On the Feasibility of User De-Anonymization from Shared Mobile Sensor Data

Nicholas D. Lane[†], Junyuan Xie[‡], Thomas Moscibroda[†], Feng Zhao[†]
[†]Microsoft Research Asia, [‡]University of Science and Technology of China

Abstract

Underpinning many recent advances in sensing applications (e.g., mHealth) is the ability to safely collect and share mobile sensor data. Research has shown that even from seemingly harmless sensors (e.g., accelerometers, gyroscopes, or magnetometers) an ever expanding set of potentially sensitive user behavior can be inferred. Providing robust anonymity assurances is a principal mechanism for protecting users when data is shared (e.g., with medical professionals or friends). In this paper, we study the feasibility of user de-anonymization from mobile sensor datasets routinely collected on commodity devices (e.g., smartphones). We perform a systematic investigation to quantify the threat of de-anonymization using existing sparsity-based techniques adapted to exploit mobile sensor data characteristics. This preliminary study indicates significant threats to user anonymity exist within shared mobile sensor data and further investigation is warranted.

1 Introduction

Datasets containing information about actions or preferences of individuals have become increasingly publicly available. And even though the privacy-risks of publishing such user data are well-known and research has tried to mitigate these risks, unexpected privacy leaks nevertheless occur frequently. In the infamous AOL query logs [3] or the Netflix challenge incidents [12], for example, it was possible to de-anonymize individual user records with alarming ease in spite of conscious efforts taken by the data provider.

On mobile devices, these privacy risks are magnified by two overlapping trends. First, advances in sensing platforms (e.g., smartphones) are making the collection of large-scale multi-modality datasets by users more commonplace, and second, the trend towards storing personal data, and sharing it among friends, medical experts, or other users, for example in social networks, mHealth, games, etc. In other words, as sensors allow us to conveniently capture more and more aspects of our lives, and thus enable a rich data characterization of our activities, emotions, habits, etc, sharing such sensory data (regardless of application) will make it increasingly difficult to protect mobile user privacy and anonymity.

For many years, the mobile computing research community has studied questions that relate to privacy and sensor data, predominantly in the context of *location privacy* [16]. However, researchers are discovering location-oriented sensors are not the only source of concern and finding other sensors modalities can also introduce a variety of new privacy threats (e.g., [15, 9]). Similarly, in this study we find that sensors, such as accelerometers, gyroscopes, magnetometers, or barometers, which at first glance may appear innocuous, can lead to significant new challenges to user *anonymization*. Research on activity recognition has shown that, collectively, these sensors are capable of inferring an ever-growing range of human activities, and even moods (e.g., [7, 10, 17]). We complement this research by showing that such sensor data can also be used to de-anonymize users, i.e. to obtain sensitive information about specific users from anonymized datasets.

In this paper, we quantify the feasibility of *de-anonymization* attacks based on the kind of sensor data that is commonly collected with commodity smartphones. Our study is based on a systematic analysis of a large representative activity recognition dataset from the ALKAN project [11]. We build on analysis frameworks that had been used in the database and security communities to study de-anonymization properties in non-mobile datasets such as transaction logs, purchase records or preference/ranking lists. We first show that mobile sensor data exhibits similar structural sparsity properties as these non-mobile datasets, and that therefore popular algorithmic approaches for de-anonymization can be successfully applied to mobile sensor data as well. Specifically, we consider two scenarios (1) de-anonymization with auxiliary information, and (2) de-anonymization by linking two identities from separate datasets. In the first case, we show that an adversary can – even with very limited background knowledge about an individual – use the sensor data to identify an individual within an anonymized dataset with high probability, and to thereby learn (potentially sensitive) information about him or her. In the second scenario, we recognize when two identities in dif-

PhoneSense'12, November 6, 2012, Toronto, ON, Canada.

ferent datasets are likely generated by the same individual – in such cases, a weakly protected identity, associated with otherwise innocuous data, can be used to de-anonymize a carefully protected private dataset.

2 Threat Model

In what follows, we describe the definitions and approaches that frame our investigation into the de-anonymization risk posed by shared mobile sensor data.

Shared Mobile Sensor Data. Societal norms for mobile sensing applications are still in a state of flux; we do not yet know precisely what will become commonplace in terms of: sensor data collected, inferences made and how – and with whom – this information will be shared. In this work, rather than focus on a set of specific applications (e.g., mHealth, sensor-enhanced social networking) we consider the following generic data-centric assumptions in which de-anonymization threats can manifest:

Users participate in one or more sensing applications. Each application makes use of mobile sensors, such as those embedded in smartphones [13] to gather data about the user. Select events in the user's life (e.g., transportation mode, physical activity) are inferred from this data, as required by application needs. Some representation of these inferences are subsequently shared with other people – friends, medical practitioners, insurance companies – which potentially include an adversary. We focus our study on the dangers of sharing these inference-based representations. While clearly representations will be application-specific, two fundamental representation types we consider are: (1) event-oriented publishing of inferences and their derivatives (e.g., Nike+ [1] and similar exercise applications which publicly publish user information including jogging route, event time, duration); and, (2) periodic publishing of a summary of behavior (e.g., the Snapshot discount program from Progressive [2] which requires a sensor-based summary of driver behavior to be shared with the company).

Importantly, the mechanism for sharing within the applications must maintain a level of user anonymity appropriate for (1) the sensitivity of the data and (2) the type of relationship between the user and the people with whom the data is shared. Various forms and degrees of anonymity will be required, for example: the use of a digital persona that although tied to a stream of shared data is deliberately separated from a real-world identity. The objective of an adversary is to weaken the original level of user anonymity under which mobile sensor data is shared and at the extreme connect the data to a real-world identity.

De-anonymization by Sparsity. Researchers have discovered that maintaining user anonymity within sparse datasets is surprisingly difficult [6, 12, 8, 19]. Within this research community, the term sparsity refers to datasets in which an individual user or identity can be distinguished from others in a dataset by only a few select rarely occurring user attributes. This characteristic has been found to enable a variety of de-anonymization approaches based on differing adversary assumptions. Typically, approaches assume the availability of small amounts of what is termed, `auxiliary information` – additional information beyond what is contained in the dataset, usually focused on the target user of interest to the adversary. The potential danger of auxiliary information is not surprising; however, the key finding in this research is that sparsity allows dangerous forms of de-anonymization while only requiring alarmingly weak assumptions regarding auxiliary information.

Sparsity-based de-anonymization algorithms have been evaluated against data, such as, movie preferences and viewing habits or search queries logs and shopping purchase history. Shared mobile sensor data has many similarities with such datasets and is likely prone to sparsity. One seminal case study of sparsity-based de-anonymization was performed on the Netflix challenge dataset [12] containing movie ratings; users were found to be vulnerable to an adversary using auxiliary information such as, user movie preferences, titles of movies watched by the user, or even just the dates of when movies were watched. Similar opportunities may exist in sensor data as a mixture of infrequently occurring events or user characteristics may be equally distinctive – especially when data spans an extended period of time.

Relationships between Diverse User Activities. The previous study of sparsity-based de-anonymization has largely focused on datasets that contain a narrow range of user behavior (e.g., user preferences movies or restaurants). Unlike these other datasets, shared mobile sensor data will capture a wider variety of everyday user activities (e.g., social activity, exercise, commuting) which are more greatly influenced by a set of high-level user characteristics and constraints. As a result, the opportunity for relationships to exist between different types of activities captured in mobile sensor data is increased – even if initially the captured activities do not appear to closely related. For example, event-based transportation mode choices by a user (such as, subway, car or bike usage) can correlate a coarse periodic summary of exercise (such as, walking, biking or jogging usage).

This is important to sparsity-based de-anonymization for two reasons. First, in practice it increases the levels of sparsity between users as related events – unlike those that are independent – can combine together to more strongly differentiate one user from another. Second, it broadens the range of viable auxiliary information for an adversary to use; this is because of the increased opportunity for relationships to exist between data available to the adversary and data contained within shared mobile sensor data.

3 De-Anonymization Framework

We use the following two-stage framework to explore sparsity-based de-anonymization. In this framework, we adapt the `SCORE` algorithm [12] from prior de-anonymization research by incorporating activity relationship mining.

By using this framework, an adversary with auxiliary information concerning a target user can identify an activity stream belonging to the target within an anonymized set of streams collected from other users. We believe this framework is representative of a wider class of de-anonymization approaches applicable to shared mobile sensor data.

Stage One - Relationship Mining. Our framework begins by extracting a set of rules that connect different aspects of two captured activity types. Each of these rules

should provide some small hint that two activities either were likely performed by the same person, or just as usefully – the two activities were likely *n*ot performed by the same person. Rules may discover, for instance, that two activities – such as, a long afternoon run and a morning commute to work – seldom occur on the same day. We find in our experiments that the most effective rules for cross-domain activities are based on connecting anomalies. For example, consider a weekday but during a vacation period – anomalous observation in certain activity types (e.g., late night prolonged conversations and music) can be related with others, such as, anomalous sleep duration.

In practice, this stage requires training data in which cross-domain activities are collected from each user. Sources of such data include: (1) multi-modality sensor datasets – e.g., audio and accelerometer data – from which a variety of activities can be automatically extracted; and (2) users who are not overly concerned with anonymity and share various activities with weak (or no) forms of identity protection.

Although we focus on rules between different activity types useful rules can still exist between the same type of activity. For example, rules that capture symmetry in activities, such as, commuting – a morning subway ride is commonly paired with a subway ride later that day; with transportation mode and duration rarely being vastly different within the same day. Or another example is activity counting, in which certain activities occur within a particular range of intra-day frequency – such as, prolonged exercise is typically once a day or meal frequency is often around three.

In our initial experiments (see Section 4) we use simple association rule mining [4] to learn activity relationships. We mine over a series of descriptors (e.g., day of week, discretized duration) for each instance of an activity type, with some descriptors designed to capture basic forms of anomalies (e.g., a binary indicator variable tied to when activity duration is two standard deviations beyond average).

Stage Two - `SCORE` Algorithm. Informally `SCORE` functions as follows. The adversary provides `SCORE` with two inputs: (1) auxiliary information about the target user and (2) a stream of one or more activity types regarding a single anonymized user; the algorithm will output a score for the input pair – a high score indicates the stream is likely generated from the target user, while a low score suggests the opposite. An adversary can then proceed to test their auxiliary information against all anonymized streams in a user population, allowing them to identify a likely mapping between a target user and a specific stream based on the test pair with the highest score. It will not always be necessary for the adversary to map a target user to a single stream, thus the adversary may ultimately decide to use the $k-$highest scores – which indicates the user is within this group of k streams – rather than select only one.

The function within `SCORE` is a weighted sum of separate similarity measurements between auxiliary information and a candidate activity stream. The key to its operation is how similarity is defined. First, both adversary information and the candidate activity stream are normalized into a vector representation based on the information they contain. Vector elements are data dependent, for example, binary indicators of if an activity occurred (e.g., did the user ride the subway) or if an activity *type* occurred (e.g., was a transportation mode used) – time (e.g., duration, time of day, day of week) and location semantics can also be added. This initial vector is then extended through the use of the association rules learned in stage one. The purpose is to create additional vector elements that have values in both the auxiliary information and candidate stream. The similarity between these two vectors is then simply the L-2 norm with the inclusion of two additional terms. First, to penalize weak association rules (those rules prone to error) we incorporate the rule confidence value provided by the association rule miner – if the vector element is not based on an association rule this term is set to 1, otherwise it expresses how frequently the rule is correct. Second, to emphasize rare vector elements (that are highly discriminative) we compute a weight based on how often the element value appears in the dataset.

To avoid false positive results we must also estimate the confidence in output of `SCORE`. We estimate confidence by the difference in score between the highest and second highest candidate streams normalized by the number of user streams in the dataset. Experimentally, a threshold can be learned to indicate if a proposed pairing by `SCORE` is reliable.

4 Feasibility Experiments

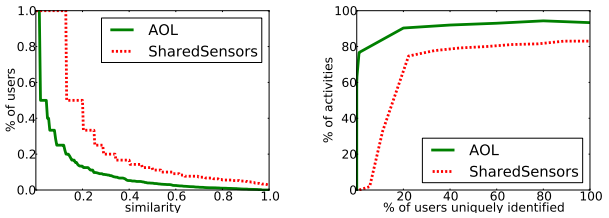
In this section, we describe a series of feasibility experiments that investigate the de-anonymization threat posed by shared mobile sensor data.

Datasets. We use a public activity recognition dataset from the ALKAN project [11], which we subsequently refer to as `SharedSensors`. This dataset contains hand-labeled data from more than 200 users and consists of over 35,000 activities. All data is gathered from a combination of iOS and Android mobile devices. A diverse range of activities is included, specifically: social events (e.g., eating, party, conversation, coffee), physical activities (e.g., run, yoga, strength training) and transportation mode (e.g., bus, car, bike, plane). We use user labeled events in our analysis, along with activities extracted automatically from unlabeled portions of the data, namely: physical activities and transportation modes.

In our sparsity comparison experiments we use an additional dataset, AOL query logs [3], that have been previously studied within the de-anonymization literature. We refer to this dataset as `AOL`. The dataset contains the anonymized search query logs of approximately 650,000 users spanning three months. `AOL` has been shown to be sufficiently sparse to be vulnerable to a variety of de-anonymization strategies – making it an interesting baseline to compare against `SharedSensors`.

Permuted Activity Representations. Given the uncertainty as to how mobile sensing applications will use and share sensor data, for all experiments, we permute the representation of activities within `SharedSensors`. Each experiment is repeated for each permutation. Thus, our results are agnostic to a particular application scenario. `SharedSensors` is permuted as follows:

All user activities are represented as (1) an activity event along with a timestamp; (2) a binary indication of activity



(a) Sparsity (b) Uniqueness

Figure 1. We find `SharedSensors` to be sparse, and so susceptible to sparsity-based de-anonymization.

type occurring within some time interval; (3) a frequency count of an activity type within a time interval and, (4) the total duration of all occurrences of the activity within a time interval. We rely solely on duration to report an aggregate for activities as it generalizes easily to many other aggregates (e.g., calories, speed, distance) – that are often computed by applying a scaler to duration. During the permutations, we consider time intervals of an hour, day and week. Although we use the term "stream" when reporting results as our permutations include duration, frequency and time intervals of day and week; thus we still test both the de-anonymization of activity *summaries* along with event-based activity streams.

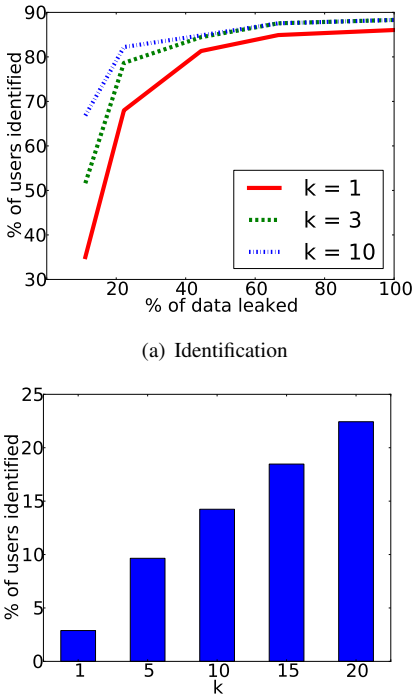
Threat Scenarios. Our experiments consider two specific de-anonymization threats commonly found in previous de-anonymization studies. Both threats utilize the framework described in Section 3 but differ in (1) the type and source of auxiliary information and (2) the form of de-anonymization achieved by the adversary.

Threat 1: Identification. In our first threat, the assumed auxiliary information of the adversary is a collection of activities performed by the target user. The adversary's objective is to identify which of the anonymized activity streams belongs to the target user. Auxiliary information may be collected by observing the target user directly or from already available public sources (e.g., twitter, foursquare).

Threat 2: Linkability. In the second threat, the adversary's auxiliary information is another stream of activities itself. The adversary uses the de-anonymization framework to discover if two different streams of activities correspond to the same real-world user. This threat is serious as it may (1) allow additional information (time, location) to be connected with user activities in a way never anticipated by the user; and, (2) different activity streams may have varying levels of anonymity protection – a weakly protected stream (easily tied to a real-world identity) may be linked to an anonymous stream of sensitive activities. This threat also exists even if the real-world identities of the streams are unknown.

Data Sparsity Comparison. We begin our experiments by investigating data sparsity. As previously described, sparsity is a key property that underpins a variety of existing de-anonymization approaches. Evidence of similar levels of sparsity in `SharedSensors` (relative to `AOL`) would suggest that it is prone to existing forms of de-anonymization.

Figure 1(a) is a CDF of the sparsity across the whole user population, we compute similarity as described in Section 3. From Figure 1(a) we learn `SharedSensors` is indeed sparse,



(a) Identification

(b) Linkability

Figure 2. Accuracy results for de-anonymization assuming two threat scenario – Identification and Linkability

and in-fact has similar levels of sparsity to `AOL`. For the majority of activity streams, across all permutations of activity representation, there is not a single stream with a similarity above 0.18 in the entire dataset. Figure 1(b) shows what fraction of the activities stream in `SharedSensors` is required to uniquely distinguish the stream on average. We find that for around 79% of streams (again, regardless of activity representation) only 20% of the user activities are necessary before the stream is unique within the dataset.

Threat 1: Identification. Our next experiment examines how easily users can be de-anonymized when an adversary is aware of a small number of activities from an activity stream.

Figure 2(a) shows, on average, the relationship between auxiliary information and the accuracy of de-anonymization. In this experiment we assume the adversary has auxiliary information consisting of a few isolated activities from the stream of the target user, we then vary the amount of auxiliary information and repeatedly attempt to de-anonymize. Each line in Figure 2(a) represents a different value of k (viz. 1, 3, 10), which allows the user to be identified within a k-sized group of streams. As this result is sensitive to how common/rare an activity happens to be within the larger dataset we repeat it for all combinations of activities – both for stream content and auxiliary information – and report the average. We find that if k is set to 3 only 10% of the sensor stream is required by an adversary to successful identify the correct stream for 53% of the users.

Since our de-anonymization framework reports many false positives the estimation of confidence in any de-anonymization result is crucial. We find in this experiment

by carefully selecting a confidence threshold we can eliminate 70% of false positives while having negligible impact on de-anonymization accuracy.

Threat 2: Linkability. Our final experiment considers the linkability de-anonyimization threat, i.e., how successfully an adversary can link one stream to another.

Figure 2(b) shows the fraction of streams successfully linked for varying values of k. We see that between 10% and 23% of the population of streams can be tied to k other sensor streams in groups around 5 and 20 (within a total user population of 200 people). During experiments, we observe that some activity representations of `SharedSensors` even allow up to 32% of the streams to be linked (k set to 20). Again, due to the sensitivity of results to which activities are included in streams, we permute the content of streams so that all combinations of activity types are tested.

Under this threat, we again find our confidence estimator is reasonably effective. We find again by careful selection of a threshold we can eliminate 81% of false positives while maintaining similar accuracy levels as reported above.

5 Discussion

We briefly describe the limitations of our exploratory study before discussing potential solutions.

Study Limitations. Our results are intended only to highlight a potential threat to the future adoption of mobile sensing applications. While we utilize one of the largest datasets of its kind to perform our experiments, it still contains a relatively small number of people (≈ 200). As such, we make no claims as to the result of repeating our experiments with much larger-scale datasets. Our experiments identify certain scenarios of shared sensor data (i.e., representations of activity, combinations of shared activity) where as many as 32% of users are de-anonymized. However, it remains difficult to predict which of these scenarios will become popular with the public. For this reason, we present aggregate experiment results that permutate: (1) which activities are shared and (2) how activities are represented.

Towards Solutions. Developing solutions to threats such as de-anonymization will be challenging. Conventional solutions, such as perturbing the data, also diminish the utility gained from shared data. This situation is complicated by the variety of different activity inferences that are possible from the same sensor data, making it hard to control the information adversaries may gain. Perhaps a first step will be the development of new metrics that can quantify the threat of a particular combination of shared sensor data and user populations. Such metrics could allow users to understand the implications of using certain applications, or even support automated safety features built into future smartphones.

6 Related Work

Concern for maintaining mobile user privacy and anonymity is mounting as we are faced with rapid progress in sensing technology (e.g., [18]). Historically, the majority of this concern has been placed on location-oriented sensors and the microphone. However, awareness of the threat posed by less obvious sensors has recently been gaining greater appreciation [15]. For example, [5] shows even low-level sensors can expose certain demographic attributes of a users.

The rise of participatory sensing has also highlighted key system design questions [14] as to how sensor data can be collected and shared while protecting contributing users. Towards safer data sharing, participatory sensing researchers have investigated, for instance, how aggregate anonymous analysis can occur while limiting the leakage of user data [9].

This study is perhaps most closely related to de-anonymization investigations performed on non-sensor data. Our results build on existing frameworks that exploit data sparsity [6, 12, 8, 19]. However, we examine a new type of data and de-anonymization scenarios previously not studied.

7 Conclusion

In this paper, we have presented initial evidence of the de-anonymization threat posed by shared mobile sensor data. We have examined this problem through the lens of existing methods of de-anonymization that leverage dataset sparsity. Preliminary results of our study suggest the same challenges to maintaining user anonymity – that have been verified in datasets including movie ratings and purchase histories – are likely to exist when mobile sensor data is shared. Our findings indicate further study of this problem is warranted.

8 References

[1] Nike+. `http://nikeplus.nike.com/`.

[2] Snapshot. `http://www.progressive.com/auto/snapshot.aspx`.

[3] NYTimes. AOL Removes Search Data on Group of Web Users. `http://www.nytimes.com/2006/08/08/business/media/08aol.html`.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, August 2006.

[5] S. Brdar, D. Ćulibrk, V. Crnojević. Demographic Attributes Prediction on the Real-world Mobile Data. In *MDC '12*.

[6] A. Datta, D. Sharma, A. Sinha. Provable De-Anonymization of Large Datasets with Sparse Dimensions. In *POST '12*.

[7] A. Ali, S. Hossain, K. Hovsepian, M. Rahman, K. Plarre, S. Kumar. mPuff: Automated Detection of Cigarette Smoking Puffs from Respiration Measurements. In *IPSN'12*.

[8] D. Frankowski, D. Cosley, S. Sen, L. Terveen, J. Riedl. You Are What You Say: Privacy Risks of Public Mentions. In *SIGIR '06*.

[9] R. Ganti, N. Pham, Y. Tsai, T. Abdelzaher. Poolview: Stream Privacy for Grassroots Participatory Sensing. In *Sensys '08*.

[10] R. LiKamWa, Y. Liu, N. Lane, L. Zhong Can Your Smartphone Infer Your Mood? In *PhoneSense'11*.

[11] Y. Hattori, S. Inoue, G. Hirakawa. A Large-scale Gathering System for Activity Data with Mobile Sensors. In *ISWC '12*.

[12] A. Narayanan, V. Shmatikov. Robust De-Anonymization of Large Sparse Datasets. In *S&P '08*.

[13] E. Miluzzo, J. Oakley, H. Lu, N. Lane, R. Peterson, A. Campbell Evaluating the iPhone as a Mobile Platform for People-centric Sensing Applications In *UrbanSense'08*.

[14] K. Shilton, J. Burke, D. Estrin, R. Govindan, M. Hansen, J. Kang, M. Mun. Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing. In *TPRC '09*

[15] A. Raij, A. Ghosh, S. Kumar, M. Srivastava. Privacy Risks Emerging from the Adoption of Innocuous Wearable Sensors in the Mobile Environment. In *CHI '11*

[16] J. Krumm. A Survey of Computational Location Privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[17] N. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A. Campbell, F. Zhao. Enabling Large-scale Human Activity Inference on Smartphones using Community Similarity Networks (CSN). In *UbiComp '11*.

[18] R. Caceres, L. Cox, H. Lim, A. Shakimov, A. Varshavsky. Virtual Individual Servers as Privacy-preserving Proxies for Mobile Devices. In *MobiHeld '09*.

[19] E. Zheleva, L. Getoor. To Join Or Not To Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *WWW '09*.