Towards Privacy-preserving Mobile Location Analytics

Marius Gassen
Technische Universität Darmstadt
Hochschulstraße 10, 64289 Darmstadt, Germany
marius.joerg.gassen@googlemail.com

Hervais Simo Fhom
Fraunhofer Institute for Secure Information Technology
Rheinstrasse 75, 64295 Darmstadt, Germany
hervais.simo@sit.fraunhofer.de

ABSTRACT

Mobile Location Analytics (MLA) is enjoying increased attention. Typical businesses eager to exploit the opportunities offered by this emerging form of location-based services are venues of various types and size including retail stores, shopping malls, airports, hotels, and theme parks. MLA relies on applying statistical inference methods to sensory data constantly generated by mobile devices of (potential) customers/visitors or data collected by a variety of in-door sensors in order to generate useful insights into people's behavior and interests. While providing venue operators and (potential) customers with many benefits, MLA also raises significant privacy concerns, given the sensitive nature of the data being collected and transferred to remote entities for further processing. In this paper, we offer a vision for building privacy and data protection into MLA. We argue for a holistic and user-centered approach, i.e., one enabling individuals whose data are collected and processed by MLA services to be aware of and understand the associated data flows, the resulting privacy risks, and appropriated options to restrict the access to and (downstream) usage of their data. The building blocks of our approach are highlighted. Our proposal rests on a comprehensive set of privacy and data protection requirements which in turn is a result of a thorough analysis of attack surfaces available in the context of MLA, the associated threat model and the privacy risks they might entail. The compiled set of privacy and data protection requirements is tailored to the specific needs of embedded systems which are key enablers of MLA.

Keywords

Mobile Location Analytics, Privacy, AlterEgo

1. INTRODUCTION

With the belief that emerging wireless and mobile networks in combination with the increasing connectivity of mobile handheld devices will provide new business opportunities, a growing industry is promoting a new form of location-based services - Mobile Location Analytics (MLA). A currently prominent example of MLA service is Retail Tracking (a.k.a. In-store Tracking)[1]. Although there is still no commonly agreed upon definition of MLA, the term is often used to describe a set of technological solutions that integrate retail infrastructures with sensor technologies aiming to continuously track and analyze individuals' activities in or near a venue of interest. MLA services make use of various types of medium or short range sensors embedded in physical objects and/or in the venue itself to collect a variety of data about users i.e., visitors and passersby. This is typically done by capturing wireless signals emitted by their mobile devices, recording their interactions with products using the in-store cameras, and relying on the fact that smartphone users hardly ever let go of their devices [2]. The data collected is typically transferred in real-time to an MLA backend for further processing. The backend is either on-premise or a cloud-based. Relying on smart algorithms, components in the MLA backend can translate wireless signals and visitors' movement and activity data into actionable insights such as who the customers are, how much time they spend in specific areas in the venue and how often they return, what their interests are, what their current geo-location is, and more (cf. [3]). Note that, depending on the technologies the retailer relies on, e.g., GPS, WiFi or Bluetooth, current geo-location may refer to the customer's indoor location, a location a block away from the venue or a location anywhere else on the planet. Based on such insights, brick and mortar stores for instance can, among other things, better understand their customers' shopping patterns, which in turn could enable them to improve customer engagement, optimize product placement strategies [4][5], improve marketing strategies and run targeted campaigns towards user groups sharing similar profiles. Such profiles can be built based on estimations of visitors' interests, location and demographics, which includes ethnicity, age, marital status, size of family, income, education and employment [6]. We use the generic term store to refer to commercial sites and venues of various type and size, including brick and mortar stores, shopping malls, airports, theme parks, and other.

Against this background, we argue that indoor tracking approaches in MLA can be roughly classified into the three general categories: 1) *venue-initiated tracking*, where radio-frequency detection equipments and electronic sensors in the physical space of interest continuously intercept signals emitted by shoppers' smartphones and in-store cameras record the whereabouts, activities, and possibly biometrics of all individuals inside the store [7]; 2) *individual-initiated track-*

ing, where shoppers, through a "loyalty card" app running on their smartphones, are incentivized to interact with the store both offline (e.g., with tagged products and smart shelves) and online (receiving personalized notifications and contextual offers while shopping in competitors' locations); and 3) *hybrid tracking*, an approach combining the features of individual-initiated and those of venue-initiated tracking approaches (cf. [8]). Here, the tracking is performed relying on sophisticated IT-equipment (possibly a combination of optical camera-based systems and other types of sensors) within stores and specialized software on users' smartphones, and assuming the ability of the visitor and the entity operating the tracking infrastructure to cooperate with each other. Depending on the tracking approach, different data types are self-reported by the users themselves or collected, leaked, derived and processed by MLA systems. Indeed, from the signals emitted by the user smartphone, persistent/temporary identifiers, e.g., WiFi MAC addresses, can be determined and the location of an individual device inferred accurately [9][10]. When deployed, video cameras provide real-time footage from inside and outside the store for processing and analysis. If not exclusively built to create heatmaps and thermal imaging, footage from cameras embedded in the physical environments might capture individual's biometric. Relying on a branded app, additional data voluntarily disclosed by the users themselves when registering for loyalty programs or using their loyalty cards for purchases, including credit or debit card details, email addresses, date of birth, name, and home address, are collected and transferred to the MLA cloud [3]. Before further processing, this data is anonymized. However, several research efforts have recently cast doubt on the effectiveness of the anonymization techniques currently being employed to protect some of this information, e.g., MAC addresses [11].

In the light of the above, the growing popularity of MLA raises new and significant security and privacy challenges (cf. [12]), as seamless, continuous, and non-transparent tracking of users' offline activities and movements might i) be easily abused to place a large portion of the population under surveillance; and ii) prevent individuals from exercising their right to informational self-determination and to protect their privacy.

Our contribution: In this paper, we aim to i) provide a comprehensive analysis of the main privacy and data protection issues in MLA systems; and ii) discuss the foundation of a novel technological approach to solve them. Acknowledging that in-store tracking is enabled by a composition of and interplay between different wireless networks of sensors and actuators and remote data analytics platforms, we argue for a holistic user-centered approach to privacy protection in MLA, i.e., an approach that aims to address security and privacy risks at all subsystems of the MLA infrastructure. We discuss key requirements for designing such a novel approach towards user-centered privacy management framework. More importantly, we present the main features of the envisioned privacy solution and identify the main research challenges, especially focusing on privacy and usability.

The rest of the paper is organized as follows. In section II, potential security issues in and privacy implications of MLA are analyzed. In section III, the corresponding security and privacy requirements are presented and our proposal for a holistic user-centered approach to privacy-preserving MLA is discussed. Section IV concludes the paper.

2. PRIVACY THREATS AND IMPLICATIONS

Given the aforedescribed system model, this section presents the associated threat model (i.e., the associated attack surfaces and types of adversaries) and discusses key consequential privacy implications.

2.1 Attack surfaces and vulnerabilities

As stores and venues are increasingly evolving into sensor-enriched and networked environments able to bridge the real world and the virtual world of networking and computing, they also expose exploitable vulnerabilities and attack surfaces, hence providing the opportunity for new privacy attacks. Attack surfaces present in MLA systems can be roughly classified into three categories, depending on whether they affect or are related to the medium-/short-range wireless interfaces [13][14], to the software (e.g., embedded operating systems/firmware and applications) [15][16], or the hardware in relevant subsystems [17][18].

Attack surfaces related to wireless interfaces include Bluetooth, NFC, RFIDs, WiFi, and other dedicated short-range communication protocols. Adversaries with technical skills to reverse-engineer and/or exploit flaws in wireless communication protocols can passively eavesdrop on the communication between visitors' smartphones and the store infrastructure or drop, delay, replay and forge data packets [13][14].

Software attack surfaces typically include exploitable vulnerabilities in both in-store sensors' firmware and backend applications, and weak OS-level access control mechanisms in smartphones [15]. Relying on software vulnerabilities, adversaries can inject malicious code into a poorly secured device, take control of it, and use it to illegitimately obtain access to visitors' data stored in the backend system.

At a lower level, *vulnerabilities can be found and exploited in chip sets and other types of hardware.* Such hardware vulnerabilities may be shared among smartphones, in-store sensors/tracking equipment, and backend servers. Focusing on hardware attack surfaces, an adversary can hack into the sensing equipment [17] and/or plant a backdoor designed as hardware feature in the MLA infrastructure [18].

2.2 Attacker/Adversarial Model

In order to understand the risks associated with the collection, processing and sharing of user-related data more accurately, an adequate characterization of the attacker model is necessary. Such a specification should be based on adversaries' goals, capabilities, and relationship to (parts of) MLA infrastructure at hand. Borrowing from the standard taxonomy of attacker model in computer security literature, we can distinguish adversaries in MLA contexts along the following main classes: passive adversaries, active adversaries, internal adversaries, and external adversaries.

Passive adversaries are interested in eavesdropping on one or multiple in-store wireless communication channels, exploiting the fact that wireless signals broadcasted by the user's device, prior to an established connection with the wireless gateway, is typically not encrypted. Examples of passive adversaries include i) curious stores and hackers who can set-up sniffing equipment inside or near a physical space of interest and secretly capture as many visitor-related wireless signals as possible, and ii) omnipresent listeners, legitimate or not, such as the MLA provider[1] and mobile network

[1]A term we use interchangeably in this paper with MLA

operators increasingly eager to leverage data about their subscribers to deliver MLA services [19]. As an omnipresent adversary, an MLA provider can gather information from a multitude of signals. As such, omnipresent adversaries and the MLA provider in particular can correlate identifiers that are typically observable in single separated signals, e.g., MAC address and Credit Card Metadata [20]. Recall that such correlation may also consider identifiers captured from different physical locations, enabling omnipresent adversaries to continuously track the whereabouts of individuals across stores and over large areas.

Active adversaries go beyond eavesdropping and are also able to capture, modify or inject data packages over wireless communication channels. Indeed, in contrast to passive adversaries who rely solely on their ability to listen to, collect and analyze broadcasted signals, active adversaries can exploit vulnerabilities in communication protocols [13][14], in-store cameras [17], and the (cloud-based) analytics backend [16]. As a result, they can successfully inject arbitrary privacy-invasive software into both customers' smartphones and the MLA backend. Relying on such software, active adversaries can secretly extract a variety of information about the user from the infected IT components and upload it to a third-party server. The adversary can then combine and analyze all data in his possession, legitimately obtained or not, to infer additional information about a targeted user. For instance, an active adversary can rely on a privacy-invasive app to monitor information flows in smartphone's operating systems and infer sensitive details about a user's identity, health condition, and habits without that user's permission [21]. Moreover, the adversary may infer details about the user's social ties by correlating information contained in WiFi probe requests [9].

Internal adversaries refer to any entity in the MLA ecosystem that i) aim to obtain, store and/or processed identity details or any personally identifiable information without an explicit consent by the individual, or ii) might misuse or sell legally obtained/recorded visitors' data. Examples of internal adversaries include rogue store's employees (resp. MLA provider's employees) able to illegally deploy sensing equipments or tamper with critical hardware and data.

External adversaries are entities with no legitimate system roles. They may be either passive or active. Possible external adversaries include intrusive law enforcement agencies and any other illegitimate third party interested in visitors' data, e.g. data brokers, insurance companies, etc.

2.3 Privacy issues and implications

If improperly designed, MLA would raise a range of potential privacy risks and threats. Arguing that MLA is a use case resulting from the convergence between context-aware computing, the Internet of Things and cloud computing-based big data analytics, we focus on 5 possible privacy implications of MLA which we detailed as follows:

1. Lack of Transparency/ Opaqueness of data collection. Being surrounded by sensors embedded in their physical environment and capable of recognizing and responding to people's presence in a seamless and often invisible way, users may find themselves in situations in which they are not aware of such collection, do not know which information about them is collected, who (including adversaries as discussed above) is collecting their data, how it is being used, or with whom it may be shared down the road. In addition, details about both the MLA infrastructures, e.g., the exact location and capabilities of the sensors, and possible consequences (e.g. profiling) associated with the handling of their data are often opaque. Indeed, it remains unclear whether visitors are fully aware of the fact that products and other kinds of objects in their physical environment are tagged with sensor technologies able to continuously observe their movements and interactions and upload that data to an analytics backend able to infer actionable insights about habits and preferences. Furthermore, users might lack a clear understanding of the ways these insights are actually used, e.g., through opaque automated decision processes, to influence their (purchasing) choices and decisions. Such a lack of transparency may undermine the ability of the user to effectively anticipate privacy risks associated with the collection and processing of his or her data, and subsequently take adequate countermeasures.

2. Loss of Individual Control over Personal Data. Regardless of whether the data handling in the MLA ecosystem is made transparent to the user or not, complex data flows resulting from the interactions between visitors' mobile devices, networked sensors integrated into the store environment, and the backend analytics platform on the one hand, and from the MLA's ability to recognize users and anticipate their needs on the other hand, are hardly manageable by individuals, especially when relying on mechanisms provided by today's smartphones. Indeed, as the passive non-transparent data collection in the context of MLA is becoming more ubiquitous, so do the risks stemming from undesirable exposure, unintended disclosure and oversharing.

3. Ineffective Consent Management. Current approaches for requesting and managing consent remain questionable practices - individuals tend to pay limited attention to privacy notices, recall recent controversies around signposting at *Nordstrom* [3]. Further doubts about the practicability of existing consent management mechanisms stem from the fact that the continuous collection of data about potential store visitors is primarily performed in invisible ways and by sensors with limited/without human-machine interfaces, depriving users of their ability to explicitly specify and if necessary withdraw consent decisions.

4. Intrusive Profiling and Unwanted Inferences. Various adversaries, incl. MLA services, collect and analyze data about when customers come and go, where they have been before, monitor every move they make inside the store, what products they look at and how they react. However, applying smart algorithms on the collected data or performing data matching across repositories might also reveal details about shopping habits, preferences and interests for products and places. This new, inferred and clearly sensitive information is often used to create comprehensive customer profiles which are then used to track users' behavior over time, even when they are outside the store. Such profiles can be enriched with identity attributes and other sensitive information easily obtainable from online social network sites. The resulting aggregated profile can reveal a lot more about a customer's private life and intimacy, including details about his or her psychology, lifestyle, finances, and health [9][22]. Note that the profiling and tracking can also be performed by (un-)authorized third party entities, e.g. business partners of the MLA provider, possibly with-

operator and MLA company.

out knowledge or consent of the user (cf. [23]). Moreover, while allowing store operators to align customer preferences for specific brands with their offers, the correlation of seemingly non-sensitive data and use of increasingly smart algorithms also facilitates the surveillance of customers, creating high risks of unfair/unlawful discrimination (cf. [24]) and consequences such as chilling effects (cf. [25]).

5. Security of customer Data. Sensitive data collected and processed in MLA is not limited to sensor data but may also include personally identifiable information. Because this data is typically stored in databases accessible through the Internet, or increasingly shared with third parties, e.g. business partners, the risk of data breaches increases [16].

3. TOWARDS PRIVACY-PRESERVING MLA

3.1 Privacy and Security Requirements

To address the challenge of improving customer privacy and trust while preserving the full benefits of MLA to both individuals and businesses, a coalition of privacy groups and technology companies propose a set of guidelines for privacy and security protection (cf. [26]). The set of proposed guidelines (thereafter referred to as FPF Code of Conduct) prescribe limitations of how the data collected by MLA companies can be used or shared, and how long it may be retained. They explicitly aim at restricting discriminatory uses of data, for instance, by prohibiting the use of collected data for employment, health care or insurance purposes. In addition, the FPF Code of Conduct mandates de-identification and opt-in consent when personal information is collected. Further recommendations include a call for the display of noticeable signage by retailers, a clearly specified and understandable privacy policy, and providing customers with a mean of opting-out. Although not explicitly referring to MAC addresses as personal information, the code of conduct views it as sensitive data that needs adequate protection. The result is a self-regulatory framework focusing primarily on MLA companies operating in the USA. Nevertheless, the proposed principles have strong similarities with well established privacy principles such as those of the OECD[27] and the Fair Information Practices Principles [28] which are common to most privacy laws, regulations and standards around the world, e.g. the U.S. federal privacy laws and the European Data Protection Directives EC/95/46, 2004/52/EC and 2009/136/EC. Although focusing solely on WiFi-based in-store tracking and primarily addressing to US-based companies, the FPF Code of Conduct can be enhanced with privacy properties increasingly embraced by privacy officials in Europe [29] and long promoted in the privacy research community [30]. Our aim here is to propose a comprehensive set of requirements essential to achieve privacy and data protection in various types of in-store tracking. Indeed, based on the FPF Code of Conduct and given the aforedescribed attack surfaces and adversarial model, we derive the necessary requirements for privacy-preserving MLA as follows:

R.1 Improved awareness & transparency of data practices. Awareness refers to the users i) knowing and understanding that networked sensors and other mobile and stationary optical devices embedded in the retail environments continuously collect and process data about them while they are performing routine activities; ii) being aware of the resulting privacy implications, and; iii) comprehending the availability and limits of control options. On the other hand, transparency aims at making all privacy-relevant data processing clear and understandable to users. That is, the user must be informed when in-store sensing instruments and/or a dedicated app on his or her smartphone collect data about the device or him- or herself. The MLA provider should inform users about when and how data is gathered, what kind of data is gathered, what is happening to this data and whether data might be shared with third parties. The overall data practice and precise purposes for which sensitive data is collected and processed must be clearly expressed. Recall that as of today, MLA companies often communicate their data practices to users through signposting or app's privacy policy [3]. However, here is clear evidence that users do not always understand and tend to ignore both forms of notices [31]. Accordingly, privacy notices should be clearer, shorter, and understandable by both humans and machines. Moreover, transparency about inference algorithms run by the MLA provider is required. Indeed, MLA users should be able to understand how profiles about them are constructed and used to provide personalized services.

R.2 Unlinkability and data anonymization by design. In order to reduce the risk of (re-)identification of people visiting the store, unlinkability is required. That is, the ability of data processors or any third party to link sensitive data about visitors without their knowledge or legal basis should be restricted by technical means. Enabling unlinkability would require providing individuals with the ability to rely on pseudonyms when interacting with stores and brands through the in-store tracking infrastructure. We refer to this form of unlinkability as identifier unlinkability. In this context, the creation of pseudonyms should take place on user devices and be a non-reversible process, and its use ensures that independent signals originating from the same user are unlinkable across stores or at least across MLA domains. This should allow the user to hide his or her real (device) identifier, while enabling tracking through the aliases and thus helping to address the current controversy regarding cryptographic hash functions to de-identify MAC addresses. Given the risk of re-identification arising from correlations between anonymous data and external, easily accessible datasets (cf. [32]), the requirement for unlinkability should be extended to include process unlinkability. The latter prescribes that the processing of privacy-relevant data for MLA purposes should be designed in such a way that for an adversary two or more events observed during process execution either remain as much unrelated to each other as they did before the process execution, or cannot be combined for purposes other than those communicated at the collection time.

R.3 User control. The concept of Control aims at empowering the user to supervise and actively influence the collection, flow, and use of personal data. In particular it implies the user's ability to object to the collection and processing of personal data, to determine what piece of data about him or her is about to be disclosed, to whom, for what purposes, and under which conditions. It aims to provide the user with the ability to check the correctness of and to update or delete data about him or her, if necessary or desired. A key instrument for user control is consent management which calls for in-store tracking infrastructures that allow visitors to express informed, specific, and unambiguous consent decisions related to the automated collection

and processing of sensitive data by networked sensors. In addition, MLA infrastructures should enable visitors to exercise their right to withdraw previously given consent. Consent management should support "Opt-In" as part of a set of default privacy control options. That is, data collection and processing should not occur unless the visitor explicitly allows it or requests it. We argue, contrary to the recent MLA code of conduct proposals [26] but in the spirit of EU law, that opt-out is not an adequate approach to obtain customers' informed consent. However, coarse-grained control options such as consent management, often reflect the "take it or leave it" dualism, even when designed to be simpler to understand and rely on. Accordingly, fine-grained control options, e.g., selective disclosure of device sensor data, are required to complement aforementioned coarse-grained control options. Fine-grained control options aim at empowering individuals to be able to specify fine-grained preferences that should govern how in-store sensors collect and upload visitors' data, or how privacy sensitive data should be handled down the road, while not jeopardizing the service. Users' fine-grained preferences need to be enforceable in various parts of the MLA ecosystem. However, users' preferences may conflict with terms of the MLA policy, e.g., it may be the case that the user wants hier or her profile not to be shared at all, not even in an anonymized form, while the policy on the MLA provider's side allows sharing customer profiles with partners. To avoid conflicting privacy rules, terms enforceable in the ecosystem should be dynamically generated from the visitor's preferences and the MLA provider's policy, requiring visitor and MLA provider to negotiate appropriated privacy rules in an automated manner. Moreover, users should be provided with a reliable way to verify the privacy claims by the operators. Indeed, reliable and verifiable enforcement of individual's preferences in various subsystems of the MLA infrastructure is crucial when aiming to increase public acceptance of MLA. Furthermore, ensuring high expressiveness and confidentiality of user's preferences is a challenging endeavor in itself [33].

R.4 Data, process and infrastructure security. Operators of in-store tracking infrastructures should put adequate technical and organizational measures in place in order to prevent unauthorized access, alteration or loss of sensitive customer data (profiles, real identities,$\cdots$). This implies deploying measures to protect and restrict access to sensitive data across its lifecycle. It includes preventing data leakage i) by branded apps; ii) during transmission to the cloud-based backend; iii) when data is stored and processed "in the cloud". In all cases, flow of sensitive data to unauthorized parties should be prevented. Moreover, sensitive information resulting from advanced data processing and statistical analysis is required to be well protected from both the cloud operator and other third parties.

R.5 Performance requirements. Implementing enhanced transparency and privacy control, especially at sensor level, for instance, in the form of hardware accelerators for cryptographic algorithms, may decrease both the performance of the overall system and MLA's utility. Accordingly, privacy solutions for MLA need to be lightweight enough in order to be supported by resource-constrained devices like iBeacons or cameras. In addition, the communication, storage, and computational overhead that might result from embedding privacy in such devices should be kept to the minimum. Specifically, new challenges might emerge when designing privacy mechanisms that aim at covering the heterogeneity of (wireless) communication technologies used in MLA. This includes constraints related to standardized maximum network bandwidth, protocol and frame format.

R.6 Usability. When building privacy and transparency enhancement into MLA infrastructures and services, system designers will be ill-advised not to include usability considerations in the specification of the new systems. More precisely, technological options to fulfill the aforedescribed requirements should be designed and implemented in a way that by bringing key fact to the user's attention, he or she would become aware of both cognitive biases and possible consequences of seamless, continuous data collection, while avoiding overburdening him or her. Usability requirements along with privacy are especially challenging aspects of designing MLA services and should therefore be taken into account from the earliest steps in their conception.

3.2 User-centered Privacy Protection

Here we illustrate how the aforementioned vulnerabilities, threats, and attacks can be mitigated and privacy requirements for in-store tracking can be realized. Our proposal is a trusted ubiquitous service we refer to as *"AlterEgo"*. The approach aims to empower users' eager to effectively manage their privacy and regain appropriate control over their data as they navigate through sensor enriched store environments. More specifically, the AlterEgo as an smart agent acting on behalf of the data subject, aims to allow him or her to be aware of sensors embedded in range, to intuitively and effectively control how these sensors collect and process data about the device or the user, and restrict the flow to and further processing in the MLA backend. Leveraging scalable data mining methods and access to a user's data, the AlterEgo can determine privacy risks associated with the collection and processing of sensitive data. Through user-friendly visualizations and adequate interactions, the AlterEgo aims not only to help the user understand the ways in which his or her data is collected and processed, but also the privacy ramifications thereof. In order to provide these features to the user, we argue for a design and implementation of the AlterEgo as a mobile app that relies on both the smartphone's system-level services provided by a component we refer to as *AlterEgo Middleware* and the privacy APIs provided by remote services. Figure 1 provides an overview of the functional architecture of the resulting AlterEgo framework.

AlterEgo app. In the context of this paper, we consider two possible modi for running the AlterEgo app: a passive modus and an active modus. We imagine the app in the passive modus as being a background process continuously listening to and interpreting specific signals emitted by intelligent objects in the user's physical surroundings. Such signals might contain advertisements about the presence of sensing technologies and notifications about the data practices of their respective operators. Upon detecting signals from the surrounding environment and based on details about the mobile device context, the AlterEgo app decides whether to notify the user, for instance, through vibration alerts and/or displays notifications. Reacting to the notification the user can then decide whether the app should adapt the privacy profile of his or her mobile device. For instance, the app may help switch from a *low-profile* in which all sensor-based interactions with the surrounding world are

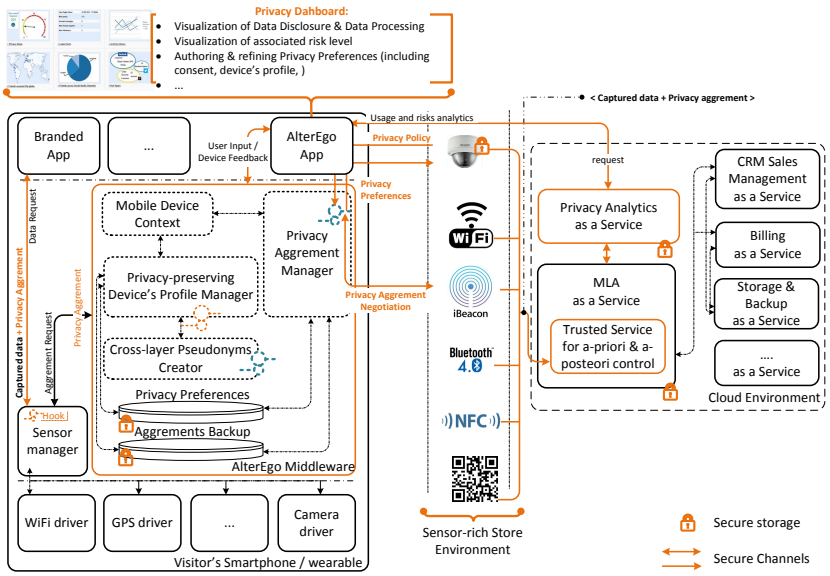

Figure 1: An Architectural Overview of the AlterEgo Framework

completely blocked to a *high-profile* in which information about the user and device is selectively disclosed. Data disclosure, resp. collection of data is performed in accordance to the terms of an on-the-fly negotiated privacy agreement between each user and the MLA provider.

Privacy Agreement Negotiation. To support privacy negotiation, MLA operators might rely on technological approaches that leverage wireless and non-wireless signaling mechanisms detectable by user mobile devices. Relying on similar mechanisms, the user device can signal her privacy preferences to MLA operators. Our proposal builds upon and extends various approaches towards preference and policy signaling from the literature including privacy beacons [34][35] and Roesner's concept of world-driven access control that relies on QR codes, ultrasound and audio [36]. In our design, the device, once switched in the *high-profile*, starts to negotiate a privacy agreement with operators of MLA sensing technologies detected in the user's vicinity. The privacy policy communicated by the MLA operator and captured by the AlterEgo app is forwarded to the Privacy Agreement Manager (PAM), a trusted module in the AlterEgo Middleware. The PAM creates the privacy agreement, i.e., a machine-readable set of overlapping rules (about allowed usage and obligations) and non-binding options resulting from the matching between user preferences and the operator policy. The terms of the created privacy agreement are then communicated to the operator and a copy is stored locally. The local copy can later be reused in three possible ways, provided the device context and the operator's privacy practice remain unchanged. First, the terms of the privacy agreement might be enforced directly on the user device before any sensitive data is transferred to the MLA operator, resulting, for instance, in a cross-layer unlinkable pseudonym being created and broadcasted instead of revealing the WiFi/Bluetooth MAC-address (satisfying R.2). Second, the privacy agreement might be used as a *sticky policy* attached to each single piece of sensitive information leaving the user's device towards the operator controlled infrastructure. In the third scenario, the agreement is attached to obfuscated but still re-identifiable data, before it is transferred to the infrastructure controlled by the MLA provider. Accordingly, objects in sensor-enriched store environments, especially wireless sensor readers and in-store cameras, need to be able to evaluate the privacy agreement (satisfying R.4), and if needed enforce its terms on the sensor itself, before uploading the data they are capturing to the cloud-based MLA backend. Related to in-store cameras for instance, this implies that a privacy engine implemented as a software or hardware extension either performing pre-filtering, watermarking or encryption on raw images; or attaching the agreement to sensitive raw image prior to its transfer to the MLA cloud. The realization of the privacy engine could leverage "Trusted Execution Environments", such as ARM TrustZone [37] or hardware-based solutions specifically designed to meet strict real-time requirements such as SMART [38]. Both approaches provide means for the isolated execution of applications and secure storage of credentials on semi-trusted embedded platforms. By aiming at a practical implementation of the *sticky policy paradigm* [39], our proposal puts the user in the driver's seat, providing him or her with improved control abilities over the complete life cycle of sensitive data in particular as it travels across multiple domains (satisfying R.3). Nevertheless, a challenging issue remains yet to solve: how to enable users to specify and communicate fine-grained privacy preferences to sensors embedded in physical environments given limitations of wireless and non-wireless signaling mechanisms mentioned in the literature? For instance, could we leverage standard header fields in IEEE 802.1X management frames to encode and communicate users' preferences regarding off-/online tracking and data handling in a way that truly enables fine-grained preference specification, i.e., that goes beyond binary decision to consent to or disapprove a particular data practice?

Supporting ex-ante and ex-post transparency. As

mentioned, effective individual control over sensitive data requires an improvement of user awareness about data disclosure, improved transparency over data processing, and users' ability to assess associated privacy risks. To this end, our AlterEgo app provides users with a *Privacy Dashboard*. Such an envisioned Dashboard is an integration point for a variety of ex-ante and ex-post transparency features (cf. [40]), all of which leverage data flow monitoring and visualization services provided by trusted modules in the AlterEgo Middleware. More precisely, leveraging trusted modules in the AlterEgo Middleware, the app logs data requested by the branded app and data broadcasted by the user's smartphone and interceptable by sensors in its vicinity. Through overview charts, indicators and scores integrated in the Dashboard, the user can visualize sensors around, data collected by the branded app, data broadcasted by the smartphone and inferable information such as the user's favorite location, among others (satisfying R.1). In addition, the Privacy Dashboard provides end-user interfaces to create privacy preferences, visualize subsequent privacy consequences of data disclosure, and take part in the semi-automated process of negotiating privacy agreements with the MLA operator. Indeed, upon becoming aware of sensitive data being collected and what the privacy implications might be, the AlterEgo app helps the user activate a variety of control options. Examples of such options include the automated context-dependent modification of the device's privacy profile (e.g. by switching off all or a subset of relevant device sensors based on geo-fencing events) and the specification of privacy preferences that should govern subsequent use of her data. Acknowledging the fact that even fine-grained specification and enforcement of privacy agreements do not totally eliminate the risk of data re-purposing, we envision the AlterEgo framework to be realized as a tool that supports both ex-ante and ex-post transparency. That is, in addition to supporting transparency and control prior to the disclosure of user data to an MLA operator, the Privacy Dashboard also provides transparency after data disclosure. Being an ex-post transparency tool, the Privacy Dashboard provides visualizations and interfaces aiming to make data handling practices and user profiling by the MLA provider as well as privacy consequences thereof transparent to previously potentially unaware users. It lets the users track and understand how sensor data emitted by their mobile devices, photo and video footage captured by in-store cameras is used e.g., to create and deliver personalized context-aware offerings. In addition, we envision a Dashboard designed to interface with mechanisms to detect discrepancy between observed data handling practices and the terms of the privacy agreement with the MLA provider. For this purpose, it relies on trusted components in the cloud environment to obtain insight into the actual processing of sensitive data being performed by the MLA provider. The result of such a discrepancy check, including the main causes of the violation of the privacy agreement, is then presented to the user as charts and reports within the Dashboard. Moreover, the Privacy Dashboard offers risk assessment and risk visualization features. To do so, it interacts with a remote cloud-based Privacy Analytics service which actually assesses privacy risks by applying inference algorithms on data obtained from the trusted data handling logging service hosted in the MLA cloud. Informed by the level and possible consequences of privacy risks, the user (or her AlterEgo app) can undertake additional corrective and preventive measures, including consent withdraw, renegotiation of the terms of the custom privacy agreement with the MLA provider, or service termination and request to delete all related data. The dual nature of our Privacy Dashboard, i.e., it providing both ex-ante and an ex-post transparency, raises significant design challenges. Chief among them is the fact that a proper implementation of our proposal might require a certain level of firmware and/or hardware modification (e.g., to realize secure storage and management of privacy preferences/agreements on embedded mobile devices and sensors) and improvements of current wireless signaling protocols (e.g., to support signaling of fine-grained specified preferences). Another set of challenges emerges especially from the need to provide effective ex-post transparency, i.e., to reliably record, analyze, and visualize details about the actual use of sensitive data by the cloud-based MLA service. For instance, it raises the question of how to reliably track and log data handling by the MLA operator, given the fact that the user typically does not own nor control the backend. Moreover, the question of how to design the data handling logging service in a way that balances the user's need for privacy (to avoid logging service being exploited as a surveillance tool by unauthorized third parties) and the MLA provider's interest to restrict access to complex and proprietary algorithms they often treat as trade secrets remains largely unanswered.

4. CONCLUSION

Today's realizations of MLA exhibit characteristics and features that might be considered as unlawful under EU privacy laws. Designing MLA with privacy and data protection in mind would not only help to achieve compliance goals, it might also solidify user trust in an emerging industry, thus playing a crucial role in acceptance and widespread deployment of new MLA scenarios and applications. Our position, against this background, is that privacy and data protection in MLA requires a holistic and user-centered framework. This implies i) addressing privacy issues across MLA subsystems or domain boundaries, and ii) enabling the user to effectively shape the ways in which sensor-enriched venues collect and use data about her. We believe that besides unlinkability and data anonymization by design as well as means to ensure infrastructure security, empowering users with transparency-enhancing tools that leverage scalable data mining/machine learning algorithms and visualizations is the way forward. The proposed framework supports users in assessing and visualizing privacy risks arising from their interactions with an MLA infrastructure, and in negotiating and enforcing terms that should govern the collection and usage of their (device) data. While additional research is clearly needed, we hope our preliminary results will enhance understanding of privacy issues in the context of MLA, may inform the design of the next generation of MLA technology, and pave the way for new approaches to privacy management in sensor-enriched environments.

Acknowledgment

This work has been supported in part by the German Federal Ministry of Education and Research (BMBF) within the project "Forum Privatheit und selbstbestimmtes Leben in der Digitalen Welt" (http://www.forum-privatheit.de/).

5. REFERENCES

[1] C. Smith, "How beacons – small, low-cost gadgets – will influence billions in us retail sales." 2015.

[2] IDC, "Always connected how smartphones and social keep us engaged." 2012.

[3] S. Clifford and Q. Hardy, "Attention, shoppers: Store is tracking your cell." *NYT*, 2013.

[4] C. Matthews, "Private eyes: Are retailers watching our every move?" *Business Time*, 2012.

[5] Future of Privacy Forum. (2015) About mobile location analytics technology.

[6] F. Rosa, S. Sillani, F. Nassivera, and M. Vasciaveo, "Language, ethnical identity and consumer behavior: A cross-cultural study of marketing communication in the region fvg," *Proceedings in Food System Dynamics*, pp. 197–219, 2014.

[7] P. Rubens. (2015) Facial recognition: Shop where everybody knows your name.

[8] V. Corporation. (2014) Http://www.videomining.com/technology.

[9] M. Cunche, M. A. Kaafar, and R. Boreli, "I know who you will meet this evening! linking wireless devices using wi-fi probe requests," in *World of Wireless, Mobile and Multimedia Networks, 2012 IEEE International Symposium on a.* IEEE, 2012, pp. 1–9.

[10] Y. Gwon, R. Jain, and T. Kawahara, "Robust indoor location estimation of stationary and mobile users," in *INFOCOM 2004.*, vol. 2. IEEE, 2004, pp. 1032–1043.

[11] L. Demir, M. Cunche, and C. Lauradoux, "Analysing the privacy policies of wi-fi trackers," in *Proceedings of the 2014 workshop on physical analytics.* ACM, 2014, pp. 39–44.

[12] J. Gallinaro, "Meet your new big brother: Weighing the privacy implications of physical retail stores using tracking technology." *George Mason Law Review 2015-2015 Vol. 22 No. 2*, 2015.

[13] J. Priest and D. Johnson, "Covert channel over apple ibeacon," in *Proceedings of the International Conference on Security and Management*, 2015, p. 51.

[14] C. Miller, "Exploring the nfc attack surface," *Proceedings of Blackhat*, 2012.

[15] N. Asokan, L. Davi, A. Dmitrienko, and S. Heuser, *Mobile Platform Security.* Morgan & Claypool Publishers, 2013.

[16] C. Isidore, "Target: Hacking hit up to 110 million customers." 2014.

[17] C. Heffner. (2013) Exploiting network surveillance cameras like a hollywood hacker.

[18] M. Tehranipoor and F. Koushanfar, "A survey of hardware trojan taxonomy and detection," 2010.

[19] J. Leber, "How wireless carriers are monetizing your movements," 2013.

[20] Y.-A. de Montjoye, L. Radaelli, V. K. Singh *et al.*, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, 2015.

[21] X. Zhou, S. Demetriou, D. He, M. Naveed, X. Pan, X. Wang, C. A. Gunter, and K. Nahrstedt, "Identity, location, disease and more: Inferring your secrets from android public resources," in *Proceedings of the 2013 ACM SIGSAC CCS.* ACM, 2013, pp. 1017–1028.

[22] E. Dwoskin, "What secrets your phone is sharing about you - businesses use sensors to track customers, build shopper profiles," *Wall Street Journal*, 2014.

[23] J. Yap, "User profiling fears real but paranoia unnecessary," 2011.

[24] H. S. Fhom, "Big data: Opportunities and privacy challenges," *arXiv preprint arXiv:1502.00823*, 2015.

[25] D. Lyon, *Surveillance as social sorting: Privacy, risk, and digital discrimination.* Psychology Press, 2003.

[26] Future of Privacy Forum. (2013) Mobile location analytics code of conduct.

[27] OECD, "Oecd guidelines on the protection of privacy and transborder flows of personal data," Tech. Rep., 1980.

[28] U.S. Federal Trade Commission, "Privacy online: Fair information practices in the electronic marketplace: A federal trade commission report to congress," FTC, Washington, DC, USA, Tech. Rep., 2000.

[29] V. Reding, "The eu data protection reform 2012: Making europe the standard setter for modern data protection rules in the digital age," in *Innovation Conference Digital, Life, Design Munich*, vol. 22, 2012.

[30] G. Danezis, J. Domingo-Ferrer, M. Hansen, J.-H. Hoepman, D. L. Metayer, R. Tirtea, and S. Schiffner, "Privacy and data protection by design-from policy to engineering," *arXiv:1501.03726*, 2015.

[31] F. T. Commission *et al.*, "Protecting consumer privacy in an era of rapid change," *FTC Report, Washington, DC*, 2012.

[32] A. Narayanan and E. W. Felten, "No silver bullet: De-identification still doesn't work," 2014.

[33] P. Samarati and F. Gey, "Final research report on next generation policies," 2011.

[34] N. Davies, A. Friday, P. Newman, S. Rutlidge, and O. Storz, "Using bluetooth device names to support interaction in smart environments," in *Proceedings of the 7th international conference on Mobile systems, applications, and services.* ACM, 2009, pp. 151–164.

[35] M. Langheinrich, "A privacy awareness system for ubiquitous computing environments," in *UbiComp 2002.* Springer, 2002, pp. 237–245.

[36] F. Roesner, D. Molnar, A. Moshchuk, T. Kohno, and H. J. Wang, "World-driven access control for continuous sensing," in *Proceedings of the 2014 ACM SIGSAC Conference CCS.* ACM, 2014, pp. 1169–1181.

[37] J. Winter, "Trusted computing building blocks for embedded linux-based arm trustzone platforms," in *Proceedings of the 3rd ACM workshop on Scalable trusted computing.* ACM, 2008, pp. 21–30.

[38] K. Eldefrawy, G. Tsudik, A. Francillon, and D. Perito, "Smart: Secure and minimal architecture for (establishing dynamic) root of trust." in *NDSS*, vol. 12, 2012, pp. 1–15.

[39] G. Karjoth, M. Schunter, and M. Waidner, "Platform for enterprise privacy practices: Privacy-enabled management of customer data," in *Privacy Enhancing Technologies.* Springer, 2003, pp. 69–84.

[40] S. Fischer-Hübner and S. Berthold, "Privacy-enhancing technologies," *Morgan Kaufmann Publishers*, 2013.