Privacy in Indoor Positioning Systems: A Systematic Review

Sylvia Holcer*,†, Joaquín Torres-Sospedra*,†, Michael Gould*,‡, and Inmaculada Remolar*,‡

*Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain
†UBIK Geospatial Solutions S.L., Castellón, Spain

Abstract—This article proposes a systematic review of privacy in indoor positioning systems. The selected 41 articles on location privacy preserving mechanisms employ non-inherently private methods such as encryption, k-anonymity, and differential privacy. The 15 identified mechanisms are categorized and summarized by where they are processed: on device, during transmission, or at a server. Trade-offs such as calculation speed, granularity, or complexity in set-up are identified for each mechanism. In 40% of the papers, some trade-offs are minimized by combining several methods into a hybrid solution. The combinations of mechanisms and their levels of offered privacy are suggested based on estimated user mobility cases.

I. INTRODUCTION

The Global Navigation Satellite System (GNSS) provides accurate location readings when outdoors, but it is not effective in indoor environments [1]. People on average spend 90% of their time indoors [2] yet there is no standardized Indoor Positioning System (IPS) fitting in all possible scenarios. Research in IPS is improving in accuracy, energy efficiency, and calculating speed [3], but privacy continues to lack definitive solutions [4]. The same level of privacy should be applied to location data as any other demographic data such as age, gender, income, education level, occupation, etc. Any combination of demographic data with location data, even coarse location data such as a postal code, might be enough to personally identify an individual.

Privacy is a growing concern as the number of wearable and Internet of Things (IoT) devices collecting location data continues to grow [5]. De Montjoye et al. [6] studied the mobility traces of smartphones and state that human mobility is highly unique. They conclude that four randomly chosen spatio-temporal points are enough to uniquely identify 95% of the individuals. Loss of location privacy has serious implications. Location information reveals home addresses, company travel, and visits to sensitive areas such as medical clinics, client locations, political events, etc [7]. This underlines the necessity to research privacy in IPS.

Corresponding Author: S. Holcer (holcers@uji.es)

The authors gratefully acknowledge funding from European Union’s Horizon 2020 Research and Innovation programme under the Marie Skłodowska Curie grant agreement No. 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, http://www.a-wear.eu/). J. Torres-Sospedra acknowledges funding from Ministerio de Ciencia, Innovación y Universidades (INSIGNIA, PTQ2018-009981).

The General Data Protection Regulation1 (GDPR) was created by the European Union to lawfully protect the personal data of its citizens. It states that personal data should be processed in a fair and transparent manner, for its intended purposes, keeping only what is necessary, with justified storage times, in a secure, confidential, accurate and accountable manner. The regulation defines personal data to include location data, which can identify a natural personal directly or indirectly. Therefore, privacy of location data guarantees the user that either they control the access to their data by others, or that their data gets processed in order to not contain any personally identifiable information. Liu et al. [8] look at all applications of location privacy. Their review summarizes location information as a three-part tuple <identity, position, time>, yet it is possible to lose privacy based on spatial information without time, using frequency alone to determine the likelihood of revisits by a user. They posit that users need to be guided to help them select the most appropriate Location Privacy Preserving Mechanisms (LPPM), and that there has been some research about automatically determining or recommending personalized privacy settings. Most of them rely on previous social media privacy settings.

The most prominent IPS technology is Wi-Fi [9] because it is relatively quick to implement, especially when using the fingerprinting method. In this case, privacy is two-fold. Allowing the Localization Server (LS) access to the user’s measurements gives it the possibility of tracking the user within the building. This may include continuous tracking, keeping historical records of the user’s location, and sharing these locations to third parties without the user’s knowledge. On the other hand, if the LS sends its database (also called a radio map) and algorithms to the user to let them calculate their location on their own, then the LS loses its privacy and can be abused by an adversary. Building and room layouts and all Access Points (APs) locations might be confidential to the operations of a military, hospitals, airports, government offices, etc.

Due to the ubiquitous presence of IPSs and location-based services (LBSs) in our personal devices, such as smartphones and wearables, we consider necessary to review the different mechanisms to enhance location privacy on those devices. Thus, this paper aims to systematically review all LPPMs.

1https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504
in IPS in order to discuss the current trends and analyze the possible lines for future work. The review is based on the PRISMA guidelines proposed in [10] to assess the pros and cons of a health care intervention with a wide array of systematic reviews and meta-analyses.

The remainder of this paper is organized as follows. Section II describe the methodology used for the systematic review and the datasets considered for the search. Section III introduce the main results retrieved from the search of related literature. Section IV discusses the current solutions and draws the lines for future work.

II. METHOD

The literature review follows a systematic review scheme proposed in the PRISMA guidelines. The search was performed on the Scopus and Web of Science databases. The results were combined (360 + 351 = 711) and the 229 duplicates were removed. Afterwards, 122 totally unrelated titles were removed from the combined list. The inclusion criteria are that the articles must use privacy preservation mechanisms in their work about indoor positioning systems. Exclusion criteria are sources that are not in the English language, are published before 2015, and that are not articles or conference papers. The search queries and results are reported in Table I. Filtered audio, video, or device-free indoor positioning systems are inherently private because they do not contain any personally identifiable information in order to operate, therefore the papers using these also were excluded.

| Web of Science | Scopus |
|----------------|--------|
| Search terms: TS = (indoor AND (privacy OR ethic*) AND (locali* OR locati* OR navigat* OR posit* OR track*)) | Search terms: TITLE-ABS-KEY (indoor AND (privacy OR ethic*) AND (locali* OR locati* OR navigat* OR posit* OR track*)) |
| Results: 360 documents in English | Results: 436 documents in English |

| Category | Web of Science | Scopus |
|----------|----------------|--------|
| Meeting | 182 | Conf. paper | 208 |
| Article | 178 | Article | 120 |
| Other | 18 | Conf. review | 87 |
| Patent | 14 | Review | 9 |
| Clinical Trial | 2 | Book chapter | 2 |
| Editorial | 2 | Book | 2 |
| Review | 2 | Editorial | 1 |
| Book | 1 | Note | 1 |
| Early access | 1 | Undefined | 2 |

III. RESULTS

A. Overview

41 articles fit the previously described inclusion and exclusion criteria [9, 11–50]. Several dimensions of this literature were explored: the technology used, the localization method, and the LPPMs. Wi-Fi was used in 70% of the papers. The remaining 12 papers either did not explicitly mention which technology was used, had used several, or had too few counts to consider meaningful correlations. The localization algorithms yielded similar results. Received Signal Strength (RSS) fingerprinting was used for most Wi-Fi localization, and there were only 2 papers that used trilateration, therefore these dimensions were not pursued.

All the LPPMs can be categorized into one of three groups based on the processing of the location data: on-device, during transmission, and at the server (see Fig. 1). Each of the methods will be summarized below.

B. On device

One way of dealing with keeping location information private from the LS is by keeping all localization calculations on the device itself. Schauer et al. [45] concentrate on passive Wi-Fi readings to estimate indoor location on the user's device in a method they call beacon-based fingerprinting. A model-based signal propagation algorithm was devised in [43] with specially developed firmware for Wi-Fi modules.

The PL-Protector middleware [25] is built between the platform component layer and the application layer. It prevents Google's fused location service from reaching the application location request, and instead apply privacy rules on cached locations. It is the only privacy solution that seriously considers the seven tenets of the Privacy-by-Design framework proposed for developers for socially acceptable and user-friendly privacy. The middleware’s drawbacks include being exclusive for Android systems, and initial set up requiring some technical knowledge which might be outside the scope of the technical abilities of some users. Locally computed positions and middleware are complex to implement because they require the technical background knowledge.

C. Transmitted data

Fig. 2 shows that encryption is the most popular mechanism, probably because it is a common solution for securely transmitting data. The papers that use encryption aim to balance semi-honest security models with good estimation accuracy and low computational overhead. Pseudo-certificates in [34] rely on trusted third parties (Certificate Authorities) for their protocol. The IMAKA-Tate method [16] is built upon a three-way handshake, using encrypted public keys exchanged between each side. In the OTPri method [50], the user’s mobile locally computes its location with an oblivious transfer. In this process it reveals several vicinity AP identifiers, which exposes a coarse-level location that can still be abused by
an inference attack. The PILOT method in Jarvinen et al. [20] combines RSS quantization and an outsourcing protocol with semi-trusted third parties to make an efficient localization scheme for large-scale deployment. These encryption methods take up time and resources to set up, therefore are not easy to implement. Perhaps a more secure but computationally heavy approach is applying the Paillier cryptosystem. It allows for addition operations on encrypted location information against an inference attack. The PILOT method in Jarvinen et al. [44] build upon previous K-anonymity attempts by creating dummy signal strength data that model human mobility behaviour with a Gauss-Markov mobility model. Their work is incomplete as it does not consider indoor physical constraints such as walls. This knowledge can be exploited by an adversary to filter out unrealistic dummy signals. Furthermore, any form of anonymization cannot effectively protect users from inference attacks. It has been proven that auxiliary information can be used to re-identify users. Netflix released an anonymized database of 100 million movie reviews of 500,000 users. In 2008, researchers demonstrated that by linking the data with movie rating from Internet Movie Database (IMDB), a movie database website, 99 of the unique records were identified with 8 movie ratings (allowing 2 to be wrong) and dates that have up to a 14-day error [51].

Spatial obfuscation (or cloaking) reports a different area to the LS than the actual one. The work by [38] has each user work collaboratively by sending their RSS measurements to a chosen leader, which then adds specially adjusted noise to the data before sending it to the LS. The use of the collaboration prevents inference attacks, but also should use a trust system within the network to deter malicious agents.

Randomization of Media Access Control (MAC) addresses consists of sending the LS a fingerprint with frequently changing device identification, to prevent the LS from gathering a history of readings from a single device ID. However, randomization itself is not a simple mechanism. Armengol et al. [17] mention that there are issues with of address collision and network disruptions. In another paper, [52] demonstrate that BLE-based location tracking and analytics are possible even when the MAC addresses are randomized. The trackability is possible due to the low frequency of MAC address changing, and the original information contained in the UUID and the probe request field.

Permutation adds controlled or random noise to the RSS data. A specific use of permutation is used in differential privacy. Differential privacy is a mathematical method of releasing aggregate statistics of a database for analysis without the release of personal information. It satisfies the condition that any sequence of responses to database queries are almost equally likely to occur, regardless of the presence or absence of any individual. There are many algorithms to achieving differential privacy. Their main task is to add random choices and the level of privacy is set by the epsilon \( \epsilon \) parameter. The randomness is determined with a Laplacian or exponential mechanism. The smaller the \( \epsilon \), the better the privacy will be, but since more randomness is added, the accuracy of the output decreases. In [33] the user sends a sample of the AP sequence to the LS. Then, the sequences reference points are grouped into k clusters, and differential privacy is used to mask the real centers of the AP clusters.

### E. Hybrid methods

Privacy is difficult to implement because there is no ideal solution. Table III summarizes the disadvantage of each method. Of the 41 papers, 40% use two or more LPPMs. Eshun et al. [40] develop a system to allow the LS to query the user’s position without them losing their privacy, for example to track employees in a work environment. They assume that both parties are distrusting, therefore it is a secure multi-party computation problem. Their solution is to use a probabilistic data structure called the Spatial Bloom filter (SBF) with an
efficient decision algorithm that is then encrypted using the Paillier cryptosystem. They design a system that allows the user to hide their location from the SP when in a sensitive area. They also include some permutation of the filter so the server cannot reconstruct it after decrypting it. Armengol et al. [17] use two algorithms to reduce the communication overhead of data encrypted with the Paillier cryptosystem. The paper [15] relies on protecting the privacy of the crowdsourcing users providing RSS measurements for the offline fingerprinting phase by receiving their data perturbed with differential privacy and encrypted with the Paillier cryptosystem.

Most of proposed solutions combine two LPPMs to enhance the location privacy. However, there is no winner combination as each author combines different approaches (see Table IV).

### TABLE IV: HYBRID SOLUTIONS

| LPPM 1                | LPPM 2                | Ref.   |
|-----------------------|-----------------------|--------|
| encryption            | encryption            | [32]   |
| encryption            | Hilbert curve         | [46]   |
| encryption            | local calculations    | [50]   |
| k-anonymity           | bloom filter          | [21]   |
| k-anonymity           | differential privacy  | [28]   |
| k-anonymity           | Hilbert curve         | [48]   |
| local calculations    | obfuscation           | [37]   |
| local calculations    | differential privacy  | [18, 30, 39] |
| local calculations    | randomization         | [14]   |
| Paillier cryptosystem | differential privacy  | [13, 15] |
| Paillier cryptosystem | local calculations    | [36]   |
| Paillier cryptosystem | spatial Bloom Filter  | [40]   |
| perturbation          | randomization         | [22]   |

### F. Other approaches

A couple of papers included in the review focus on breaking existing privacy mechanisms. In [9], the PrWFL method is proven to be faulty. A malicious client can fabricate queries to the LS with RSS values set to zero for the APs that are presumed to be far away from the user. The LS does not notice that the query is not genuine because it is encrypted with a Paillier cryptosystem. This way the attacker can extract the entire Wi-Fi fingerprint database from the LS.

Zheng et al. [53] develop a location inference attack using smartphone’s inertial sensors, deploy BLE beacons to obtain the readings and for labelling sensitive areas, and mining techniques for the movement patterns and environment data. Side channel attacks are possible sources of threats to security and privacy. Zhang et al. [35] propose a map as a countermeasure for channel state information-based attacks. They direct a user to a location where Channel State Information (CSI) readings are difficult to analyze. In another study [19], small COTS drones are deployed in an indoor environment to detect and map all present IoT devices. Such information is useful to find rogue devices or tracking personal employee devices which might not be permissible in certain private environments (operation rooms or corporate meetings).

### IV. DISCUSSION AND FUTURE WORK

#### A. Criticism

While the concepts of security and privacy overlap, they are not the same. Data security ensures users that their data is not seen by anyone with unauthorized access. It should be distinguished from data privacy, which is an active method of controlling the access to personally identifiable information. Data that is properly secured through encryption can still reveal a user’s identity by being shared or sold to third parties. For example, a NBC News article from March 7th, 2020 reported that Google sent a notification to a user that police have obtained a warrant to receive all location data from his device on the premise of being at the same time and place as a crime scene under investigation².

#### B. Privacy settings

Many of the papers excluded from this survey rely on device-free positioning by analysing acoustic, infrared, Radio-frequency Identification (RFID), or ultrasound signals to estimate locations. A significant application of this method is in Ambient Assisted Living, where patients require non-invasive motion analysis to infer activeness and physical positions. This demonstrates that avoiding privacy issues altogether is possible in many cases where the purpose of localization is unrelated to personal mobile devices.

The reason that many of these indoor positioning systems are developed is to support indoor LBSs. The survey of all ongoing evolutions of LBSs [54] mentions that some indoor LBSs providers and LS providers are the same entity. It is for these situations the following suggestions will make the most impact. Other LBS applications rely on Google’s fused location Application Programming Interface (API) or one of Apple’s location services, in which only a system override such as the middleware solution would be able to control the information sent from the location request to the LBS.

Different LBS require different kinds of location data; therefore, the level of privacy varies between them. The suggested levels of privacy in Table V show that certain privacy pre-sets might be appropriate based on the type of user. Additionally, a user type can have either a high (H) or a low (L) probability of using the indoor LBS.

²https://www.nbcnews.com/news/us-news/google-tracked-his-bike-ride-past-burglarized-home-made-him-n1151761
The first level of privacy is the most basic one, that is, mostly inherently private. It is assumed that to find things and places in a user’s vicinity, the LBS provider should only require the location data without any other information. If the LBS is collecting user information with permission, it should do so in a fast and secure manner. In a marketing scenario, there is a high trade-off between sharing privacy and motivation. Chances are that companies want to collect fine resolution trajectory data about its customers in return for sales discounts. Many have already employed loyalty reward systems that collect personal data with purchase history and store-level localization. These work on an opt-in basis, which should be carried over to indoor localization scenarios. When customers connect to store Wi-Fi, they should be notified in concise and simple language that their location is being shared with the company. In a fitness tracking scenario, if the user wishes to keep their data private, exact measurements should be kept locally, while if the provider wishes to collect user data for analytics, it should do so in a differentially private manner.

The highest level of privacy is applied to those services that require the most personal data. Social networks have user locations, private conversations, photos, likes, and other highly identifiable data that require complex privacy tools. These need to be applied in IoT environments as well. Emergency services that use localization need trusted servers to securely manage the sensitive data. In research settings, where location and demographics may be collected, differential privacy is suggested.

Generic user privacy profiles based on mobility aspects of users were explored with possible LBS functions that they might require. At first, the probability of the user using a certain LBS function was estimated, then the levels of privacy were applied to those with high probabilities to establish a score. The conclusion of Table VI is that the more mobile a user is, the more they will explore less-known areas and require more functionality and more privacy from LBSs. This hypothesis could be tested in future studies.

### Table V

| Levels of privacy: | Suggested LBS functions | Suggested LPPM |
|--------------------|-------------------------|---------------|
| 1: No personal data is needed in the service. | item retrieval, navigation, POI suggestion | Homomorphic encryption hybrids |
| 2: Personal data adds value to the service. | fitness tracking, marketing | Local calculations, differential privacy |
| 3: Personal data is needed in the service. | research, emergency, social networking | Differential privacy, encryption, trusted servers |

### Table VI

| Profile          | Soc. Network | Marketing | Navigation | Item retrieval | Research | Emergency | Fitness Tracking | POI Suggestion | Total |
|------------------|--------------|-----------|------------|---------------|----------|-----------|-----------------|---------------|-------|
| Tourist          | H,3          | H,2       | H,1        | L,-           | H,3      | H,3       | H,2             | H,1           | 15    |
| Static Worker    | L,-          | L,-       | L,-        | H,1           | L,-      | H,3       | L,-             | L,-           | 4     |
| Moving Worker    | H,3          | H,2       | H,1        | L,-           | H,3      | H,3       | H,2             | H,1           | 15    |
| Student          | H,2          | H,2       | L,-        | L,-           | L,-      | H,3       | H,2             | H,1           | 11    |
| Elderly          | L,-          | L,-       | H,1        | H,1           | L,-      | H,3       | L,-             | L,-           | 5     |

### References

[1] H. Li, L. Sun, H. Zhu, *et al.*, “Achieving privacy preservation in WiFi fingerprint-based localization,” *IEEE INFOCOM*, pp. 2337–2345, 2014.

[2] N. E. Klepeis, W. C. Nelson, W. R. Ott, *et al.*, “The National Human Activity Pattern Survey,” *Lawrence Berkeley National Laboratory*, vol. 11, no. 3, pp. 231–252, 2001.

[3] A. Yassin, Y. Nasser, M. Awad, *et al.*, “Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications,” *IEEE Communications Surveys and Tutorials*, vol. 19, no. 2, pp. 1327–1346, 2017.

[4] R. P. Minch, “Location privacy in the era of the internet of things and big data analytics,” *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 1521–1530, 2015.

[5] A. J. Perez and S. Zeadally, “Privacy Issues and Solutions for Consumer Wearables,” *IT Professional*, vol. 20, no. 4, pp. 46–56, 2018.

[6] Y. A. De Montjoye, C. A. Hidalgo, M. Verleysen, *et al.*, “Unique in the Crowd: The privacy bounds of human mobility,” *Scientific Reports*, vol. 3, pp. 1–5, 2013.

[7] M. Duckham and L. Kulik, “Location privacy and location-aware computing,” in *Dynamic &Mobile GIS: Investigating Change in Space and Time*, 2006, pp. 1–20.

[8] B. Liu, W. Zhou, T. Zhu, *et al.*, “Location Privacy and Its Applications: A Systematic Study,” *IEEE Access*, vol. 6, pp. 17 066–17 624, 2018.

[9] Z. Yang and K. Jarvinen, “The Death and Rebirth of Privacy-Preserving WiFi Fingerprint Localization with Paillier Encryption,” *IEEE INFOCOM*, vol. 2018-April, no. Infocom, pp. 1223–1231, 2018.

[10] A. Liberati, D. G. Altman, J. Tetzlaff, *et al.*, “The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration,” *BMJ*, vol. 339, 2009. eprint: https://www.bmj.com/content/339/bmj.b2700.full.pdf.

[11] R. M. Góes, B. C. Rawlings, N. Recker, *et al.*, “Demonstration of Indoor Location Privacy Enforcement using Obfuscation,” *IFAC-PapersOnLine*, vol. 51, no. 7, pp. 145–151, 2018.

[12] M. Caesar and J. Steffan, “A location privacy analysis of Bluetooth,” in *ACM International Conference Proceeding Series*, 2019.

[13] L. Xiang, B. Li, and B. Li, “Privacy-preserving inference in crowdsourcing systems,” in *2017 IEEE Conference on Communications and Network Security*, CNS 2017, vol. 2017-Janua, 2017, pp. 1–9.

[14] S. Kim, S. G. Yoo, and J. Kim, “Privacy protection mechanism for indoor positioning systems,” *International Journal of Applied Engineering Research*, vol. 12, no. 9, pp. 1982–1986, 2017.

[15] S. Li, H. Li, and L. Sun, “Privacy-preserving crowdsourced site survey in WiFi fingerprint-based localization,” *Earusip Journal on Wireless Communications and Network*, 2019.

[16] M. F. Sadikin and M. Kyas, “IMAKA-Tate: secure and efficient privacy preserving for indoor positioning applications,” *International Journal of Parallel Emergent and Distributed Systems*, vol. 30, no. 6, SI, pp. 447–463, 2015.

[17] P. Armengol, R. Tobkes, K. Akkaya, *et al.*, “Efficient Privacy-Preserving Fingerprint-based Indoor Localization using Crowdsour-
