Report from Dagstuhl: The Liberation of Mobile Location Data and its Implications for Privacy Research *

Gennady Andrienko[a] *gennady.andrienko@iais.fraunhofer.de*
Aris Gkoulalas-Divanis[b] *arisdiva@ie.ibm.com*
Marco Gruteser[c] *gruteser@winlab.rutgers.edu*
Christine Kopp[a] *christine.kopp@iais.fraunhofer.de*
Thomas Liebig[d] *thomas.liebig@tu-dortmund.de*
Klaus Rechert[e] *klaus.rechert@rz.uni-freiburg.de*

[a]Fraunhofer IAIS, St. Augustin, Germany
[b]IBM Research, Dublin, Ireland
[c]Rutgers University, New Brunswick, U.S.A
[d]TU Dortmund University, Dortmund, Germany
[e]University of Freiburg, Freiburg i. B., Germany

With the emergence of the mobile app ecosystem, user location data has escaped the grip of the tightly regulated telecommunication industry and is now being collected at unprecedented scale and accuracy by mobile advertising, platform, and app providers. This position paper is based on discussions of the authors at the Dagstuhl seminar on Mobility Data Mining and Privacy. It seeks to highlight this shift by providing a tutorial on location data flows and associated privacy risks in this mobile app ecosystem. Moreover, it reflects on the implications of this shift to the mobile privacy research community.

I. Introduction

The phones we carry around as we go about our daily lives do not only provide a convenient way to communicate and access information but also pose privacy risks by collecting data about our movements and habits. For example, they can record when we get up in the morning, when we leave our homes, whether we violate speed limits, how much time we spend at work, how much we exercise, whom we meet, and where we spend the night. The places we visit during our everyday activities allow inferences about not just one but many potentially sensitive subjects: health, sexual orientation, finances or creditworthiness, religion, and political opinions. For many, such inferences can be embarrassing, even if they are untrue and simply misinterpretations of the data. For some, this movement data can even pose a danger of physical harm, such as in stalking cases, or may lead to financial damage, such as in cases of burglaries due to knowledge of peoples' absence from certain locations.

These risks have been amplified by the emergence of smartphones and the app economy over the last few years. We have witnessed a fundamental shift in mobility data collection and processing from a selected group of tightly regulated cellular operators to a complex web of app providers and Internet companies. This new ecosystem of mobility data collectors relies on a more sophisticated mix of positioning technologies to acquire increasingly precise mobility data. In addition, smartphones also carry a much richer set of sensors and input devices, which allow collection of a diverse set of other data types in combination with the mobility data. Many of these types of data were previously unavailable. While individual aspects of these changes have been highlighted in a number of articles as well as in a string of well-publicized privacy scandals, the overall structure of current mobility data streams remains confusing.

This position paper intends to survey this new mobility data ecosystem and to discuss the implications of this broader shift. The survey includes the types of data collected, the positioning technologies involved, and the purpose of the collection as well as privacy threats resulting from such data. We begin in Section II by reviewing how cellular networks have to monitor the location of subscriber phones to be able to route incoming calls and to provide the ability to locate emergency callers (known as E911 in the US). We survey the technologies used by operators to determine the position of a phone, describe how this location information at operators is stored, and how it is accessed by law enforcement entities [28, 27] and selected application service providers. We then describe how smartphone apps can directly acquire position and movement information from the handset, without assistance from cellular operators. This can involve different positioning technologies based

*Revised, enhanced Dagstuhl Report [11] – Working Group on Cellular Data. Authors are listed in alphabetical order.

on crowdsourced maps of WiFi access points and cell sectors, which poses additional privacy risks. We further discuss several classes of apps, such as location-based applications that collect position information to deliver location-targeted information and advertising-supported apps that collect position and mobility information for targeted ads. In addition to the commercial use of mobile network and application data, we present the scientific point of view on the data.

In Section III we discuss privacy threats and risks that arise from data collection. In particular, we distinguish risks for the three types of collected data. First, we assume that location information is available along with a personal identifier. Second, we relax this notion and assume (a series of) anonymous location data observations. Finally, we consider how location information about a user may be derived even though no georeference is included in the collected data. We conclude our work with a section on the manifold implications of this rapidly evolving mobility data ecosystem. We find that it is difficult to understand the data flows, apparently even for the service providers and operators themselves [7, 18, 25]. There appears to be a much greater need for transparency, perhaps supported by technical solutions that monitor and raise awareness of such data collection. We find that location data is increasingly flowing across national borders, which raises questions about the effectiveness of current regulatory protections. We also find that applications are accessing a richer set of sensors, which allows cross-referencing and linking of data in ways that are not yet fully understood.

II. Location Data Collection

In order to assess privacy risks posed by location and mobility data, in a first step an overview on current data collection practices and the general characteristics of such data is given. We distinguish three groups of data collectors (observers): mobile network operators (MNO), mobile platform service providers (MPSP), and application service providers (ASP). While for the first group of observers (MNOs), location data is generated and collected primarily due to technical reasons, i.e. efficient signaling, in the case of MPSP and ASPs location information is usually generated and collected to support positioning, mapping, and advertising services and to enable the offering of various kinds of location based services. Figure 1 provides a schematic overview on location data generated by mobile phones but also highlights the specific components and building blocks of mobile phones, which are controlled by the different entities. Furthermore, available location data originating from the aforementioned layers may be re-used to support various new (third-party) businesses. Typically the data is then anonymized, or aggregated, in some way before shared with to third parties.

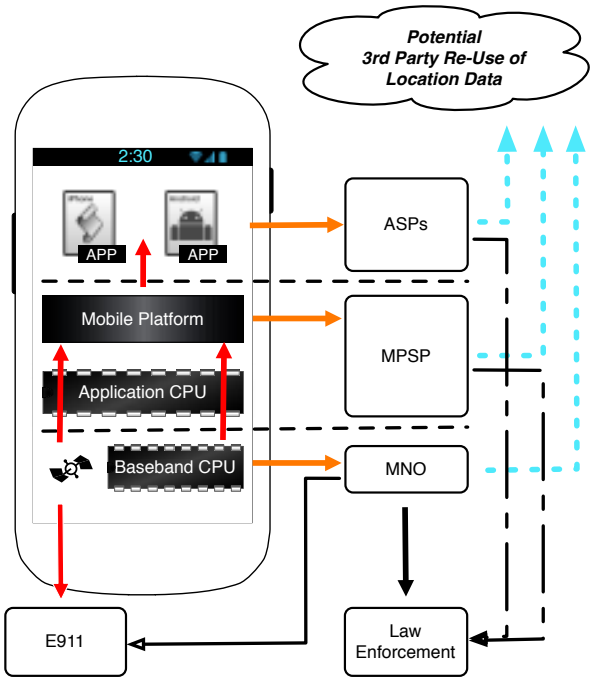

Figure 1: A schematic overview on a today's smartphone, its essential building blocks and their controllers illustrating the generation and information flow of location data.

The primary – and usually most accurate – source for location information that we associate with a mobile phone is a Global Navigation Satellite System (GNSS) as, for example, GPS. However, any mobile radio-based communication system can be either used to acquire location data (e.g. through triangulation) or location data is implicitly generated (e.g. by cell / network association). Section II.A discusses details on location information in mobile telephony networks. WiFi-based positioning and various combinations thereof are briefly discussed in Section III.A.

In today's smartphones the GNSS unit is usually bundled with a so-called baseband processor, which is an autonomous CPU running the mobile network stack, e.g., handling calls, network attachment, etc. A MNO is required by regulation to provide a device's location within 50 - 300 meters in the case of an emergency (E-911) [1]. Due to the bundling of GNSS and baseband CPU, accurate positioning of an individual

[1] FCC Enhanced 911 Wireless Service, http://www.fcc.gov/pshs/services/911-services/enhanced911.

device becomes possible even in situations where cellular network positioning is difficult or too inaccurate (e.g. in rural areas).

Baseband and application CPU are both conceptually and physically separated. On today's smartphones the user interacts solely with the mobile platform, i.e. a sophisticated mobile operation system, which runs on the application (generic) CPU. The mobile platform concentrates various location information sources into a single location API and further implements and enforces user-defined privacy policies. The application layer (so-called apps) connects to such an API in order to retrieve and use location data.

II.A. Collection and Usage of Mobile Telephony Network Data

As an example for mobile telephony networks we discuss the widely deployed GSM infrastructure, as its successors UMTS (3G) and LTE (4G) have a significantly smaller coverage and share most of its principal characteristics. A typical GSM network is structured into cells, each served by a single base transceiver station (BTS). Larger cell-compounds are called location areas. To establish a connection to the mobile station (MS), e.g. in the case of an incoming connection request, the network has to know if the MS is still available and in which location area it is currently located. To cope with subscriber mobility the location update procedure was introduced. Either periodically or when changing the location area, a location update is triggered. The time lapse between periodic location updates is defined by the network and varies between infrastructure providers.

Additionally, the infrastructure's radio subsystem measures the distance of phones to the serving cell to compensate for the signal propagation delay between the MS and BTS. The timing advance (TA) value (8-bit value) is used to split the cell radius into virtual rings. In the case of GSM these rings have a size of roughly 550 m in diameter. The TA is regularly updated and is sent by the serving infrastructure to each mobile phone. In the following, we provide details on the different positioning methods available to MNOs.

II.A.1. Active Positioning

To obtain a position of an idle phone, active communication, e.g., voice/text/data transmission but also protocol related communication like location updates, IMSI attach, etc., is required. Thus, the network either has to wait for the next active period of the MS (e.g. phone call, location update) or has to trigger MS activity. This can be achieved by transmitting a so-called *silent text message* to force an active communication without raising the user's awareness.

Active positioning methods yield immediate and more accurate results than passive methods. These methods work without special requirements on the mobile station and achieve a positioning accuracy of up to 50 m in urban areas (TDOA) [38]. However, there are additional costs involved (e.g. network utilization, suitable infrastructure as, e.g., SLMC) and, therefore, an incentive and a dedicated target is required. In general, active GSM positioning methods are not suitable for location tracking of masses, but they are a quite accurate tool to track individuals.

II.A.2. Call Data Records

For billing purposes, the so-called call data records (CDR) are generated. This datum usually consists of the cell-ID where a call has been started (either incoming or outgoing), the cell where a call has been terminated, the start time, duration, ID of the caller, and the phone number called. A typical GSM cell size ranges from a few hundred meters in diameter to a maximum size of 35 km. In a typical network setup a cell is further divided into three sectors. In this case the sector ID is also available and part of a call record. CDRs are usually not available in real-time. However, MNOs store CDRs for a certain time span, either because of legal requirements (e.g. EU data retention directive [14]) or accounting purposes.

II.A.3. Alternative, Non-Standard Positioning Methods

Another, (non-standard) method to determine the location of a MS is to make use of received signal strength measurement results. Usually based on databases derived from signal propagation models used during the planning phase of the infrastructure, this data can be exploited to create a look-up table for signal measurements to determine the MS's location. Based on the cell, TA and received signal strength of the serving cell as well as the six neighboring cells, Zimmermann et al. achieved a positioning accuracy of below 80 m in 67% of cases and below 200 m in 95% of cases in an urban scenario [42]. In a recent study using RSSI in combination with map information and movement prediction Anisetti et al. achieved in 50% of cases less than 19 m and in 95% of cases less than 64 m of accuracy [6]. While these methods seem cost effective (no changes to infrastructure or protocols re-

quired) and technically feasible, they are usually not yet available in current deployments.

II.A.4. Re-Use of Cellular Data

Mobile telephony networks and their physical characteristics are able to help locating mobile phone users in the case of an emergency and may be a valuable tool for search and rescue (SAR) [29]. For instance, Bengtsson et al. analyzed post-disaster populations displacement using SIM-card movements to improve the allocation of relief supplies [8].

Furthermore, location information gathered through mobile telephony networks is now a standard tool for crime prosecution and is used by the EC Data Retention Directive with the aim of reducing the risk of terror and organized crime [14]. As an example, law enforcement officials seized CDRs over a 48 hour timespan resulting in 896,072 individual records containing 257,858 call numbers after a demonstration in Dresden, Germany, became violent [27]. However, the degree of data collection is jurisdictionally disputed, as a subsequent district court decision in the Dresden case showed [35]. Further, the police of North Rhine-Westphalia issued 225,784 active location determinations on 2,644 different subjects in 778 preliminary proceedings in 2010 [33]. While in principle law enforcement could also collect location- and movement-data from MPSP and ASPs, difficulties arise if such data is stored outside of the respective jurisdiction.

Additionally, commercial services are based on the availability of live mobility patterns of larger groups. (e.g. for traffic monitoring or location-aware advertising [26]). Thus, location information of network subscribers might be passed on to third parties. Usually, subscribers are neither aware of the extent of their information disclosure (just by carrying a switched-on mobile phone), nor of how and by whom the collected data is used. Even sporadic disclosure of location data, e.g. through periodic location updates, can disclose a user's frequently visited places (i.e. preferences) in an accuracy similar to continuous location data after 10-14 days [36].

II.B. Collection and Usage of Data Through MPSPs and ASPs

Mobile advertising is among the most rapidly growing advertising media, most likely doubling its yearly revenue over the next years by a prediction of [15]. The availability of new powerful mobile devices (e.g. Smartphones) in combination with comprehensive and affordable mobile broadband communication has given rise to this new generation of advertising media, which makes it possible to deliver targeted information in a context-aware and personalized manner. Personal information, such as a user's current location and personal preferences, are prerequisite for a tailored advertisement delivery. Consequently, MPSPs and ASPs are interested to profile users, and personal data is disclosed in an unprecedented manner to various (unknown) commercial entities which poses serious privacy risks.

In general, mobile advertising is similar to online advertising involving the four main entities: *advertiser*, *publisher*, *client* and *broker* [20]. The advertiser is interested in promoting his goods or services and provides the ad content. Publishers (e.g. websites, apps) provide the space to place an advertisement (e.g. in form of a banner), and clients refer to the devices that receive the advertisement. Brokers play the main role in the advertising ecosystem as they connect advertisers, publishers and clients. In a mobile advertising setting they are also known as *ad networks*. Their task is the optimal placement of advertisements given the available pool of advertising space, advertisements, and users that can be addressed. Smartphones are ideal for targeted advertising because they are typically used by a single user. In consequence, brokers are highly interested in personal information stored or available on smartphones, since it allows to infer a user's interests. As MPSPs and ASPs are potentially able to access such information, they have become a major player in the advertising ecosystem. Often, they even subsume the roles of publisher and broker. For instance, Google offers advertising space in its search engine and owns the mobile advertising company called AdMob. Similarly, Apple acquired the mobile advertising platform iAd.

Over the past years, researchers and journalists have started to analyze apps and mobile operating systems w.r.t. the collection of personal data [17, 23, 12, 1, 37, 39, 7]. The analyses show that sensitive information is accessed and transmitted. The following sections provide an overview on how data is collected and examples about personal information that is or was collected by MPSPs and ASPs.

II.B.1. MPSP – Positioning & Profiling Services

There are typically three reasons for MPSP to collect location information: positioning, mapping, and advertising. These services can also be provided by third parties but in practice they have become so important

for the mobile ecosystem that they are usually linked into the mobile platform itself and offered by the mobile platform service provider.

For instance, mobile platform providers utilize their installation base to create or support new (commercial) services based on crowdsourced data. Even though a mobile phone may not be equipped with GPS, a position may be obtained by approximate location determination based on mobile telephony infrastructure or WiFi. The sensor data is sent to external services and external information sources are used to improve (i.e. speed-up) the determination of the user's current location. Thus, the modeling of the user's context is not conducted solely on the user's device anymore. Just as well as the user's location, data necessary for calculation and displaying requested information are usually not stored on the user's device anymore. It is downloaded to the user's device only on demand. Therefore, the user's whereabouts (as well as the user's preferences) have to be transmitted to the service provider frequently.

By aggregating location information of many users, such information could improve or enable new kinds of services. For instance, Google Mobile Maps makes use of user contributed data (with the user's consent) to determine and visualize the current traffic situation.

In Spring 2011 it was found that Apple's iPhone generates and stores a user's location history, more specifically, data records correlating visible WiFi access-points or mobile telephony cell-ids with the device's GPS location on the user's phone. Moreover, the recorded data-sets are frequently synchronized with the platform provider. Presumably, this data is used by MPSPs to improve database-based, alternative location determination techniques for situations where GNSS or similar techniques are not available or not operational. Thus, re-visited locations are stored on the phone irregularly. Therefore they are not suitable for identifying frequently visited places and providing semantic interpretations to routine trips and activities [2]. Nevertheless, the stored locational information is sufficient for inferring which places have been visited by a phone owner or, in contrast, which places were not attended. Such information can also be harmful to personal privacy when a person was expected to visit some places due to his or her obligations.

II.B.2. ASP – Personalization of Mobile Context

In order to perform dedicated tasks, apps also access other data such as the user's contacts, calendar, and bookmarks as well as sensor readings (e.g. camera, microphone). If these apps have access to the Internet, they are potentially able to disclose this information and are a serious threat to user privacy [23]. Usually, advertisement libraries (e.g., as part of an app) require access to the phone information and location API [17] in order to obtain the phone's IMEI number and geographic position.

For instance, Apple Siri records, stores, and transmits any spoken request to Apple's cloud-based services, where it is processed through speech recognition software, is analyzed to be understood, and is subsequently serviced. The computed result of each request is communicated back to the user. Additionally, to fully support inferencing from context, Siri is "expected to have knowledge of users' contact lists, relationships, messaging accounts, media (songs, playlists, etc) and more" [2], including location data to provide the context of the request, which are communicated to Apple's data center. As an example of Siri's use of location data, users are able to geo-tag familiar locations (such as their home or work) and set a reminder when they visit these locations. Moreover, user location data is used to enable Siri to support requests for finding the nearest place of interest (e.g., restaurant) or to report the local weather.

III. Privacy Threats and Risks

From a business perspective, mobility data with sufficiently precise location estimation are often valuable for enabling various location-based services; from the perspective of privacy advocates, such insights are often deemed a privacy threat or a privacy risk. Location privacy risks can arise if a third-party acquires a data tuple (user ID, location), which proves that an identifiable user has visited a certain location. In most cases, the datum will be a triple that also includes a time field describing when the user was present at this location. Although in theory there are no location privacy risks if the user cannot be identified or if the location cannot be inferred from the data, in practice it is difficult to determine when identification and such inferences are possible.

Recently, several location privacy incidents were reported in the media. A famous incident regards the case of Apple [9], where 3G Apple iOS devices were reported to store the location of their mobile users' in unencrypted form for a period of over one year. This precise location information was stored without

[2] http://privacycast.com/siri-privacy-and-data-collection-retention/, Online, Version of 9/6/2012

the knowledge of the users and was transmitted to the iTunes application during the synchronization of the device. According to Apple, the stored location information was not used to track the users but was attributed to a programming error which was later fixed with a software update.

Google was also reported to be using precise location data, collected from users' mobile devices, to improve the accuracy of its navigation services [21], while Microsoft [31] recently admitted that their camera application in Windows Phone 7 ignored the users' privacy settings to disable transmitting their location information to Microsoft. In response to this incident, the company issued a software update.

Although the above-mentioned privacy incidents did not lead to actual harm caused to the individuals due to the lack of location privacy, the continual flurry of such breaches is worrying as it becomes evident that sensitive location information may easily fall into the wrong hands [10]. In the following subsections, we elaborate on different types of privacy risks leading to user identification or to sensitive location inferences.

III.A. Collection of location information with assigned user ID

This is the most trivial case, as long as the location of the user is estimated with sufficient accuracy for providing the intended LBS. In case where the location is not yet precise enough, various techniques (e.g. fusion of several raw location data from various sensors) allow for improving the accuracy.

- **Example 1.1:** MNO routinely stores tuples of the form (cell ID and sector ID, user ID), e.g. within the CDR data.
- **Example 1.2:** ASP gets the GPS-location for a user who has already been identified, e.g. by his/her log-in to the ASP or by a payment transaction.
- **Example 1.3:** From a smartphone, ASP receives the IDs and signal strengths of several nearby transmitters (base stations, WiFi devices,...). Based on previously established maps of these transmitters, the ASP is able to estimate a more precise location.

Additionally, ASP may have direct access to a variety of publicly available spatial and temporal data such as

- geographical space and inherent properties of different locations and parts of the space (e.g. street vs. park)
- various objects existing or occurring in space and time: static spatial objects (having particular constant positions in space), events (having particular positions in time), and moving objects (changing their spatial positions over time).

Such information either exists in explicit form in public databases like OSM, WikiMapia or in ASP's data centers, or can be extracted from publicly available data by means of event detection or situation similarity assessment [3][4]. Combining such information with positions and identities of users allows deep semantic understanding of their habits, contacts, and lifestyle.

III.B. Collection of anonymous location information

When location data is collected without any obvious user identifiers, privacy risks are reduced and such seemingly anonymous data is usually exempted from privacy regulations. It is, however, often possible to re-identify users based on quasi-identifying data that have been collected. Therefore, the aforementioned risks can apply even to such anonymous data.

The degree of difficulty in re-identifying anonymized data depends on the exact details of the data collection and anonymization scheme as well as on the adversaries, access to background information. Consider the following examples:

Re-identifying individual samples. Individual location records can be re-identified through observation attacks [30]. The adversary knows that user Alice was the only user in location (area) l at time t, perhaps because the adversary has seen the person at this location or because records from another source prove it. If the adversary now finds an anonymous datum (l, t) in the collected mobility data, the adversary can infer that this datum could only have been collected from Alice and has re-identified the individual. In this trivial example, there is actually no privacy risk from this re-identification because the adversary knew a priori that Alice was at location l at time t, so the adversary has not learned anything new.

There are, however, three important variants of this trivial case that can pose privacy risks. First, the anonymous datum may contain a more precise location l' or a more precise time t' than the adversary knew about a priori. In this case, the adversary learns this more precise information. Second, the adversary

may not know that Alice was at l but simply know that Alice is the only user who has access to location l. In this latter case, also referred to as restricted space identification, the adversary would learn when Alice was actually present at this location. Third, the anonymous datum may contain additional fields with potentially sensitive information that the adversary did not know before. Note, however, that such additional information can also make the re-identification task easier.

Re-identifying time-series location data. Re-identification can also become substantially easier when location data is repeatedly collected and time-series location traces are available. We refer to time-series location traces, rather than individual location samples, when it is clear which set of location samples was collected from the same user (even though the identity of the user is not known). For example, the location data may be stored in separate files for each user or a pseudonym may be used to link multiple records to the same user.

Example 2.1: A partner of the MNO has obtained anonymized traces of a user, e.g. as a sequence of CDRs where all user IDs have been removed. While this looks like anonymous location data, various approaches exist to re-identify the user associated with these mobility traces. One approach is to identify the top 2 locations where the user has spent most of its time. This corresponds in many cases to the home and work location of a certain user.

Empirical research has further observed that the pair (home location, work location) is often already identifying a unique user [16]. A recent empirical study [41] explains various approaches for re-identification of a user. Another paper has analyzed the consequences of increasingly strong re-identification methods to privacy law and its interpretation [34].

Further re-identification methods for location data rely on various inference and data mining techniques.

III.C. Collection of data without location

Even in absence of actual location readings provided by positioning devices, location disclosures may occur by means of other modern technologies. Recent work by Han et al. demonstrated that the complete trajectory of a user can be revealed with a 200 m accuracy by using accelerometer readings, even when no initial location information is known [24]. What is even more alarming is that accelerometers, typically installed in modern smartphones, are usually not secured against third-party applications, which can easily obtain such readings without requiring any special privileges. Acceleration information can thus be transmitted to external servers and be used to disclose user location even if all localization mechanisms of the mobile device are disabled.

Another example of privacy disclosures in mobile devices regards the monitoring of user screen taps through the use of accelerometer and gyroscope readings. Recent work by Miluzzo et al. demonstrated that user inputs across the display and the letters of a mobile device can be silently identified with high precision through the use of motion sensors and machine learning analysis [32]. Their prototype implementation achieved tap location identification rates of as high as 90% in accuracy, practically demonstrating that malevolent applications installed in mobile devices may severely compromise the privacy of the users.

Last but not least, several privacy vulnerabilities may be exposed through the various resource types that are typically supported and communicated by modern mobile phone applications. Hornyack, et al. examined several popular Android applications which require both internet access and access to sensitive data, such as location, contacts, camera, microphone, etc. for their operation [23]. Their examination showed that almost 34% of the top 1100 popular Android applications required access to location data, while almost 10% of the applications required access to the user contacts. As can be anticipated, access of third-party applications to such sensitive data sources may lead to both user re-identification as well as sensitive information disclosure attacks, unless privacy enabling technology is in place.

- **Example 3.1:** During a vacation, a user has shot many photos, which are all tagged with a time-stamp but not geotagged. There are, however, techniques to assign to most of these photos a geo-location, as long as these photos contain some unique features. Similarly, there are techniques to assign real names to most persons on these photos, e.g. by using tools or crowdsourcing as provided e.g. by a social network or other platforms to store photos. Having time and places of a photo stream one might reconstruct precise trajectories.

- **Example 3.2:** An app is able to continuously read the accelerometer of a hand-set. Then it can reconstruct a 3D trace of the user's movements.

III.D. Specifics of Episodical Movement Data

Most of the data collected by MNO, MPSP and ASP are referred to as "Episodical Movement Data": data about spatial positions of moving objects where the time intervals between the measurements may be quite large and therefore the intermediate positions cannot be reliably reconstructed by means of interpolation, map matching, or other methods. Mainly three types of uncertainty distinguish episodic from continuous movement data which were identified in [5]. First, the most common type of uncertainty is the lack of information about the spatial positions of the objects between the recorded positions (continuity), which is caused by large time intervals between the recordings and by missed recordings. Second, a frequently occurring type of uncertainty is imprecision of the recorded positions (accuracy). Due to these two types of uncertainty, episodic movement data cannot be treated as continuous trajectories, i.e., unbroken lines in the spatio-temporal continuum such that some point on the line exists for each time moment. Third, the number of recorded objects (coverage) may also be uncertain due to the usage of a service or due to the utilized sensor technology. For example, one individual may carry two or more devices, which will be registered as independent objects. On the other hand, some techniques only capture devices which are turned on. The activation status may change while a device carrier moves.

Figures 2 and 3 emphasizes the differences between continuous and smooth GPS-based trajectories and discrete and abrupt phone-based trajectories. Two images on the left show a map (Figure 2) and so-called space-time cube representation (Figure 3) of a one-day car trajectory in Milan, Italy. Similarly, two images on the right show a map and a space-time cube of a single-day phone-based trajectory.

As discussed above, the information encoded in episodic data is much smaller than in continuous movement data. Many of the existing data analysis and privacy preservation methods designed for dealing with movement data are explicitly or implicitly based on the assumption of continuous objects movement between the measured positions and are therefore not suitable for episodic data. However, due to the increased availability of mobile phone data, analysis methods for episodic movement data and the retrieval of data for unobserved locations are rapidly evolving. On the one hand such techniques pose a privacy risk, on they other hand they would help us understand what sensitive information can be extracted from location traces.

IV. Implications

Potentially sensitive location data from the use of smartphones is now flowing to a largely inscrutable ecosystem of international app and mobile platform providers, often without knowledge of the data subject. This represents a fundamental shift from the traditional mobile phone system, where location data was primarily stored at more tightly regulated cellular carriers that operated within national borders.

A large number of apps customize the presented information or their functionality based on user location. Examples of such apps include local weather information, location-based reminders, maps and navigation, restaurant rating, and friend finders. Such apps often transmit the user location to a server, where it may be stored for a longer duration.

It is particularly noteworthy, however, that mobile advertisers and platform providers have emerged as an additional entity that aggregates massive sets of location records obtained from user interactions with a variety of apps. When apps request location information, the user location can also be disclosed to the mobile platform service provider as part of the wireless positioning service function. Even apps that do not need any location information to function, often reveal the user location to mobile advertisers. The information collected by these advertising and mobile providers is arguably more precise than the call data records stored by cellular carriers, since it is often obtained via WiFi positioning or the GPS. In addition, privacy notices by app providers often neglect to disclose such background data flows [1]. While the diversity of location-based apps has been foreseen by mobile privacy research to some extent—for example, research on spatial cloaking [19] has sought to provide privacy-preserving mechanisms for sharing location data with a large number of apps—this aggregation of data at mobile platform providers was unexpected. In essence, this development goes back to economic reasons. Personal location information has become a tradable good: users provide personal information for targeted advertising in exchange for free services (quite similar to web-based advertising models). The advertising revenue generated from such data finances the operation of the service provider. Because of this implicit bargain between users and service providers, there is little incentive to curb data flows or adopt stronger technical privacy protection as long as it is not demanded by users or regulators.

Figure 2: Comparison of continuous GPS (left) vs. episodic phone-based trajectories (right)

Figure 3: Time-space cube comparison of continuous GPS (left) vs. episodic phone-based trajectories (right)

We suspect, however, that many users are not fully aware of this implicit bargain. Therefore, we believe that it is most important from a privacy perspective to create awareness of these data flows among users, which is not incidentally the very first core principle of the fair information practice principles [40]. It is well understood that lengthy privacy disclosures, if they exist for smartphone apps, are not very effective at reaching the majority of users, and even the recent media attention regarding smartphone privacy [3] does not appear to have found a sufficiently wide audience as our workshop discussions suggest. Raising awareness and empowering users to make informed decisions about their privacy will require novel approaches, user-interfaces, and tools.

When using smartphones, users should not only be aware of *what* data they are revealing to third-parties and of *how frequently* it is revealed; they should also be able to understand the *potential risks of sharing* such data. For instance, users/subscribers in the EU are currently entitled to a full copy of their personal data stored by a commercial entity [4] but such voluminous datasets can currently only be analyzed by experts. Even then, it will be difficult to judge what sensitive information can be learned from this dataset when it is linked with other data about the same person or when it is analyzed by a human expert with powerful visual analysis tools [2]. Is the precision of a loca-

[3] For instance, `http://blogs.wsj.com/wtk-mobile/` Retrieved 2012/10/18.

[4] For example, an Austrian student requested all personal data from Facebook and received a CD [22]

tion record sufficient to determine the building that a user has entered? Is it possible to reconstruct the path a user has taken between two location records? How easily can one infer habits or health of a person based on the location records collected from smartphones?

As another example: some service providers claim to collect location data only in anonymous form. The methods for re-identification, however, have evolved quickly. When can "anonymized" time-series location data really qualify as data that is not personally identifiable information and remain outside of the most current privacy regulations? Finally, even non-georeferenced data provided by the sensors embedded in a smartphone (camera, accelerometer, microphone, etc.) as well as the files stored in the internal memory (photos, music, playlists) allow extracting knowledge about a person's location and mobility. Overall, it appears necessary to investigate what associations can be established and what inferences can be made by a human when the data is considered in context and how such information can be conveyed to users of services.

Users should also be able to learn in which countries their data is stored or processed, since this can have important implications for the applicable legal privacy framework. While the European Union has achieved some degree of harmonization of privacy standards for exported data from its citizens through the safe harbor provisions [13], differences still exist, for example, with respect to law enforcement access to user data. We believe that providing transparency of cross-border data flows would lead to a more meaningful public discussion of data protection policies. For example, when data is handled by multi-national corporations, should data subjects be given a choice of where their data is processed and stored?

We hope that the research community will help address these questions and will, in collaboration with data protection authorities and policy experts, actively define privacy for this mobility data ecosystem.

References

[1] Mobile apps for kids: Current privacy disclosures are dis*app*ointing. Technical report, Federal Trade Commission, 2012. http://www.ftc.gov/os/2012/02/120216mobile_apps_kids.pdf.

[2] G. Andrienko and N. Andrienko. Privacy issues in geospatial visual analytics. In Georg Gartner, Felix Ortag, William Cartwright, Georg Gartner, Liqiu Meng, and Michael P. Peterson, editors, *Advances in Location-Based Services*, Lecture Notes in Geoinformation and Cartography, pages 239–246. Springer Berlin Heidelberg, 2012.

[3] G. L. Andrienko, N. V. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In *IEEE VAST*, pages 161–170. IEEE, 2011.

[4] G. L. Andrienko, N. V. Andrienko, M. Mladenov, M. Mock, and C. Pölitz. Identifying place histories from activity traces with an eye to parameter impact. *IEEE Trans. Vis. Comput. Graph.*, 18(5):675–688, 2012.

[5] N. Andrienko, G. Andrienko, H. Stange, T. Liebig, and D. Hecker. Visual analytics for understanding spatial situations from episodic movement data. *KI - Künstliche Intelligenz*, pages 241–251, 2012.

[6] M. Anisetti, C. A. Ardagna, V. Bellandi, E. Damiani, and S. Reale. Map-based location and tracking in multipath outdoor mobile networks. *Wireless Communications, IEEE Transactions on*, 10(3):814 –824, march 2011.

[7] C. Arthur. iphone keeps record of everywhere you go. *The Guardian*, 2011. http://www.guardian.co.uk/technology/2011/apr/20/iphone-tracking-prompts-privacy-fears.

[8] L. Bengtsson, X. Lu, A. Thorson, R. Garfield, and J. von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in haiti. *PLoS Med*, 8(8):e1001083, 08 2011.

[9] N. Bilton. 3G Apple iOS devices are storing users location data. The New York Times, Published: April 20, 2011, 2011.

[10] N. Bilton. Holding companies accountable for privacy breaches. The New York Times, Published: April 27, 2011, 2011.

[11] Christopher W. Clifton, Bart Kuijpers, Katharina Morik, and Yucel Saygin. Mobility Data Mining and Privacy (Dagstuhl Seminar 12331). *Dagstuhl Reports*, 2(8):16–53, 2012.

[12] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. In *Proceedings of the 9th USENIX conference on Operating systems design and implementation*, OSDI'10, pages 1–6, Berkeley, CA, USA, 2010. USENIX Association.

[13] European Parlament and European Council. Directive 95/46/ec on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities*, (L281), 1995.

[14] Council European Parliament. Directive 2006/24/ec of the european parliament and of the council of 15 march 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending directive 2002/58/ec. *Official Journal of the European Union*, L 105:54 – 63, 2006.

[15] Gartner, Inc. Gartner says worldwide mobile advertising revenue forecast to reach $3.3 billion in 2011, 2011. http://www.gartner.com/it/page.jsp?id=1726614.

[16] Philippe Golle and Kurt Partridge. On the anonymity of home/work location pairs. In Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. Brush, and Yoshito Tobe, editors, *Pervasive Computing*, volume 5538 of *Lecture Notes in Computer Science*, pages 390–397. Springer Berlin / Heidelberg, 2009.

[17] M. C. Grace, W. Zhou, X. Jiang, and A.-R. Sadeghi. Unsafe exposure analysis of mobile in-app advertisements. In *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks*, WISEC '12, pages 101–112, New York, NY, USA, 2012. ACM.

[18] A. Greenberg. Phone 'rootkit' maker carrier iq may have violated wiretap law in millions of cases. *Forbes*, 2011. http://www.forbes.com/sites/andygreenberg/2011/11/30/phone-rootkit-carrier-iq-may-have-violated-wiretap-law-in-millions-of-cases/. Retrieved 2012/10/18.

[19] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services*, MobiSys '03, pages 31–42, New York, NY, USA, 2003. ACM.

[20] H. Haddadi, P. Hui, T. Henderson, and I. Brown. Targeted advertising on the handset: Privacy and security challenges. In Hans Jrg Mller, Florian Alt, and Daniel Michelis, editors, *Pervasive Advertising*, Human-Computer Interaction Series, pages 119–137. Springer, 2011.

[21] M. Helft. Apple and Google use phone data to map the world. The New York Times, Published: April 25, 2011, 2011.

[22] K. Hill. *Forbes*, 2012. http://www.forbes.com/sites/kashmirhill/2012/02/07/the-austrian-thorn-in-facebooks-side/. Retrieved 2012/10/18.

[23] P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS '11, pages 639–652, New York, NY, USA, 2011. ACM.

[24] H. Jun, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang. Accomplice: Location inference using accelerometers on smartphones. In *Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on*, pages 1 –9, 2012.

[25] D. Kravets. An intentional mistake: The anatomy of googles wi-fi sniffing debacle. *Wired*, 2011. http://www.wired.com/threatlevel/2012/05/google-wifi-fcc-investigation/. Retrieved 2012/10/18.

[26] J. Krumm. Ubiquitous advertising: The killer application for the 21st century. *Pervasive Computing, IEEE*, 10(1):66 –73, jan.-march 2011.

[27] S. Landtag. Druchsache 5/6787. Sächsischer Landtag 5. Wahlperiode, 2011.

[28] E. Lichtblau. Police are using phone tracking as a routine tool. The New York Times, Published: March 31, 2012, 2012.

[29] C. Ling, M. Loschonsky, and L. M. Reindl. Characterization of delay spread for mobile radio communications under collapsed buildings. In *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 329–334, Sept. 2010.

[30] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao. Privacy vulnerability of published anonymous mobility traces. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, MobiCom '10, pages 185–196, New York, NY, USA, 2010. ACM.

[31] D. McCullagh. Microsoft collects locations of Windows phone users. CNet News, Published: April 25, 2011, 2011.

[32] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury. Tapprints: your finger taps have fingerprints. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, MobiSys '12, pages 323–336, New York, NY, USA, 2012. ACM.

[33] Ministerium für Inneres und Kommunales NRW. Funkzellenauswertung (FZA) und Versenden "Stiller SMS" zur Kriminalitätsbekämpfung. MMD 15/3300, 2011.

[34] P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, Vol. 57, p. 1701, 2010*, 2009.

[35] ZEIT Online. Landgericht erklärt funkzellenabfrage auf demo für rechtswidrig. *Die Zeit*, 2013. `http://www.zeit.de/digital/datenschutz/2013-04/funkzellenabfrage-dresden-landgericht`.

[36] K. Rechert, K. Meier, B. Greschbach, D. Wehrle, and D. von Suchodoletz. Assessing location privacy in mobile communication networks. In H. Li X. Lai, J. Zhou, editor, *ISC 11*, LNCS 7001, pages 309–324. Springer, Heidelberg, 2011.

[37] E. Smith. iphone applications & privacy issues: An analysis of application transmission of iphone unique device identifiers (udids). Technical report, PSKL, 2010. `http://www.pskl.us/wp/wp-content/uploads/2010/09/iPhone-Applications-Privacy-Issues.pdf`.

[38] G. Sun, J. Chen, W. Guo, and K. J. R. Liu. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs. *Signal Processing Magazine, IEEE*, 22(4):12 – 23, July 2005.

[39] S. Thurm and I. Yukari Kane. Your apps are watching you. *The Wall Street Journal*, 2010. `http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html`.

[40] W. H. Ware. Records, computers, and the rights of citizens. Secretary's Advisory Committee on Automated Personal Data Systems, Department of Health, Education and Welfare, Washington, D.C., 1973.

[41] H. Zang and J. Bolot. Anonymization of location data does not work: a large-scale measurement study. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, MobiCom '11, pages 145–156, New York, NY, USA, 2011. ACM.

[42] D. Zimmermann, J. Baumann, A. Layh, F. Landstorfer, R. Hoppe, and G. Wolfle. Database correlation for positioning of mobile terminals in cellular networks using wave propagation models. In *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, volume 7, pages 4682 – 4686 Vol. 7, 2004.