM classifier).

of under- and over-share is limited for values of α greater than 5 and it plateaus after 10, which partially justifies the values of the parameters used in Fig. 12.

6. Conclusion and future work

Mobile social networks enable users to share an increasing number of contextual information, such as their location, their activity and their co-presence with others. To simplify the sharing process and improve usability, the research community has been studying sharing preferences and developing applications that, based on several contextual features, can automate to some extent the sharing process. Machine-learning approaches have been used for specific instances of information (mostly location) or for online social network (without the notion of context).

In this paper, we have presented and evaluated a novel privacy-preserving information-sharing system (SPISM) that decides in a (semi-)automated fashion whether or not to share different types of contextual information and to what level of detail. Beyond information sharing on mobile social networks, SPISM can be applied in order to dynamically control access, by mobile apps and websites on smartphones, to personal and contextual information. Using a personalized online user-study involving 70 participants, we have shown that SPISM significantly outperforms both individual and general user-defined sharing policies, achieving up to 90% of correct sharing decisions, with only a limited cost for the user in terms of initial setup due to active learning. We also show that the system has a slight bias toward over-sharing information, which can be mitigated by introducing error-penalties for this kind of error. This has a limited effect on performance. Furthermore, our results provide significant insight into two other crucial aspects of studies related to ubiquitous computing: The users' reasons behind their sharing decisions and the participants' confidence in them. We show that the type of the requested information, in addition to the social ties of the requester, is an influential feature in the decision process.

As part of our future work, we intend to perform a large-scale real-world field study, based on our on-going Android implementation on top of the XMPP communication protocol, to gain further insight into the user's sharing decisions. Such a field study would enable us to collect more realistic data and to take into account more features (including the time since the last request, as suggested in [26], the familiarity with the requester, and the semantics of the current location), compared to the what-if scenarios used in our online survey. In addition, such a field study would enable us to analyze other interesting aspects of information sharing, such as the evolution of the users' sharing attitudes over time. Finally, we intend to investigate more sophisticated machine learning algorithms to better take into account the specificity of information-sharing in mobile social networks.

Acknowledgments

We would like to thank Vincent Etter, Peng Gao, Jens Grossklags, Urs Hengartner, Mathias Humbert, Aylin Jarrah Nezhad, Mohamed Kafsi, Bradley Malin, Xin Tang and Romain Tavenard for the fruitful discussions. This work was partially funded by

the Swiss National Science Foundation with grant 200021-138089. Parts of this work were carried out while Igor Bilogrevic and Kévin Huguenin were with EPFL.

References

[1] I. Smith, S. Consolvo, A. Lamarca, J. Hightower, J. Scott, T. Sohn, J. Hughes, G. Iachello, G. Abowd, Social disclosure of place: from location technology to communication practices, in: Proc. of Pervasive'05, pp. 134–151. http://dx.doi.org/10.1007/11428572_9.
[2] E. Toch, J. Cranshaw, P.H. Drielsma, J.Y. Tsai, P.G. Kelley, J. Springfield, L. Cranor, J. Hong, N. Sadeh, Empirical models of privacy in location sharing, in: Proc. of ACM UbiComp'10, pp. 129–138. http://dx.doi.org/10.1145/1864349.1864364.
[3] M. Benisch, P.G. Kelley, N. Sadeh, L.F. Cranor, Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs, Pers. Ubiquitous Comput. 15 (2011) 679–694. http://dx.doi.org/10.1007/s00779-010-0346-0.
[4] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, J. Rao, Understanding and capturing people's privacy policies in a mobile social networking application, Pers. Ubiquitous Comput. 13 (2009) 401–412. http://dx.doi.org/10.1007/s00779-008-0214-3.
[5] H. Wu, B.P. Knijnenburg, A. Kobsa, Improving the prediction of users' disclosure behavior…by making them disclose more predictably? in: Proc. of PPS'14, pp. 1–7.
[6] E. Toch, J. Cranshaw, P. Hankes-Drielsma, J. Springfield, P.G. Kelley, L. Cranor, J. Hong, N. Sadeh, Locaccino: a privacy-centric location sharing application, in: Proc. of ACM UbiComp'10 (Adjunct Papers), pp. 381–382. http://dx.doi.org/10.1145/1864431.1864446.
[7] G. Danezis, Inferring privacy policies for social networking services, in: Proc. of ACM AISEC'09, pp. 5–10. http://dx.doi.org/10.1145/1654988.1654991.
[8] L. Fang, K. Lefevre, Privacy wizards for social networking sites, in: Proc. of ACM WWW'10, pp. 351–360. http://dx.doi.org/10.1145/1772690.1772727.
[9] J. Xie, B.P. Knijnenburg, H. Jin, Location sharing privacy preference: analysis and personalized recommendation, in: Proc. of IUI'14, pp. 189–198. http://dx.doi.org/10.1145/2557500.2557504.
[10] K. Tang, J. Hong, D. Siewiorek, The implications of offering more disclosure choices for social location sharing, in: Proc. of ACM CHI'12, pp. 391–394. http://dx.doi.org/10.1145/2207676.2207730.
[11] B.P. Knijnenburg, A. Kobsa, H. Jin, Preference-based location sharing: are more privacy options really better? in: Proc. of ACM CHI'13, pp. 2667–2676. http://dx.doi.org/10.1145/2470654.2481369.
[12] L. Barkhuus, The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI, in: Proc. of ACM CHI'12, pp. 367–376. http://dx.doi.org/10.1145/2207676.2207727.
[13] I. Bilogrevic, K. Huguenin, B. Ağır, M. Jadliwala, J.P. Hubaux, Adaptive information-sharing for privacy-aware mobile social networks, in: Proc. of ACM UbiComp'13, pp. 657–666. http://dx.doi.org/10.1145/2493432.2493510.
[14] O. Riva, C. Qin, K. Strauss, D. Lymberopoulos, Progressive authentication: deciding when to authenticate on mobile phones, in: Proc. of USENIX Security'12, pp. 1–16.
[15] M. Miettinen, N. Asokan, Towards security policy decisions based on context profiling, in: Proc. of ACM AISEC'10, pp. 19–23. http://dx.doi.org/10.1145/1866423.1866428.
[16] X. An, D. Jutla, N. Cercone, C. Pluempitiwiriyawej, H. Wang, Uncertain inference control in privacy protection, Int. J. Inf. Secur. 8 (2009) 423–431. http://dx.doi.org/10.1007/s10207-009-0088-z.
[17] D. Anthony, T. Henderson, D. Kotz, Privacy in location-aware computing environments, IEEE Pervasive Comput. 6 (2007) 64. http://dx.doi.org/10.1109/MPRV.2007.83.
[18] C. Mancini, K. Thomas, Y. Rogers, B.A. Price, L. Jedrzejczyk, A.K. Bandara, A.N. Joinson, B. Nuseibeh, From spaces to places: emerging contexts in mobile privacy, in: Proc. of ACM UbiComp'09, pp. 1–10. http://dx.doi.org/10.1145/1620545.1620547.
[19] T. Buchanan, C. Paine, A. Joinson, U. Reips, Development of measures of online privacy concern and protection for use on the Internet, J. Am. Soc. Inf. Sci. Technol. 58 (2006) 157–165. http://dx.doi.org/10.1002/asi.v58:2.
[20] J. Wiese, P. Kelley, L. Cranor, L. Dabbish, J. Hong, J. Zimmerman, Are you close with me? Are you nearby?: investigating social groups, closeness, and willingness to share, in: Proc. of ACM UbiComp'11, pp. 197–206. http://dx.doi.org/10.1145/2030112.2030140.
[21] J.C. Tang, N. Yankelovich, J. Begole, M. Van Kleek, F. Li, J. Bhalodia, ConNexus to awarenex: extending awareness to mobile users, in: Proc. of ACM CHI'01, pp. 221–228. http://dx.doi.org/10.1145/365024.365105.
[22] G. Hsieh, K.P. Tang, W.Y. Low, J.I. Hong, Field deployment of IMBuddy: a study of privacy control and feedback mechanisms for contextual IM, in: Proc. of ACM UbiComp'07, pp. 91–108. http://dx.doi.org/10.1007/978-3-540-74853-3_6.
[23] K.P. Tang, P. Keyani, J. Fogarty, J.I. Hong, Putting people in their place: an anonymous and privacy-sensitive approach to collecting sensed data in location-based applications, in: Proc. of ACM CHI'06, pp. 93–102. http://dx.doi.org/10.1145/1124772.1124788.
[24] E. Miluzzo, N.D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S.B. Eisenman, X. Zheng, A.T. Campbell, Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application, in: Proc. of ACM SenSys'08, pp. 337–350. http://dx.doi.org/10.1145/1460412.1460445.
[25] G. Bigwood, F.B. Abdesslem, T. Henderson, Predicting location-sharing privacy preferences in social network applications, in: Proc. of AwareCast'12, pp. 1–12.
[26] R. Schlegel, A. Kapadia, A.J. Lee, Eyeing your exposure: quantifying and controlling information sharing for improved privacy, in: Proceedings of the Seventh Symposium on Usable Privacy and Security, p. 14. URL: http://dl.acm.org/citation.cfm?id=2078846.
[27] B. Liu, J. Lin, N. Sadeh, Reconciling mobile app privacy and usability on smartphones: could user privacy profiles help? in: Proc. of WWW'14, pp. 201–212. http://dx.doi.org/10.1145/2566486.2568035.
[28] P. Saint-Andre, Extensible messaging and presence protocol (XMPP): core, in: RFC 6120 (Proposed Standard), 2011. URL: http://www.ietf.org/rfc/rfc6120.txt.
[29] A.J. Brush, J. Krumm, J. Scott, Exploring end user preferences for location obfuscation, location-based services, and the value of location, in: Proc. of ACM UbiComp'10, pp. 95–104. http://dx.doi.org/10.1145/1864349.1864381.
[30] S. Consolvo, I. Smith, T. Matthews, A. LaMarca, J. Tabert, P. Powledge, Location disclosure to social relations: why, when, & what people want to share, in: Proc. of ACM CHI'05, pp. 81–90. http://dx.doi.org/10.1145/1054972.1054985.
[31] B. Priyantha, D. Lymberopoulos, J. Liu, Littlerock: enabling energy-efficient continuous sensing on mobile phones, IEEE Pervasive Comput. 10 (2011) 12–15. http://dx.doi.org/10.1109/MPRV.2011.28.
[32] J. Ross, L. Irani, M.S. Silberman, A. Zaldivar, B. Tomlinson, Who are the crowdworkers?: Shifting demographics in mechanical turk, in: ACM CHI'10 (Extended Abstracts), pp. 2863–2872. http://dx.doi.org/10.1145/1753846.1753873.
[33] W. Mason, S. Suri, Conducting behavioral research on Amazon's mechanical turk, Behav. Res. Methods 44 (2012) 1–23. http://dx.doi.org/10.3758/s13428-011-0124-6.
[34] S. Patil, Y. Gall, A. Lee, A. Kapadia, My privacy policy: exploring end-user specification of free-form location access rules, in: Proc. of FC'12, pp. 86–97. http://dx.doi.org/10.1007/978-3-642-34638-5_8.
[35] P.J. Lavrakas, Encyclopedia of Survey Research Methods, SAGE Publications, 2008, http://dx.doi.org/10.4135/9781412963947.