Modeling the C(o)urse of Privacy-critical Location-based Services – Exposing Dark Side Archetypes of Location Tracking

Fabian Burmeister
University of Hamburg
burmeister@informatik.uni-hamburg.de

Paul Drews
Leuphana University of Lüneburg
paul.drews@leuphana.de

Ingrid Schirmer
University of Hamburg
schirmer@informatik.uni-hamburg.de

Abstract

With the ubiquitous use of mobile devices, location-based services (LBS) have rapidly pervaded daily life. By providing context- and location-specific information, LBS enable a myriad of opportunities for individuals and organizations. However, the manifold advantages come along with a radical increase in location privacy concerns and non-transparent data flows between the various actors involved. While research often focuses on protecting the dyadic relation between the user and LBS provider, the entirety of dark sides constituting privacy violations remains hidden. In this paper, we follow the paradigm of architectural thinking to shed light on the diverse dark sides emerging in today's LBS. By drawing on a multiple case study and developing a notation for architectural maps that help understand LBS from a socio-technical and privacy-oriented perspective, we reveal six dark side archetypes of LBS.

1. Introduction

The worldwide spread of mobile devices along with technological advances that enable accurately locating a user's or object's position has led to a rapid increase in location-based services (LBS) [1]. Often realized via mobile applications, LBS provide information tailored to the location and context of a user [2]. While in the past LBS were primarily used for navigation purposes, their scope has been expanded to social networking (e.g., locating friends), advertising (e.g., promotional alerts), healthcare (e.g., fitness monitoring) and other domains (e.g., weather forecast) [2, 3]. In fact, the global LBS market size was valued at $36.2 billion in 2019 and is predicted to reach $157.3 billion by 2026 [4].

Although LBS offer several advantages, they require people to disclose their location, personal preferences, and the context they are currently facing [1, 2, 5]. In addition, LBS often come at the cost of sharing one's private identity and location data with untrusted or even unknown third parties, raising serious privacy concerns [6]. These concerns refer, for example, to the disclosure of visited locations or daily habits [2, 6]. While service providers declare they process and share personal data only in an aggregated and non-identifiable form, the idea that personal data can successfully be anonymized is controversially discussed [3, 7]. Indeed, in a 2019 study, researchers were able to correctly re-identify 99.98% of participants in an anonymized dataset [7]. In this regard, location data can be considered as particularly privacy-critical by acting as a quasi-identifier of users through a sequence of spatio-temporal constraints [6, 8]. Recently, as part of its long-term privacy project, The New York Times published a special issue on the threat posed by LBS, highlighting how easily people can be identified by only using location data, i.e., without identifiers like user's Ad ID or phone number [9]. Identifying a person requires only four time-stamped location records [10].

LBS research has created a large body of knowledge on privacy-enhancing technologies and privacy threat models, and a plethora of algorithms and methods to prevent inference attacks [1, 3, 6, 8]. However, scholars claim that most studies focus on the dyadic relationship between the user and service provider, leading to a lack of transparency about socio-technical relations between the various actors participating in LBS and a missing understanding of the associated diversity of dark sides that may impair user's location privacy [2, 11, 12]. In addition, both information systems (IS) and privacy researchers call for more design science orientation to provide approaches that help different practitioners (e.g., app developers, policy makers) capture the complexity of data sharing mechanisms today and enable a common understanding of privacy-related dark sides [11, 13].

Against this background, in this paper we aim to shed light on the heterogeneity of dark sides affecting location privacy as well as on the underlying determinants. For this purpose, we draw on the paradigm of architectural thinking [14, 15] to decompose LBS from both a socio-technical and privacy-oriented perspective. Having its origin in the enterprise architecture management (EAM), architectural thinking is a rather lightweight approach that seeks to support researchers and practitioners in understanding complex causalities through architectural

URI: https://hdl.handle.net/10125/71419
978-0-9981331-4-0

maps that visualize socio-technical elements and their relations in a simplified form [16]. Following a design science approach [17], we develop a notation for LBS-related architectural maps and use this notation as a means to identify privacy-related dark sides. Moreover, we classify these dark sides into archetypes representing recurring practices of privacy violation. We argue that exploring LBS from an architectural perspective reveals causalities for privacy-related dark sides at a detailed socio-technical level and assists both researchers and practitioners in various tasks related to location privacy (e.g., legal judgment of specific LBS). Therefore, our study deals with the following research question:

Which archetypes of privacy-related dark sides can be identified in location-based services by taking an architectural perspective?

To answer this research question, we conduct an explorative multiple case study [18] of privacy-critical cases related to LBS, which are widely reported in the public media. By collecting and analyzing data on these cases, we first develop the notation and then apply it to identify dark side archetypes in the field of LBS.

The remainder of this paper is structured as follows. In the next section, we summarize literature relevant to our research context. Then, we present our methodology and continue with a description of our results. Finally, we discuss our results and give a conclusion.

2. Related research

We identified three streams of related research. The first stream points out definitions, types, and elements of LBS. The second stream describes location privacy and related main areas of research, but also introduces dark sides and archetypes. The third stream differentiates architectural thinking from EAM and outlines models for extending architectural maps through a privacy lens.

2.1 Location-based services

In recent years, the availability and use of LBS has increased significantly [1, 4]. Roughly defined, LBS are “services that take the user’s current or past location as input to provide a service” [11, p. 148]. Other authors concretize the activities performed by LBS, defining them as “services that create, compile, select, or filter information based on the current locations of the users or those of other persons or mobile objects” [19, p. 214]. While the first definition is rather simple and lacks a specification of relevant entities, the second definition does not consider past location data. Abbas et al. define LBS as “an application that combines the location or position of a mobile device associated with a given entity (individual or object) together with contextual information, to offer a value added service and/or fulfil a particular need for the user across a wireless network” [20, p. 3]. Our work follows this definition, because it considers the context of a service as an aspect essential to judge privacy compliance of LBS [20, 21].

The literature suggests several criteria for classifying LBS. In the first instance, it can be distinguished between push-based LBS, where location-related information is proactively provided to the user when a specific event occurs, and pull-based LBS, where users directly request location-related information [19, 22]. In addition, LBS can be classified into single-target (tracking the position of a certain target) or multi-target LBS (interrelating the positions of many targets) and outdoor or indoor LBS [2, 19]. Küpper and Treu list further classifications [19].

The LBS value chain is realized by the interaction of various social and technical elements. While the former embody the different actors in LBS, such as users, LBS providers, developers, network operators, and content providers [5, 11, 19], the latter refer to the underlying information technology (IT), including devices (e.g., smartphones), communication networks (e.g., wireless local and cellular networks), positioning components (e.g., global positioning system (GPS)), and software (e.g., operating systems) [3, 5, 11]. The LBS value chain is unique insofar as a single provider is unable to make a complete offering to customers, leading to an inter-organizational matter, i.e., a situation where multi-actor collaboration is crucial [20]. Moreover, as LBS are used in dynamic and mobile environments, they are aware of the context their users are in and accordingly customize the content and presentation of information [23].

2.2 Location privacy and dark side archetypes

Research on location privacy is gaining increasing importance. As a subset of information privacy, which is “the ability of the individual to control the terms under which personal information is acquired and used” [22, p. 138], location privacy refers to “the capability of the target person to exercise control about who may access her location information in which situation and in which level of detail” [19, p. 233]. Wang and Liu stress three unique characteristics of location privacy, which impose major research challenges [8]. First, requirements for location privacy are inherently user-dependent (e.g., some users regard their location as private, while others care about service quality). Second, there is a trade-off between location privacy and utility (e.g., more precise location data leads to higher service quality). Third, location data is updated frequently and processed in real time, bearing the risk of inferring user locations [8].

In IS research, most studies on location privacy focus on the development of technical mechanisms to prevent different types of attacks [3, 6, 8]. Based on their target, these attacks can be classified into identity attacks (e.g., de-anonymization of users by their address) and location attacks (e.g., exposure of regularly visited places) [5]. As a counterpart, location privacy preserving mechanisms include cryptographic methods (e.g., encryption of user positions), anonymization techniques (e.g., suppression and generalization to achieve k-anonymity), as well as obfuscation (e.g., dummy locations mask true positions) [1, 5, 6]. While these attacks and mechanisms usually refer to the dyadic relation between a user and malicious actor, the complexity of privacy violations in today's LBS requires exploring the manifold dark sides that are unwittingly triggered by users or appear in different data sharing practices of socio-technical elements [11, 20].

Dark sides can be defined as "'negative' phenomena that are associated with the use of IT, and that have the potential to infringe the well-being of individuals, organizations, and societies" [24, p. 161]. In our context, we also understand dark sides as privacy-critical actions or mechanisms in LBS that are hidden to or not expected by actors, especially users, and are part of or constitute a privacy violation. Classifying archetypes that are "a very typical example of a certain person or thing" [25], can help identify recurring patterns of dark sides causing a privacy violation. Therefore, we use the term dark side archetype to describe typical examples of how multiple dark sides in combination or in a specific sequence lead to a violation of location privacy. Archetypes are special as they do not only occur in one case, but can be found across multiple cases. Schilling et al. explicitly call for increased attention to archetypes in IS research [26]. To identify the archetypes, we consider both the context and sequence in which dark sides appear in a case. While the literature outlines types of privacy violations in LBS, such as location-based advertising or profiling, there is a lack of knowledge about the specific causalities [12].

2.3 Architectural maps and privacy models

Architecture comprises "the fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution" [27, p. 2]. The architecture of an enterprise has business, software, and hardware layers, and is managed by the EAM, which seeks to improve business IT alignment and transparency [15]. For this, EAM refers to enterprise architecture (EA) models that address specific stakeholder concerns [14]. A prevalent EA modeling language is ArchiMate, which proposes several elements and relations to structure the EA [28].

However, recent studies claim that EAM is mainly used by IT experts and is rather formal [14, 15]. Thus, EAM should evolve to architectural thinking as a more pragmatic approach performable by non-architects [14]. Instead of providing complex EA models, architectural thinking builds upon architectural maps that visualize socio-technical elements and relations in a simple form [16]. Thereby, critical causalities can be highlighted and discussed by different types of decision-makers [14, 16].

So far, architectural maps have only been proposed at the enterprise level in form of strategic theme maps, capability maps, or value stream maps, but not in the inter-organizational privacy context [16]. Closest to such architectural maps are privacy models that classify the different actors involved in data sharing networks [11, 20, 29]. For example, Conger et al.'s model suggests differing between first, second, third, and fourth parties [29]. In the context of LBS, a few models cover some technical elements like LBS servers or mobile devices [3, 5]. However, all these models focus on explaining the actors that are generally interacting in LBS, but do capture neither the concrete socio-technical relations nor the related dark sides causing privacy violations in LBS.

Summing up, IS research on location privacy often focuses on technically improving the security between users and LBS providers, but is lacking a comprehensive understanding of the various dark sides, especially those caused by third parties, that lead to privacy violations. Moreover, pragmatic approaches are missing that enable visualizing socio-technical relations and dark sides in LBS. By addressing our research question, we aim to contribute to this research gap and support researchers and practitioners in performing different tasks related to location privacy, such as case analysis or legal judgment.

3. Methodology

In our study, we followed a design science oriented research approach [17]. We developed a notation for architectural maps that visualize LBS from both a socio-technical and privacy perspective as a means to identify dark side archetypes currently emerging in LBS. In a multiple case study [18], we identified the modeling elements for the notation exploratory while searching for criteria to classify the archetypes. We evaluated the notation in a focus group and modeled all cases to verify the archetypes. Figure 1 shows the steps of our study.

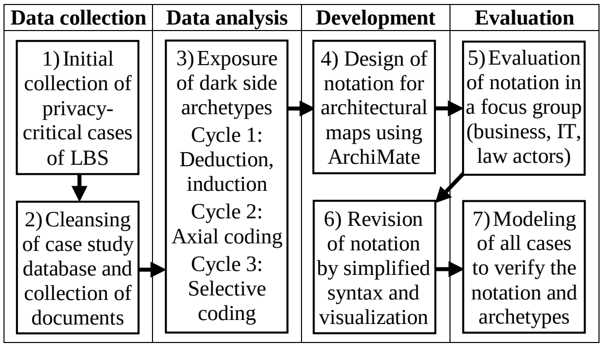
Data collection
Data analysis
Development
Evaluation
1) Initial collection of privacy-critical cases of LBS
2) Cleansing of case study database and collection of documents
3) Exposure of dark side archetypes
Cycle 1: Deduction, induction
Cycle 2: Axial coding
Cycle 3: Selective coding
4) Design of notation for architectural maps using ArchiMate
5) Evaluation of notation in a focus group (business, IT, law actors)
6) Revision of notation by simplified syntax and visualization
7) Modeling of all cases to verify the notation and archetypes

Figure 1. Methodology

3.1 Data collection

In the initial collection of privacy-critical cases, we followed a theoretical replication logic [18] where cases are selected to predict contradictory results. We hereby intended to cover a high variety of cases. Privacy-critical in our sense are cases reporting on a misuse of location data or a threat to location privacy caused by an LBS provider, but not cyberattacks or IT security incidents. Figure 2 shows our process of case selection. In the first step, we searched for news articles on four widespread news platforms (CNet, The Guardian, The New York Times, and ZDNet). We used the search term "Privacy AND Location OR GPS" and considered a publication period from January 2018 to May 2020, resulting in 1308 potential cases (n is the number of cases, i.e., in steps 1 and 2 every news article could have represented a suitable case). In the second step, we read the title of all news articles in our filter and, if applicable, their abstract. While we recorded 212 articles describing an eligible privacy-critical case in a case study database [18], we excluded the others. For example, we had to exclude hundreds of articles reporting on cyberattacks or security updates. In the third step, we studied the 212 articles, removed those not focusing on location data, and removed duplicates by assigning articles reporting on the same scandal to a unique case number, resulting in 29 cases. In the fourth step, we collected additional data material on each case, such as official responses from accused LBS providers, by conducting a backward search using the links in the articles. In the fifth step, we aggregated the data material per case and excluded eight cases exhibiting insufficient or inaccurate information. Following our aim of identifying archetypes, in the sixth step we switched to a literal replication logic [18] and selected those cases predicted to provide similar results by comparing their content and considering their impact.

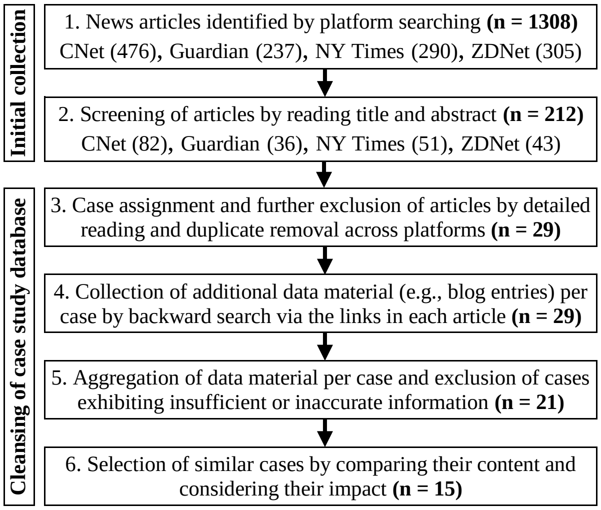
Initial collection
1. News articles identified by platform searching (n = 1308)
CNet (476), Guardian (237), NY Times (290), ZDNet (305)
2. Screening of articles by reading title and abstract (n = 212)
CNet (82), Guardian (36), NY Times (51), ZDNet (43)
Cleansing of case study database
3. Case assignment and further exclusion of articles by detailed reading and duplicate removal across platforms (n = 29)
4. Collection of additional data material (e.g., blog entries) per case by backward search via the links in each article (n = 29)
5. Aggregation of data material per case and exclusion of cases exhibiting insufficient or inaccurate information (n = 21)
6. Selection of similar cases by comparing their content and considering their impact (n = 15)

Figure 2. Process of case selection

Table 1 lists the data material collected for the final 15 cases, together with a link to an exemplary news article. In total, we collected 83 news articles, 9 blog entries, 19 tweets, 4 official responses, 2 studies, and 7 technical reports throughout the cases. The diversity of documents allowed us to triangulate data per case and thus increase the reliability and validity of our study [18].

Table 1. Data material collected per case

No.	Description	N	B	T	O	S	R	Link
1	Strava leaked secret army bases	7	1	2			1	Link
2	MoviePass tracked user locations	4		1	1		1	Link
3	Kids gaming apps share user data	6	2			1		Link
4	Polar leaked soldiers' location data	5			1			Link
5	GasBuddy sells data to Reveal Mobile	7	1	1				Link
6	Weather Channel app amasses data	6		1				Link
7	Mobile carriers sold location data	7	1	3			2	Link
8	AccuWeather shared location data	11	1	2	1	1		Link
9	Shutterfly collects visits via photos	3					1	Link
10	Netflix's Android app tracked users	3		1				Link
11	Pokémon GO data used for profiling	7	1	2				Link
12	TikTok accused of sharing user data	5	2	4	1			Link
13	Family locator Life360 shares data	4						Link
14	Ring's Neighbors leaks location data	5		2			1	Link
15	Care19 app shares sensitive user data	3					1	Link
	Σ	83	9	19	4	2	7	

Legend: N = news article, B = blog entry, T = tweet, O = official response, S = study, R = technical report

3.2 Data analysis

To identify dark side archetypes and related socio-technical elements across the 15 cases, we conducted a qualitative content analysis of the data material [30] via MAXQDA. We followed Saldaña's advice that multiple coding cycles constitute a rigorous data analysis [30], as we performed three coding cycles. In the first cycle, we combined deduction and induction. We set up a coding agenda [30] consisting of a priori codes we deductively received from the literature. These codes included LBS' basic elements (e.g., devices, users) [3, 5, 11, 19], actors and relations highlighted by privacy models (e.g., fourth parties) [11, 29], and key aspects for assessing privacy violations (e.g., context change) [21, 22]. As induction also enabled an open coding of the content, we refined our coding agenda with several codes we identified that were not covered by the selected literature. For example, in many cases software development kits (SDK) and data aggregators were considered as malicious elements, but also privacy-critical actions of LBS providers like user profiling. In the second coding cycle, we reviewed and reorganized the codes we received from induction and deduction by using axial coding. As the coding cycle evolved, we combined the codes into broader, theme-focused categories [30]. For example, while we assigned the codes "user" and "LBS provider" to the category "actors", we categorized "augmented reality game" and "SDK" as "applications". In the third coding cycle, we used selective coding to reveal cause-effect patterns in the data material [30]. For each case, we interlinked the categories and codes across the text passages to identify the relations constituting dark sides. By comparing the dark sides across the cases, we revealed six archetypes.

3.3 Development and evaluation

Based on the identified socio-technical elements and relations constituting dark sides in LBS, we developed a notation for deriving architectural maps that seek to help researchers and practitioners understand privacy-critical cases at first glance. For the maps' layout, we used the categories from our axial coding as layers and referred to the syntax and graphical notation of ArchiMate [28]. Adding an actor layer was especially important to ensure an inter-organizational perspective. To demonstrate and evaluate the notation in a focus group, we exemplarily modeled architectural maps for the cases 8 and 11, since these are described as particularly privacy-critical and exhibit different characteristics. The focus group session was attended by three lawyers, two executives, one IT expert, and one researcher. Together we discussed the usability and utility of the maps. Regarding usability, the participants highly appreciated the layered structure and simplicity of the maps. However, since ArchiMate's original notation looks slightly formal, they suggested integrating icons to distinguish the modeling elements more clearly and using different line styles to clarify the multiple types of relations. Regarding utility, the experts approved that lightweight visualizations, such as our architectural maps, are urgently needed to make privacy violations comprehensible in a simple form. Especially the lawyers emphasized the lack of such visualizations in the legal literature, since they provide a valuable basis for discussing complex causalities. After improving the notation in line with the evaluation results, we modeled all 15 cases to verify the notation's completeness. By comparing the modeled relations for similarity, we were also able to validate the correct assignment of each case to one of the six archetypes.

Table 2 lists the modeling elements of our notation. Following the claim that architectural thinking is rather lightweight, our notation covers those socio-technical elements and relations necessary to give an overview of privacy-related dark sides in LBS. Hence, architectural maps derived from our notation are not intended to give details of internal data processing, for example. The notation suggests modeling elements distributed across actors, applications, and IT infrastructure layers, as well as their different types of relations, such as data flows. The dark side element fulfills two functions: it describes a privacy-critical action and allows the tracing of a case by specifying a sequence. Timers additionally indicate the frequency with which a dark side occurs.

Table 2. Notation for architectural maps of LBS

	Element	Description	Icon
Actors	Actor	Covers users and LBS providers, but also other (malicious) parties involved, such as data aggregators or advertisers.	🧍
Applications	Software	Embodies applications, such as mobile apps providing LBS and operating systems of devices.	📱⚙
	SDK	Shows SDKs provided by third parties. SDKs are used to enhance applications by functions pre-defined in a package.	SDK
IT infrastructure	Device	Represents mobile devices of users, including smartphones, tablets, wearables, and others.	📱
	Server	Shows servers of LBS providers or third parties. Comprises technical entities like databases and is especially used for data analytics purposes.	🗄
Relations	Data flow	Illustrates any kind of data transmission between elements.	- Text - ➔
	Permission	Shows whether an element has granted permission or not.	···Text···➔
	Other	Specifies any other type of relation like ownership or integration of elements.	—Text➔
Dark Sides	Dark side	Describes a dark side caused by an element. Dark sides are bound to relations and sequenced by (n).	**(n)** *Text*
	Timer	Indicates the frequency with which a specific dark side occurs.	🕒
Clusters	Layer	Shows the actors, applications, and IT infrastructure layers.	Text
	Container	Clusters elements and reduces the number of relations needed.	**Text**

4. Results

While each case has unique characteristics, taking a socio-technical perspective in the data analysis revealed common dark sides across the 15 cases. By comparing the context and sequence in which the dark sides occur, we identified six dark side archetypes that represent prevalent violations of location privacy in LBS. We present each archetype by illustrating the architectural map of a selected case and explaining the similar cases.

4.1 Leakage of secret locations through publicly available maps offered by LBS providers

The first archetype is exposed by the cases 1 and 4. Case 1 refers to Strava, a social network for athletes, which allows its users to compare, time, and share their exercises. For this, Strava collects fitness and location data of running, cycling, and swimming routes via the Strava mobile app or fitness trackers. Strava offers a publicly available heat map that shows the activity of its users all over the world, containing 3 trillion latitude and longitude points. Military analysts noticed that the heat map is detailed enough to uncover secret army bases, as a subset of users are soldiers on duty. While these bases are invisible in LBS like Google Maps, their layout can be seen on Strava's heat map, where cities are aglow with jogging routes and foreign army bases in remote areas stand out as isolated hotspots. The heat map also reveals which bases are mostly used and which routes are taken by soldiers. In addition, soldiers can easily be tracked by cross-referencing their Strava data with other social media use. According to our analysis, three dark sides constitute this archetype (see Figure 3). First, most users (i.e., also soldiers) opt for including their data in the map, as this is Strava's default setting. Second, the map is publicly available and updated monthly. Third, the map can be misused for various malicious purposes.

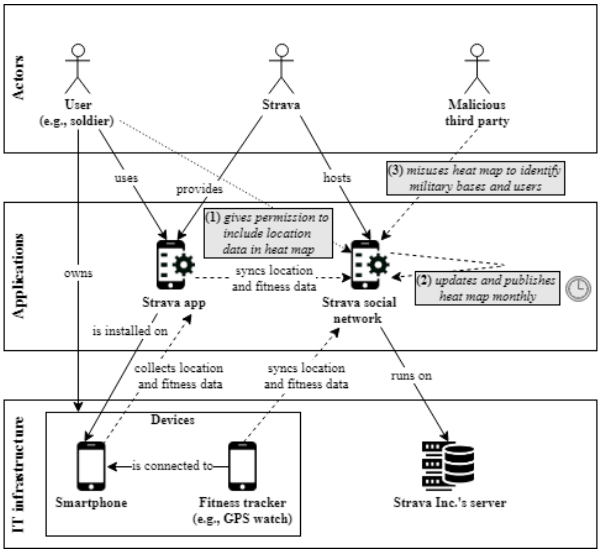

Figure 3. Architectural map of Strava leaking secret locations through its heat map (case 1)

The same happened with the fitness app Polar (case 4). For most users who set their data to public, posting their workouts on Polar's Explore map is a feature and not a privacy issue. Even if profiles are set to private, fitness activities can reveal where users live. Both cases demonstrate that location privacy not only refers to the user, but also to critical locations supposed to be secret.

4.2 Unexpected location-based advertising by sharing location data with data aggregators

The second archetype can be found in the cases 5 and 7. In case 5, the provider of GasBuddy, an app to check prices of nearby gas stations, sold data on user's latitude, longitude, and IP address together with timestamps and user's Ad ID, a code that uniquely identifies a particular person for advertising purposes, to the data aggregation and location-based marketing company Reveal Mobile. Reveal Mobile then shared the data with location-based advertisers, who were able to accurately retrace where and when users of GasBuddy have been.

However, while apps like GasBuddy can easily be uninstalled, changing the mobile carrier is cumbersome. In case 7, the four major mobile carriers AT&T, Sprint, T-Mobile, and Verizon sold their customers' real-time location data received via cell tower usage to the data aggregators LocationSmart and Zumigo, who resold the data in a prepared form to advertisers and other actors like Microbilt, a company that offers phone tracking services. Figure 4 shows the dark sides of this archetype. While GasBuddy legitimizes the collection of location data by its service provision, the mobile carriers lean on purposes of roadside assistance and fraud prevention. In this archetype, the first dark side occurs when location data is sent to a data aggregator without the expectation of users and out of the context of the service provided. Next, the data aggregators process the data and then resell it to fourth parties like advertisers. Last, the fourth parties influence people based on their location or even offer controversial phone tracking apps like Microbilt's Mobile Device Verify. In this regard, case 7 highlights another dimension of LBS: they can appear as phone tracking apps, where the target is not an object like a gas station, but the location of a specific person's device.

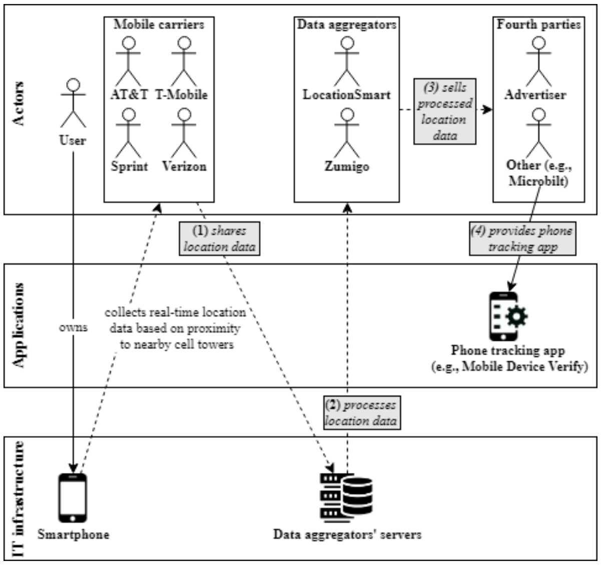

Figure 4. Architectural Map of mobile carriers selling location data (case 7)

4.3 Hidden integration of location-based services in mobile applications

Regarding the third dark side archetype, users are unaware that certain mobile apps include some kind of LBS tracking their location, since location data is not necessary to fulfill an application's actual purpose. In case 2 for example, the former movie ticketing service MoviePass collected and potentially shared the location data of its users with third parties. Users assumed that MoviePass collected data on ticket sales and movie choices, rather than detailed location data that allows tracking users before and after watching a movie. In addition, they claimed MoviePass did not disclose the location tracking in their privacy policy. Figure 5 shows the two dark sides of this archetype. First, users are tracked without their knowledge or even expectation, as location data is not necessary to fulfill the actual purpose of a service. Second, this data then might be sold to third parties that influence users based on their location.

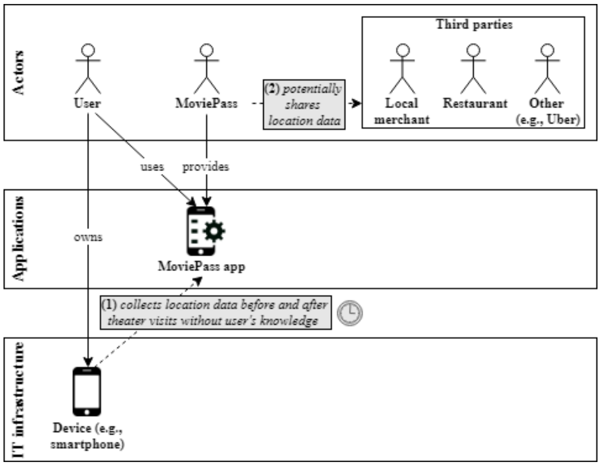
Actors
User
MoviePass
(2) potentially shares location data
Third parties
Local merchant
Restaurant
Other (e.g., Uber)
uses
provides
Applications
owns
MoviePass app
(1) collects location data before and after theater visits without user's knowledge
IT infrastructure
Device (e.g., smartphone)

Figure 5. Architectural map of MoviePass secretly tracking user locations (case 2)

Other examples for this dark side archetype are given by the cases 3, 10, and 12. While in case 3 a study found that 184 kid-targeted apps like Fun Kid Racing secretly collected and presumably shared GPS data, in case 10 Netflix's Android app tracked the location of several users without asking for permission. In case 12, TikTok, a social video app, is accused in a California lawsuit of sending personally identifiable user data, such as phone numbers and location data, to third parties like Appsflyer and Facebook. The Pentagon even classified TikTok as a security threat as it is also accused of storing this data on Chinese servers, which the government could access under Chinese law. All four cases emphasize that LBS are secretly integrated into various mobile apps without being relevant for service provision. They often appear outside the context of use and without user's knowledge.

4.4 Approximation of user locations by aggregating multiple data types

The key characteristic of the fourth archetype is the determination of the user location by collecting and aggregating different types of data, even though users explicitly denied access to their location data. In case 8, the AccuWeather app transmitted user's device data to Reveal Mobile, which approximated locations of users based on this data, although the users opted out of giving access to their location data (see Figure 6). The first dark side is that Reveal Mobile's SDK implemented in the AccuWeather app triggers the collection and transfer of device data. Next, the AccuWeather app continuously collects the device's Bluetooth status and the name and unique BSSID (basic service set identifier) of the WiFi router in use from the iOS operating system, and sends this data to Reveal Mobile every few hours. Aggregating the data, Reveal Mobile then approximates the location of users and sells this information to fourth parties.

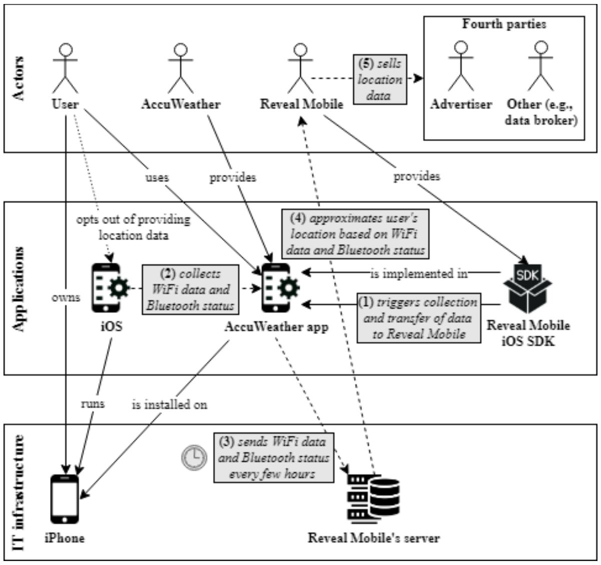
Actors
User
AccuWeather
Reveal Mobile
(5) sells location data
Fourth parties
Advertiser
Other (e.g., data broker)
uses
provides
provides
Applications
opts out of providing location data
(4) approximates user's location based on WiFi data and Bluetooth status
owns
iOS
(2) collects WiFi data and Bluetooth status
AccuWeather app
is implemented in
(1) triggers collection and transfer of data to Reveal Mobile
SDK
Reveal Mobile iOS SDK
runs
is installed on
IT infrastructure
(3) sends WiFi data and Bluetooth status every few hours
iPhone
Reveal Mobile's server

Figure 6. Architectural map of AccuWeather sending device data to Reveal Mobile (case 8)

In case 9, the photo-editing app Shutterfly defied user permissions by gleaning precise phone location data from photos and sending this data to its provider's servers. Even though the users denied access to location data, Shutterfly used the EXIF (exchangeable image file format) metadata that is generated by cameras and that integrates GPS coordinates and timestamps into photos. By aggregating this data, movements of users can be tracked. Both cases demonstrate that LBS are able to locate users even without having access to location data.

4.5 Extensive profiling by amassing user's location data

The fifth dark side archetype refers to LBS that are constantly tracking users, even when they are asleep or are actually not using the service. Based on this data deluge, LBS providers gain detailed insights about the life of their customers. A prominent example is Niantic, the creator of popular mobile augmented reality games like Pokémon GO and Harry Potter: Wizards Unite, who amasses location data of its players (case 11). Figure 7 shows the four dark sides our analysis revealed. First, while players allow Niantic's games to collect location data for proper functionality, they do not expect being tracked with such a high frequency, even when they are not playing, since this is not clarified in the privacy statements of the games. Second, the location data is sent to Niantic's server. Third, by processing the data, Niantic can discern individual patterns of user behavior and details about players, such as the number of calories they likely burned, the distance they traveled, and the promotions they engaged with. Some people might also play multiple games simultaneously, which increases the variety and precision of insights. Fourth, Niantic sells the location data and detailed profiles to third parties.

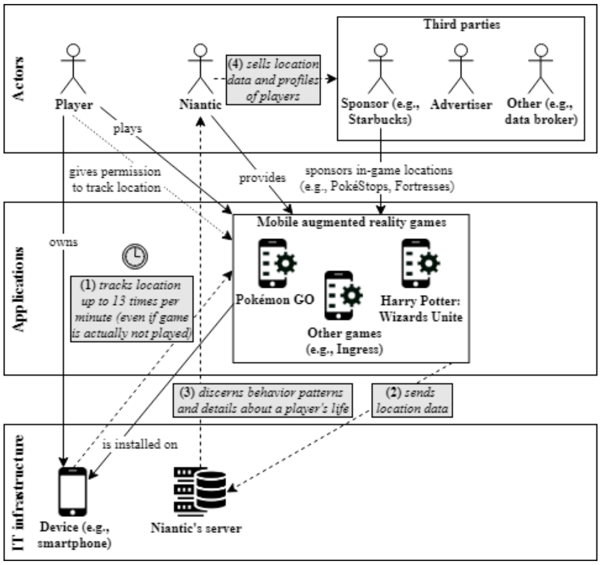
Actors
Player
Niantic
(4) sells location data and profiles of players
Third parties
Sponsor (e.g., Starbucks)
Advertiser
Other (e.g., data broker)
plays
gives permission to track location
provides
sponsors in-game locations (e.g., PokéStops, Fortresses)
Applications
owns
(1) tracks location up to 13 times per minute (even if game is actually not played)
Mobile augmented reality games
Pokémon GO
Other games (e.g., Ingress)
Harry Potter: Wizards Unite
(3) discerns behavior patterns and details about a player's life
(2) sends location data
IT infrastructure
is installed on
Device (e.g., smartphone)
Niantic's server

Figure 7. Architectural Map of Niantic amassing location data for profiling (case 11)

Another example is the Weather Company, which is accused of deceptively using its Weather Channel app to amass location data by tracking movements in minute detail, while making users believe their data would only be used to localize weather reports (case 6). According to the lawsuit, the company analyzed the data to identify daily habits, shopping preferences, and even the identity of users. Then, the profiles and location data were sold.

4.6 Misuse of protective location-based services

The latest reports we found refer to LBS that intend to provide protection for users, but have been misused for several purposes. Therefore, we identified the sixth archetype across the cases 13, 14, and 15. In case 13, the family locator Life360, which is already controversially discussed because of its child tracking function, collects not only location data of people, but also, for example, their driving speed and the battery life of their phones. The app shares the data with a risk-assessment firm that uses the data to calculate insurance prices. In case 14, a study found that the Neighbors app, which allows people to share video footage and to post on crime in their local area, leaked locations of devices via GPS coordinates not supposed to be accessible in any post. Consequently, exact positions of cameras and addresses of users were revealed. However, one of the most pressing privacy issues today is the use of contact tracing or COVID-19 apps. While they play an important role in containing the virus, there are reports of related privacy concerns like governmental surveillance. In case 15, a review found that North and South Dakota's Care19 app, created by ProudCrowd, sends data to the advertising and location technology firm Foursquare. Figure 8 illustrates the dark sides. First, Care19 has embedded Foursquare's Pilgrim SDK that supports apps by converting location data into concrete names of places, but in this case also triggers the collection and sharing of a user's Ad ID. Second, Care19 collects location data and the Ad ID from the user's device. Third, together with a unique citizen code generated by the app, the data is sent to Foursquare.

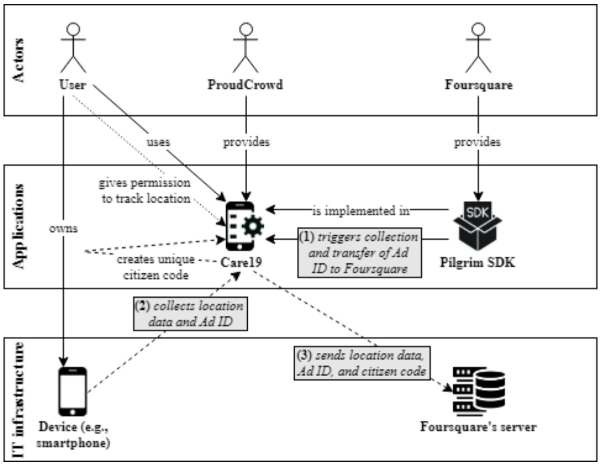
Actors
User
ProudCrowd
Foursquare
uses
provides
provides
Applications
gives permission to track location
owns
is implemented in
SDK
(1) triggers collection and transfer of Ad ID to Foursquare
Pilgrim SDK
creates unique citizen code
Care19
(2) collects location data and Ad ID
IT infrastructure
(3) sends location data, Ad ID, and citizen code
Device (e.g., smartphone)
Foursquare's server

Figure 8. Architectural Map of Care19 sharing personal data with Foursquare (case 15)

5. Discussion and conclusion

With the manifold advantages of LBS, an increase in privacy violations has been reported [9, 10]. Based on a multiple case study [18], in this paper we shed light on the diverse dark sides causing privacy violations in LBS. By following the paradigm of architectural thinking, we developed a notation for architectural maps that allow decomposing LBS from a socio-technical perspective and thereby facilitate understanding the diversity of dark sides affecting location privacy. By comparing the dark sides across 15 privacy-critical cases, we identified six dark side archetypes emerging in widespread LBS.

The archetypes highlight the various ways in which privacy violations in LBS can occur. Summarizing the dark sides per archetype, Table 3 gives an overview of each archetype's key characteristics. In comparison, certain aspects are particularly worth mentioning. First, it is remarkable that permissions to track user locations are granted in the archetypes 1, 2, 5, and 6. While users are unaware of being tracked in archetype 3, they denied access to location data in archetype 4. Second, third parties like data aggregators or advertisers are directly part of the value chain in the archetypes 2, 4, 5, and 6. While in archetype 1 third parties are not involved in the value chain, in archetype 3 it is assumed that data is sold to third parties. Third, in the archetypes 1 and 5, the LBS providers process location data on a large scale, whereas they mainly distribute data in the archetypes 2, 4, and 6. In archetype 3, it is rather unclear to what extent LBS providers process location data and for what purpose.

Table 3. Key characteristics of the identified dark side archetypes

Dark side archetype	Key characteristics
(1) Leakage of secret locations through publicly available maps offered by LBS providers	- No direct involvement of third parties like advertisers - Public map allows to infer sensitive routes and locations
(2) Unexpected location-based advertising by sharing location data with data aggregators	- Data aggregators as central intermediaries of data sharing - Frequent context changes lead to unexpected advertising
(3) Hidden integration of location-based services in mobile applications	- LBS not necessary for service provision of focal application - Hidden tracking of users
(4) Approximation of user locations by aggregating multiple data types	- Defiance of user permissions - Location is determined by combining different data types
(5) Extensive profiling by amassing user's location data	- Excessive data collection even if LBS are not actively used - Very detailed user profiling
(6) Misuse of protective location-based services	- Protective use of location data - Misuse of data for different purposes, such as advertising

Our results contribute to research and practice alike. From an academic point of view, the notation and maps complement previous research on modeling privacy-related elements in LBS. While existing privacy models often focus on classifying the different actors involved [11, 20, 29], our notation and related maps reveal dark sides causing privacy violations in LBS from a holistic architectural perspective. Moreover, differing between elements of actors, applications, and IT infrastructure layers helps understand socio-technical relations in LBS at a higher level of granularity. Our proposed notation and architectural maps additionally provide first steps towards extending architectural thinking, which has so far been discussed at the intra-organizational level [14, 15, 16], by a multi-actor perspective and to the field of LBS. Our research also exemplifies how dark sides can be identified using news articles as a primary data source and then structured via archetypes [26]. This approach can be used in other domains to reveal and classify dark sides. Above all, research often falls short of considering the various ways in which location privacy is violated. While many studies focus on anonymizing location data in the dyadic relation between a user and LBS provider [3, 6, 8], they often do not take into account the sharing of additional identifiers like the Ad ID (e.g., archetypes 2 and 6) or the inference of user locations by multiple data types (e.g., archetype 4). With our archetypes, we aim to increase sensitivity for the diversity of dark sides and underline the increasing influence of third parties.

Our results also have several practical implications. As our research is based on architectural thinking, which claims to be lightweight and pragmatic, we only included privacy-related key elements for modeling LBS in the notation. Due to their simplicity, derived architectural maps foster a common terminology and understanding between different types of practitioners. Both business and regulatory stakeholders can visualize privacy issues based on our notation and receive a discussion basis. For example, during the evaluation of our results, lawyers acknowledged that architectural maps would be useful for analyzing and debating privacy violations and that such visualizations are too scarce in the legal literature. In addition, practitioners like LBS providers can refer to our maps and archetypes to assess privacy compliance, but also to gain insights into the various ways in which location data can be collected, shared, and misused.

The results of our study are not without limitations. First, we only used 15 privacy-critical cases to identify the dark side archetypes. Considering more cases may have revealed further archetypes or led to a breakdown of those identified. Second, as we used news articles as a primary data source, some journalistic preconceptions may have affected the analysis. Nevertheless, according to Yin, news articles can serve as a source of empirical evidence [18]. Third, the results give rather an overview of privacy violations in LBS. An in-depth legal analysis per case needs to consider further aspects like purpose limitation [11], the distinction of push and pull services [22], and the evaluation of contextual integrity [21].

Future research is required to anchor architectural thinking in the domain of LBS, to improve the notation, and to complement the archetypes. Moreover, we can imagine a domain-specific modeling language for LBS or privacy research based on our notation. We encourage future research to study data sharing networks of LBS more intensively and to disclose the depth of dark sides.

6. References

[1] M. Gruteser and D. Grunwald, “Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking,” *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services (MobiSys '03)*, San Francisco, CA, 2003, pp. 31-42.

[2] H. Huang, G. Gartner, J. M. Krisp, M. Raubal, and N. Van de Weghe, “Location based services: ongoing evolution and research agenda,” *Journal of Location Based Services*, 12(2), 2018, pp. 63-93.

[3] N. Poolsappasit and I. Ray, “Towards Achieving Personalized Privacy for Location-Based Services,” *Transactions on Data Privacy*, 2(1), 2009, pp. 77-99.

[4] Allied Market Research, *LBS Market Statistics - 2026*, https://bit.ly/2ATusbA, accessed July 10, 2020.

[5] B. Liu, W. Zhou, T. Zhu, L. Gao, and Y. Xiang, “Location Privacy and Its Applications: A Systematic Study,” *IEEE Access*, 6, 2018, pp. 17606-17624.

[6] A. Khoshgozaran and C. Shahabi, “A taxonomy of approaches to preserve location privacy in location-based services,” *International Journal of Computational Science and Engineering*, 5(2), 2010, pp. 86-96.

[7] L. Rocher, J. M. Hendrickx, Y.-A. de Montjoye, “Estimating the success of re-identifications in incomplete datasets using generative models,” *Nature Communications*, 10(1), 2019, 3069.

[8] T. Wang and L. Liu, “From Data Privacy to Location Privacy,” *Machine Learning in Cyber Trust*, Springer, Boston, MA, 2009, pp. 217-246.

[9] S. A. Thompson and C. Warzel, “Twelve Million Phones, One Dataset, Zero Privacy,” *The New York Times Privacy Project*, https://nyti.ms/2YBgnHN, accessed July 10, 2020.

[10] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, “Unique in the Crowd: The privacy bounds of human mobility,” *Scientific Reports*, 3, 2013, 1376.

[11] M. Herrmann, M. Hildebrandt, L. Tielemans, and C. Diaz, “Privacy in Location-Based Services: An Interdisciplinary Approach,” *SCRIPTed: A Journal of Law, Technology & Society*, 13(2), 2016, pp. 144-170.

[12] K. Michael and M. G. Michael, “The social and behavioural implications of location-based services,” *Journal of Location Based Services*, 5(3-4), 2011, pp. 121-137.

[13] F. Bélanger and H. Xu, “The Role of Information Systems Research in Shaping the Future of Information Privacy,” *Information Systems Journal*, 25(6), 2015, pp. 573-578.

[14] R. Winter, “Architectural Thinking,” *Business & Information Systems Engineering*, 6(6), 2014, pp. 361-364.

[15] S. Aier, N. Labusch, P. and Pähler, “Implementing Architectural Thinking: A Case Study at Commerzbank AG,” *Trends in Enterprise Architecture Research – CAiSE 2015 International Workshops*, Springer, Berlin, 2015, pp. 389-400.

[16] W. Goebl, *Foundations of the Architectural Thinking Framework*, received from https://architectural-thinking.com/, accessed July 10, 2020.

[17] K. Peffers, T. Tuure, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems*, 24(3), 2007, pp. 45-77.

[18] R. K. Yin, *Case Study Research: Design and Methods*, Sage, Thousand Oaks, CA, 2009.

[19] A. Küpper and G. Treu, “Next Generation Location-based Services: Merging Positioning and Web 2.0”, *Mobile Intelligence*, Wiley, Hoboken, NJ, 2010, pp. 213-236.

[20] R. Abbas, K. Michael, and M. Michael, “The regulatory considerations and ethical dilemmas of location-based services (LBS): A literature review,” *Information Technology & People*, 27(1), 2014, pp. 2-20.

[21] H. Nissenbaum, “A Contextual Approach to Privacy Online,” *Daedalus*, 140(4), 2011, pp. 32-48.

[22] H. Xu, H.-H. Teo, B. C. Y. Tan, and R. Agarwal, “The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services,” *Journal of Management Information Systems*, 26(3), 2009, pp. 135-174.

[23] H. Huang and G. Gartner, “Current Trends and Challenges in Location-Based Services,” *ISPRS International Journal of Geo-Information*, 7(6), 2018, 199.

[24] M. Tarafdar, A. Gupta, and O. Turel, “Special issue on ‘dark side of information technology use’: an introduction and a framework for research,” *Information Systems Journal*, 25(3), 2015, pp. 161-170.

[25] Oxford Dictionaries, *Archetype*, https://www.lexico.com/definition/archetype, accessed July 10, 2020.

[26] R. Schilling, K. Haki, and S. Aier, “Introducing Archetype Theory to Information Systems Research: A Literature Review and Call for Future Research.” *Proceedings of the 13th International Conference on Wirtschaftsinformatik (WI)*, St. Gallen, Switzerland, 2017, pp. 574-588.

[27] ISO. *Architecture description*, standard 42010, 2011.

[28] The Open Group, *ArchiMate 3.1 Specification*, 2020.

[29] S. Conger, J. H. Pratt, and K. D. Loch, “Personal Information Privacy and Emerging Technologies,” *Information Systems Journal*, 23(5), 2013, pp. 401-417.

[30] J. Saldaña, *The Coding Manual for Qualitative Researchers*, Sage, Thousand Oaks, CA, 2015.