This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

A Privacy-Preserving Localization Service for Assisted Living Facilities

Francesco Buccafurri[§], *Member, IEEE*, Gianluca Lax, *Member, IEEE*, Serena Nicolazzo, Antonino Nocera
DIIES, University Mediterranea of Reggio Calabria,
Via Graziella, Località Feo di Vito, 89122 Reggio Calabria, Italy
{bucca, lax, s.nicolazzo, a.nocera}@unirc.it

[§]Corresponding Author

Abstract

In this paper, we propose a novel localization service to monitor the position of residents in assisted living facilities. The service supports a configurable balancing between precision and privacy, in such a way that the right of the residents to move freely in the environment in which they live without being tracked is preserved. However, in case of need, they can always be quickly localized. To do this, we implement, on top of an RFID-based architecture, a probabilistic model guaranteeing that the probability of identifying a person in a given (sensitive) place is at most k^{-1}, where k represents the required privacy level. This is obtained by ensuring that the EPC sent by RFID tags is not an identifier, but is equal to that of at least other $k-1$ people, each afferent to a different reader. We show that our method reaches the goal, resisting also attacks aimed at breaking privacy on the basis of humans' movement models. Importantly, privacy is guaranteed against both misuse of the administrator and client-side eavesdropping attacks.

I. Introduction

Assisted Living Facility (ALF) or assisted living residence is a housing facility for people who needs assistance with at least one of the activities of daily living. Residents of an ALF receive health assistance and are monitored by a trained staff to ensure their health, safety, and well-being. In ALFs, the possibility of knowing residents' position is very important, because it may represent a valid support to medical activities and assistance strategies. On the other hand, we are not in the case of classical nursing homes or hospitals, where residents are patients who must be continuously controlled by the staff and have limited freedom of movement. In ALFs, residents pursue their life, move freely, have their private activities, and have social relationships. Apart from specific cases, there is no special medical monitoring equipment that one would find in nursing homes, and the staff may not be fully available at all hours. Thus, we are in an intermediate situation, in which the staff cannot keep a full control on residents because it appears disproportionate w.r.t. the safeguarding of fundamental rights. However, residents' localization cannot be left entirely without monitoring, because in particular cases it is crucial to be able to quickly reach them (e.g., a resident does not show up for an essential therapy).

Therefore, we are in a case in which the trade-off between the utility of having precise information about patients' location and the right of keeping private the exact movements of patients has to be solved. Doing this, we have to take into account that in the specific application context data are strongly sensible and a large effort is required by standards and norms to enforce the prevention of health privacy violations also by abuse of authorized parties.

Which is exactly the problem we have to face? Commercial and research solutions to localize people exist, even based on patents [1], aimed at localizing patients in health assistive environments. However, few of them dealt with privacy violations coming from the administrator of the localization service, even due to data loss or theft. We argue that more advanced solutions of the above trade-off are necessary to make localization systems resistant to patients' privacy threats, which can prejudice human dignity and fundamental rights.

In this paper, we give a contribution to the above need, by proposing a privacy-preserving localization technique resisting also attacks coming from the administrator side. Our solution is based on a probabilistic framework that supports the k-*anonymity* notion [54]. Specifically, our approach leverages an infrastructure of Radio-Frequency IDentification (RFID) readers covering the entire area of the ALF, in which residents are equipped with suitably-designed RFID tags. Each RFID tag sends a quasi-identifier that does not disclose resident's identity because it is always guaranteed that at least other $k-1$ residents send the same quasi-identifier at the same stage. This allows us to guess the position of a resident with probability at most k^{-1}, for any positive integer k. Moreover, the solution guarantees that at least k RFID readers report the same quasi-identifier, so that the privacy requirement k is effective (thus resuming the concept of l-diversity [42]). However, differently from approaches leveraging the t-closeness notion [38], we do not consider the distribution of data values for private attribute, so that the t-closeness privacy requirement is not taken into account.

A probabilistic framework allows us to tune the two parameters of the proposed technique to obtain the trade-off between localization precision and privacy level. Obviously, whenever $k = 1$, our model reduces to existing localization systems (from the side of precision), but maintains privacy guaranteed against both malicious accesses of an intruder to location data and eavesdropping attacks. For $k > 1$, our k-*anonymity* localization adds the strong property of fully protecting privacy also against misuse of system administrators. Observe that, this would be not true if a (trivial) proxy solution grouping

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

together k transmissions at random is applied.

The proposal presented in this work is sound under an attacker model for which the following two high-level characteristics apply: no low-level attacks and no background knowledge (the attack model is detailed in Section V-A). These are common assumptions done in this context (see for example [29], [50]). Under this attack model, our solution preserves privacy against client and server side attacks, by guaranteeing that residents' location is guessed with probability at most k^{-1}, still fully matching the requirement of finding a resident with a small number of attempts. This is an indefeasible feature in an ALF, where a number of events requiring a rapid localization of a resident may occur (e.g., for security reasons, therapy actions, assistive requests, behavior checks, emergency). In our solution, knowing a small number of possible positions of a given resident does not allow us to know his real position for sure, but gives us the possibility to reach him through some additional information, like that obtained from the medical staff belonging to the candidate places, or by means of a physical search.

Finally, we observe that one of the goal of our IT solution is to perform a business service more efficiently and effectively, w.r.t. privacy requirements in a very challenging application scenario, thus reaching an important objective of service computing.

A. Motivation

The aim of this section is to highlight the importance of privacy in the context of ALFs, together with the relevance of the ALF scenario as application setting for innovative IT solutions. This represents a strong motivation of our work, in which we try to investigate both the above aspects in the context of people localization.

As a matter of fact, ALFs are a good example in which a challenging problem is to find the correct balancing between high services and high privacy. Since many years, researchers are facing this problem. See for example [22], in which the project team considered privacy an important aspect of the environment, as prior research found that residents had strong preferences for privacy. Still in this context, [21] classifies ALFs on the basis of the privacy level required by their patients. The results show that an increasing number of ALFs consider patient privacy as a fundamental feature and, therefore, it is included as one of the main aspect in the evaluation and classification of such environments. Moreover, the Assisted Living Facilities Association of America [4], [44] suggests a general interpretation of assisting living to include a philosophy that emphasizes some form of resident independence, autonomy, and privacy, thus recognizing the importance of residents' privacy as related to the dignity of individuals. The impact of privacy requirements and needs on the organization of ALFs is also highlighted in the book "Assisted living: Needs, practices, and policies in residential care for the elderly" by Zimmerman [81]. In this work, the author discusses the case in which the main reason for nursing home residence is not related to health care. In this context, he clinches the importance of including, in the social model of care, humanistic concerns for high order needs among which privacy is one of the most prominent. This trend of considering privacy as a main need in ALFs is also confirmed in [72], in which the privacy need is correlated to the rise of depression and depressive symptoms.

Following the observation above, our approach applied to this application scenario has, therefore, a considerable relevance. The importance of such a scenario is also witnessed by the attention toward it reserved by a lot of other researches in the IT field. Consider, for instance, the survey on IT tools to support ALFs described in [51], in which the authors report that the use of mobile and wearable sensors are becoming pervasive in such environments to improve patient security. Such technologies allows for the implementation of HAR (Human Activity Recognition) modules of fundamental importance in taking care of elderly residents. Another approach showing the importance of the use of IT tools to improve the assistance offered in ALFs is presented in [32]. Here, the benefit of using concepts for home automation environments to take decisions on the basis of variation in the value reported by sensors is proved also by focusing on privacy issues in the storage of patient's data. Other works such as [70], [73], [18] and [31] focus on the problem of improving resident's monitoring in ALFs and start to implement basic solutions to the problem of privacy but mainly focusing on how to safely storing health data and sharing it with clinicians. For instance, [70] describes a new system to handle heterogeneous data from the daily activities of residents in ALF and share it with the different clinicians and hospitals involved. [73] and [18] propose solutions to improving ALFs' residents monitoring by means of WSNs. These approaches leverage the WSN technology to both improve assistance level and proactivity and the satisfaction of basic security requirements. The approach of [31], instead, proposes a multi-agent strategy to collect resident's data from different sources and uses a private cloud to share it with doctors and clinicians so that adapted services can be delivered at any time directly in subjects' environments.

The high number of research efforts towards the improvement of services offered in ALFs together with the lack of refined solutions protecting the privacy of people living in these environments are the premises of this work, whose aim is to add a missing piece in the related scientific literature by proposing a solution to guarantee residents' privacy in ALFs still satisfying monitoring requirements.

B. Plan of the paper

The rest of the paper is organized as follows. Section II discusses about the related work. In Section III, we describe the scenario considered in our study, the addressed problem and the proposed solution. A real-life environment considered as case study to test the performance of our solution is presented in Section IV. In Section V, we introduce the threat model and analyze the security of our technique. Finally, in Section VI, we draw our conclusions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

II. Related Work

It is well known that privacy is a topic that has attracted the attention of the scientific community in several application contexts, such as databases [66], location-based services [19], [74], social networks [8], [9], etc.

In this section, we survey the literature regarding location anonymity, finally focusing on health-care environments.

The concept of k-anonymity was originally designed by Samarati and Sweeney in the field of database privacy [54], [55], [66]. According to them, a database provides k-anonymity if the explicit identifiers of all the tuples are removed from the database and, additionally, the quasi-identifiers (set of attributes leaking confidential information) of each individual in the database cannot be distinguished from those of at least $k-1$ other individuals. Hence, this approach of k-anonymity suggests the suppression and generalization of quasi-identifiers. Suppression consists in protecting sensitive information by removing it, whereas generalization involves replacing specific values (e.g., residence address) with more general ones (e.g., only ward or district of residence). Both techniques result in information loss. Always within the context of database privacy, the authors of [62] propose a probabilistic k-anonymity property, which guarantees that the probability of re-identification of a database record be the same as in k-anonymity (at most 1/k), but without explicitly requiring that the quasi-identifier attributes take identical values within each group of k records. Moreover, two computational methods, based on microaggregation and swapping, are presented to reach probabilistic k-anonymity. Our approach does not work on database tables, but it applies a *k-anonymity* localization technique, thus dealing with the issue of preventing location-based identity inference.

The general idea of k-anonymity localization [19], [74], [11] is that the position of a user is given provided that the probability of identifying him is less than k^{-1}. A similar approach to protect location data consists in creating areas of confusion where the traces of a number of users converge [46], [56], [37], [7]. Over time a wide range of application scenarios took advantage of this technique to guarantee privacy. For instance, in the context of location-based services (LBSs), k-anonymity location requires that location information inside a message sent from a mobile user to an LBS should be indistinguishable from at least $k-1$ other messages from different mobile nodes [20]. The most popular solution for designing privacy-preserving LBSs consists in obfuscating the actual location of the query by constructing cloaking regions that contain the locations of k anonymous users [13]. Another privacy-preserving mechanisms is presented in [12]. In this paper, the authors argue that the ability of generating fake contextual data can be important also in privacy preserving applications. For instance, it could be possible to add some dummy queries or fake transactions, which are indistinguishable from the real ones, so that it is hard to trace the real interests of users performing them.

Other related approaches are based on the notion of l-diversity and its extensions [60], [80], [77], a form of group based anonymization in data sets that performs a reduction of the granularity of a data representation to reach the privacy goal.

Still in the context of group based anonymization, there are approaches based on t-closeness [38], [52], [39]. The t-closeness privacy notion refines l-diversity considering the distribution of data values for the attribute. In particular, it requires that the distribution of a sensitive attribute in any equivalence class is similar to the distribution of the attribute in the overall table. t-closeness improves l-diversity because the latter ensures diversity of sensitive values in each equivalence class, without taking into account the semantical closeness of these values. These privacy notions are attempts to contrast attacks that strongly leverage the knowledge of data distribution (e.g., information on residents' habits and the knowledge on the intended use of each area inside the ALF). However, throughout this paper, as commonly done in the literature, we assume that no attacks based on this knowledge are carried out (see, for example, [29], [50], [30], [23],.

Several solutions have been proposed for indoor localization [33], [35], [41], [48], however few attention has been given to privacy issue. Only the proposal in [29] presents a system to provide indoor continuous localization to a mobile user by measuring the signal intensity of its surrounding Wi-Fi APs, with minimum energy consumption, so that a static cloud-based server exploited for localization cannot identify user's location with a probability higher than $1/k$. With respect to our solution, the proposal in [29] is based on the use of smartphones (instead of active tags) and Wi-Fi (instead of RFID technology). This solution, which has been proposed for a problem different from that addressed in this paper, appears little adequate for our application context because: (1) the collaboration of residents is required as they have to constantly bring a smartphone with them, and (2) the necessity of keeping all Wi-Fi APs always active could be little compliant with health-care facility restrictions.

The authors of [59] discuss the problem of exchanging location information amongst untrusted parties and present a solution to guarantee user's privacy. Another solution to the same problem, based on a public-key privacy homomorphism, is presented in [58]. Despite the fact that both the above solutions apply only to phone-cell-based LBSs (whereas our approach considers more fine-grained people localization), an interesting common aspect of the above papers with ours is that both do not rely on a trusted third party to anonymize the users and to guarantee their location privacy.

A new trend related to LBSs is represented by location-based social networks (LBSNs) [79], [76]. In this context, the authors of [79] formalize a framework called iGeoRec to predict the probability of a user visiting a new location through a probabilistic approach, whereas in [76] the same aim is obtained by a spatial social union approach. Still in the context of k-anonymity localization, an extension for spatio-temporal data is the (k,δ)-anonymity [2], [3], which is specifically designed for uncertain trajectories defined as the movement of an object on the surface of the Earth. This technique exploits the spatial uncertainty $\delta \geq 0$ in the trajectory recording process. In [67], the authors prove that, for any $\delta > 0$ (that is, whenever there is actual uncertainty), (k,δ)-anonymity does

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

not hide an original trajectory in a set of k indistinguishable anonymized trajectories.

Clearly, the general idea of obfuscating/confusing the final position of users to protect their privacy, is not suitable in our scenario because we need to *exactly* identify the k possible locations of a user to face emergencies. Location cloaking [10], [58], [24], [27], [36] aims to perturb location data by introducing random noise in order to guarantee user's privacy. To maintain the current locations of all users, a trusted third-party anonymizer can be employed [17], [20], [28]. It generates a Cloaking Region (CR) enclosing k users who are close to each other. However, the major drawback of these approaches is that they do not prevent from attacks on the trusted third-party. Indeed, if an attacker gains access to it, the privacy of all users is compromised. Our approach, instead, is conceived to operate with no trusted third-party.

Generally, different applications imply different privacy threats and models. For instance, in mobile phones, there are LBS applications using Global Positioning System (GPS) [16], or Wireless Networks [29]. There are LBS applications in which the user's location can be continuously detected, see for example [63], [2], also providing a user-defined level of privacy against a static server (see [29]).

In an RFID scenario, continuous detection is unlikely, because high computational efficiency and low cost are typically required. In this context, [78] presents an approach to improve the efficiency of the RFID system. An approach to tracking children in a large park still preserving their privacy is presented in [40]. Through a deep security analysis the authors show that their tracking scheme guarantees children identity privacy, unlinkable location privacy, and forward security. However, this approach does not contrast server side attacks. This limitation is not present in our approach.

The problem of balancing (1) the need of protecting health information and (2) the utility of sharing information, has received great attention in the last years [43], [71].

RFID technology and data anonymization have been widely adopted in health-care environments [25], [69], [75], [65], [14]. In particular, the authors of [69] present four lists of applications for RFIDs enabling functions (tracking, identification and authentication, automatic data collection and transfer, and sensing) and subjects (staff, patients, assets and clinical trials). They conclude that adopting RFID sensory systems to track the positions of patients, doctors, medical equipments, and devices inside a hospital, can help minimize medical errors and improve the management of patients and resources. According to the results presented in [69], our paper proposes a practical solution to the problem of localizing people in an assistive environment, to reduce risks related to hospitalization and to increase patient independence. This feature, together with the others described above (i.e., no third trusted party used, no obfuscation, administrator-side attack prevention), makes our approach relevant and novel with respect to the state of the art.

III. Privacy-preserving localization

In this section, we describe the scenario considered in our study and the proposed solution. We consider an assisted living facility in which the position of residents has to be monitored. The approach we follow to address this problem is based on the use of active RFID tags, RFID readers, and a server.

a, d	system parameters
p	number of residents
p'	the lowest prime number $\geq p$
r	number of readers
RPF	random permutation function

TABLE I
Notations

Preliminarily, we introduce the notations used in the following (they are summarized in Table I). Our solution is parametric w.r.t. two numbers d and a: d is a positive integer and a is a real number in the interval $[0, 1]$ (their role will be discussed in Section V-B). Let p be the number of residents to monitor and r be the number of RFID readers present in the environment. *RPF* is a *random permutation function* operating on the set of integers $[1, p' - 1]$, defined as:

$$RPF : \mathbb{Z}^*_{p'} \rightarrow \mathbb{Z}^*_{p'}$$

where p' is the lowest prime number such that $p' > p$, $\mathbb{Z}^*_{p'}$ is the multiplicative group of $\mathbb{Z}_{p'}$, $\mathbb{Z}_{p'}$ is the set of (equivalent classes) of integers $(\mathbf{mod}\ p')$, and $RPF(i) = i \cdot g\ (\mathbf{mod}\ p')$, for any $g \in \{1, \ldots, p'\}$. Observe that being p' prime, RPF works as a permutation because $\mathbb{Z}^*_{p'}$ is an additive cyclic group, and every $i \in \{1, \ldots, p'\}$ is a generator for $\mathbb{Z}^*_{p'}$.

We are ready to present how our proposal works. We start from the infrastructure. The first operation is to partition the ALF into *cells* by properly positioning the RFID readers in known locations. Each resident is equipped with an active RFID tag that replies to the reader requests sending a suitable number said QID (quasi-identifier). Finally, the readers communicate with a server, which handles data flow on the network.

The idea is that the QID generated by each RFID tag changes at each reading in a pseudo-random way. As a consequence, a malicious attempt to track a resident by sniffing the generated number fails, if the attacker cannot associate the generated QID with the victim. To protect user privacy, we remove any one-to-one correspondence between QIDs and residents, which is replaced by a one-to-many correspondence. However, to preserve the localization service, we require that the above correspondence guarantees the localization of a resident with a probability at most k^{-1}. In other words, our approach implements a mechanism able to guarantee, at any time, that there exist at least k different positions in which a resident could be.

The workflow of our protocol to monitor residents is iterative and, in the following, we describe what happens at the iteration $t > 0$.

1) *Query*. The server sends a *reading* request to all the r readers, which, in turn, query the tags inside their coverage area. To prevent tag reading from unauthorized devices, reader authentication is adopted. Among the several authentication techniques presented in the

literature [49], typically using classical cryptographic primitives such as hash functions or block ciphers, in our proposal we adopted the minimalist lightweight authentication protocol proposed in [47], which is low-cost, offers an adequate security level, and can be implemented even in the most limited tags as it only needs around 300 gates.

2) *QID generation*. Each RFID tag contains a 64-bit pseudo-random number generator (PRNG), whose seed is known by the server, and is able to compute the random permutation function RPF defined above. To reply the request, the tag computes $ID(t) = RPF\big(ID(t-1)\big)$, where we assume that $ID(0) = i$ for the tag of the i-th resident. In words, at the beginning ($t = 0$), each resident is identified by a progressive integer $1, \ldots, p$ and, after each iteration, it is associated with a new integer in $[1, p]$. The random permutation function assures that, at each iteration, two different tags are not associated with the same integer. Then, the tag computes QID at the iteration t as:

$$QID(t) = \begin{cases} ID(t) \bmod d & \text{if } PRNG(t) \leq a \cdot 2^{64} \\ PRNG(t) \bmod d & \text{otherwise} \end{cases}$$

where, we recall, a and d are system parameters and $PRNG(t)$ is the t-th element of the pseudo-random sequence of the tag. In words, with probability a, the tag returns a hash obtained by uniformly mapping $ID(t)$ to a domain of d elements, whereas, with probability $(1-a)$, $QID(t)$ is pseudo-randomly generated.

3) *Response*. The j-th reader (with $1 \leq j \leq r$) processes the received data, say $\{QID_1, \ldots, QID_z\}$ (i.e., there are z people under the reader coverage area). Let us denote by $v \leftarrow X$ the operation of choosing an element v from the set X according to the uniform random distribution on X. For each received QID_i, the tuple $\langle QID_i, x\rangle$ is sent to the server, where x is a positive integer obtained as:

$$x = \begin{cases} v \leftarrow \{1, \ldots, r\} \setminus \{j\} & \text{if } \exists\ w < i : QID_w = QID_i \\ j & \text{otherwise} \end{cases}$$

In words, if more tags generate the same QID, only the first is associated with j: the remaining ones are associated with other randomly chosen readers. Otherwise, if a QID is generated by only one tag, then this tag is associated with this reader (i.e., j).

4) *Storing*. The server collects the tuple received from all readers. If the number of the received tuples is lower than the number of residents, the service detects that a resident has left the ALF and throws an alert (as we will show in Section V-B, the system will continue to run correctly without need of reconfiguration). Otherwise, these tuples are stored overwriting the tuples at the previous iteration.

System parameter	Value
a	1
d	2
p	4
p'	5
r	100
t	1

TABLE II
SYSTEM PARAMETER SETTINGS FOR THE RUNNING EXAMPLE

The protocol described above gives us the possibility to locate a resident and to verify that no resident leaves the ALF.

The procedure to locate a resident u is as follows. Let's assume we are at the t-th iteration. The server can obtain the $QID(t)$ generated by u, as it can compute $PRNG(t)$ and $ID(t)$ of step *QID generation*. Then, the server performs a *reading* request to all the r readers and among all the tuples received, it filters out the set $T = \{\langle QID(t), l_1\rangle, \ldots, \langle QID(t), l_{|T|}\rangle\}$ (i.e., those referring to $QID(t)$). Now, the server can guess the location of u with probability $|T|^{-1}$: indeed, the possible $|T|$ locations of u are inside the coverage areas of the readers $l_1, \ldots, l_{|T|}$.

Concerning the procedure to guarantee that residents are confined inside the ALF, it works as follows. Periodically, the server executes a *reading* request to the readers and checks that all the p QIDs from the residents' tags are received. When the number of received QIDs is less than the expected one, an alert is generated. Moreover, from the knowledge of the expected QID, it is possible to guess (with a given probability) who is the absent resident and his/her last possible (k) positions (indeed, the server stores the last tuples received). Observe that the frequency at which tags send the signal represents an important privacy issue because a high sending frequency could allow an attacker to track residents' movements. We study this aspect in Section V-B.

A. Running Example

In this section, we sketch a simple example to show how the whole protocol works and the messages exchange among the actors.

As for the parameter settings, we choose $a = 1$, so that $QID(t)$ will be deterministically generated (see Section III). Moreover, we set the number of possible different $QIDs$ $d = 2$. The number of monitored persons p is set equal to 4 and, therefore, $p' = 5$ (i.e., the first prime number such that $p' > p$), whereas the number of readers is equal to 100. For the sake of simplicity, we omit the discussion about how the system parameters are set on the basis of the privacy requirements. In Table II, we report a summary of the settings chosen for our parameters.

In Fig. 1, we illustrate the messages exchange among the actors from the initial state ($t = 0$) to the first iteration ($t = 1$). In the first step performed by our protocol, the server sends the message *reading request* to all the 100 readers (step *Query* of Section III). This starts the iteration $t = 1$, which we are considering in this example. All the readers carry out authentication with the RFID tags within their coverage area

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

Patients	P_1	P_2	P_3	P_4
$ID(0)$	1	2	3	4
$ID(1)$	3	1	4	2
$QID(1)$	1	1	0	0

TABLE III
AN EXAMPLE OF THE TRANSFORMATION OF 4 TAG IDENTIFIERS.

and forward them the *reading request*. At this point, all RFID tags perform the step *QID generation*. Table III-A helps us to understand how $QIDs$ are generated by each monitored person: the second row of the table reports the initial value of ID associated with each person (i.e., $ID(0)$), whereas the third row reports the result of the application of the random permutation function (RPF) from step $t = 0$ to step $t = 1$, by assuming $g = 3$. The last row, instead, shows the final QID values for step $t = 1$ obtained as $ID(1)$ mod 2.

After this computation, every RFID tag sends the obtained QID to the RFID reader from which the *reading request* come. The readers collect the received $QIDs$ and process them. In particular, we assume that P_3 is under the coverage area of the reader R_1, P_4 is under the reader R_2, whereas P_1 and P_2 are under the reader R_3. Thus, both the readers R_1 and R_2 receive only one QID (from P_3 and P_4, respectively) and they send to the server the tuple $\langle QID, x\rangle$, where the first element is the QID received from the tag (0 in both cases) and the second element is the index of the reader (i.e., 1 and 2, respectively). As for the reader R_3, it receives the same QID from P_1 and P_2. For the first QID, the tuple $\langle 1, 3\rangle$ is sent to the server, whereas the second QID is not directly sent to the server. Indeed, according to the step *Response*, in case of collision of some QID on the same reader, as it happens for P_1 and P_2, only one of them is associated with this reader, whereas the remaining $QIDs$ will be mapped to other readers. In particular, a random reader is selected (we assume it is the reader R_{87} in our example) and the second QID is associated with this reader, in such a way to hide that more users with the same QID are in the same place (this could result in the violation of the privacy requirement). In practice, P_1 will be associated with reader R_3, whereas P_3 will be virtually mapped on the reader R_{87}. Finally, the server stores all the received records and checks that all expected QID are really received.

IV. CASE STUDY

In this section, we describe the real-life scenario we considered in our experimental evaluation of the performance and security features of our approach. For this purpose, we referred to an existing ALF and built a simulator based on its physical characteristics. Specifically, the considered real-life ALF stretches over about 50k square meters and hosts a maximum number of 500 monitored residents. We designed our solution by computing the number of readers to cover all the different areas occurring in this facility and their exact position therein. Observe that the position of each reader depends on the power transmission level and the read distance has to guarantee that each portion of the place is covered by at least one reader. Moreover, to face problems related to interference and overlapping coverage areas, for the spatial distribution of readers, we used the approaches proposed in [57] and [68]. This solves the problem arising when a tag is located within the range of two RFID readers, which might be falsely detected in two different places at the same time. Moreover, in our design, we considered also the impact of multipath, reader-to-tag interference and forward link fading in the estimation of reader distances by leveraging the study described in [6] and the methodology described in [34].

To satisfy all the requirements of the techniques above for reader distribution, we have that the number of readers necessary to cover the entire area is 2000. Concerning the use of the RFID technology, RFID active tags and readers use an operating frequency of 433 MHz, which can be safely used also for healthcare applications. This is an important property as some of the ALF residents may suffer from important pathologies requiring severe constraints for the electrical equipments used for their care. The (adjustable) read range of such tags has been set to cover an area of about 40 meters (i.e., about a room). Each reader can detect hundreds of tags in few seconds and based on the signal strength received, the reader reports or ignores the received EPC to avoid multiple reading of the same tag (see Section III). Tags have small size and are attached to person's wrist or ankle using standard ID straps.

This real-life scenario will be used in the next section as test bed for our experiments aimed to validate our proposal.

V. SECURITY MODEL

In this section, we describe the security model and analyze the security properties of our proposal.

A. Attack Model

As usual in this context, we realistically assume that a solution satisfying the privacy requirement exists. Consequently, we do not consider inadmissible cases, for example with very few monitored people or in which k is too high with respect to the number of residents and the possible locations.

Under this basilar assumption, we state the security properties of our protocol. We recall that we aim to obtain *k-anonymity*, meaning that the probability of localizing a person is k^{-1}, where k is a given anonymity requirement. Our approach reaches this goal by forcing that more people generate the same QID.

We identify the following actors in our scenario:

- **service provider**, the entity that implements the RFID-based solution in the environment (i.e., installs and configures the readers, wires the connections among readers, provides the monitoring software, etc.).
- **residents**, who have to be monitored and whose privacy has to be preserved.
- **(system) administrators**, who can access all data produced by the service;
- **unmonitored people**, who are present in the environment but are not monitored (e.g., nurses or visitors).

Concerning the security features of our proposal, we identify the following properties that our approach must satisfy:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

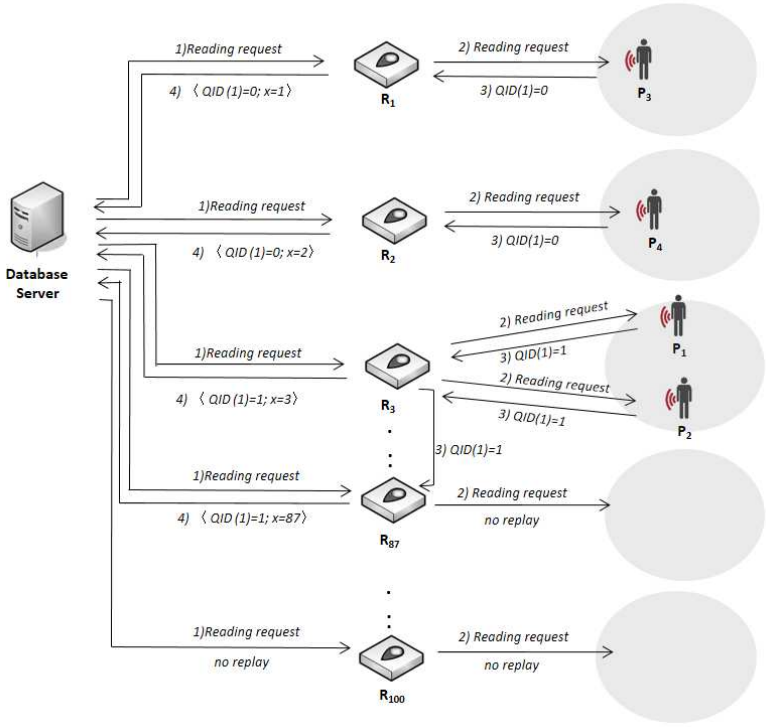

Fig. 1. An example of protocol instantiation.

1) **SP1.** Given a QID q, at least k different readers report the presence of a resident sending q.
2) **SP2.** Resistance to client-side attacks.
3) **SP3.** Resistance to server-side attacks.
4) **SP4.** Resistance to attacks based on fake tags or fake queries.

To analyze the security properties above, our threat model includes the following assumptions. Observe that these assumptions are typically adopted in this contest (see, for example, [29], [50]) and that the security analysis is done w.r.t. a parameter x (see below).

1) no low-level attacks. We assume no successful attack on tag/reader authentication and no communication tampering (the adversary can learn from whatever data is available on the system, such as messages exchanged, files, etc.).
2) no background knowledge. As shown in [45], any background information can breach privacy guarantee, so that this assumption is commonly adopted in the literature [29], [61], [30], [23]. In particular, we assume that the attacker has not any background knowledge about buildings (e.g., number of users in a building or area) or user movement. However, in our attack model, we include the case in which the attacker knows the position of at most x residents.
3) the association between RFID tag and resident cannot be compromised.

Observe that, because the compliance with the required security properties is strongly scenario-dependent, it would be infeasible to analytically prove it. Therefore, we will address this issue experimentally by simulating the application of the proposal to the ALF described in Section IV[1]. TIn our experiments, to simulate resident shifts inside a given area, we built a prototype implementing random-based mobility models described in [5]. The default model is that residents move randomly in the whole area.

B. Security Analysis

In this section, we study the security properties of our approach on the basis of the attack model described above. Recall that our target is to guarantee the privacy requirement k. First, we observe that our method is ϵ-approximate, because the privacy requirement is guaranteed with probability $1-\epsilon$, where ϵ is a small positive real. Moreover, the security is analyzed w.r.t. the parameter x defined earlier in Assumption 8. To do this, we just increase the privacy parameter k to the value $k+x$. Indeed, from the knowledge of the position of x residents, the attacker can guess the position of other residents with probability equal to $(k-x)^{-1}$ (thus the initial privacy requirement is satisfied).

Compliance with SP1 *(i.e., given a QID q, at least k different readers report the presence of a resident sending q).* The next theorem proves that our proposal guarantees the security property **SP1** described in Section V-A, i.e., given a QID q, it guarantees that at least k different readers report the presence of a resident sending q.

Theorem 1: Given an admissible privacy requirement k, there exist at least one configuration of the parameters d and a

[1] As $p = 500$, in our experiments we set $p' = 503$, because p' is the lowest prime number higher than p.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

such that k-anonymity is guaranteed with probability greater than $1-\epsilon$, for any $0<\epsilon<1$.

Proof. We need to compute two probabilities. For the first one, we introduce the random variable $\mathcal{C}_1$ defined as follows: given a QID q, $\mathcal{C}_1$ is the total number of residents who, at a given time t, generate q as QID. We call *number of QID-collisions* the value returned by $\mathcal{C}_1$. We study the probability $\mathbf{p_1}(\mathbf{c_1}) = \mathcal{P}(\mathcal{C}_1 = c_1)$, which is the probability to have exactly c_1 QID-collisions, with $0 \leq c_1 \leq p$.

First, consider the case $a=1$. We obtain that $\mathbf{p_1}(\mathbf{c_1}) = 1$ if $c_1 = \frac{p}{d}$, $\mathbf{p_1}(\mathbf{c_1}) = 0$ otherwise. Indeed, when $a=1$, $QID(t)$ is equal to the deterministic value $\widetilde{QID}(t) \bmod d$ and the function RPF, which introduces pseudo-random permutations, guarantees that the possible QIDs are uniformly distributed among residents.

Then, consider the case $a=0$. Now, the probability of having c_1 QID-collisions is: $\mathbf{p_1}(\mathbf{c_1}) = \frac{(d-1)^{p-c_1}\frac{p!}{(p-c_1)!}}{d^p}$. Indeed, d^p is the number of possible assignments of QIDs and $(d-1)^{p-c_1}\frac{p!}{(p-c_1)!}$ is the number of cases producing c_1 QID-collisions. Moreover, $\mathbf{p_1}(\mathbf{c_1}) = \frac{1}{d^{p-c_1}}\frac{1}{d^{c_1}}(d-1)^{p-c_1}\frac{p!}{(p-c_1)!} = \frac{p!}{(p-c_1)!}(\frac{1}{d})^{c_1}(1-\frac{1}{d})^{p-c_1}$.

By combining the two cases above, we obtain:

$$\mathbf{p_1}(\mathbf{c_1}) = \begin{cases} a + (1-a)\frac{p!}{(p-c_1)!}(\frac{1}{d})^{c_1}(1-\frac{1}{d})^{p-c_1} & \textit{if } c_1 = \frac{p}{d} \\ (1-a)\frac{p!}{(p-c_1)!}(\frac{1}{d})^{c_1}(1-\frac{1}{d})^{p-c_1} & \textit{otherwise} \end{cases}$$

The second probability we have to compute is related to the case in which, among the c_1 colliding $QIDs$, some of them are under the same reader. If two $QIDs$ collide under the same reader, the corresponding collision is called *local*. Let us denote by C the random variable returning the total number of local collisions (observe that $C \leq C_1$). Let $\mathbf{p(c)}$ be $\mathcal{P}(\mathcal{C} = c \mid \mathcal{C}_1 = c_1)$, which is the probability to have exactly c local-collisions given that there are c_1 QID-collisions. We have that:

$$\mathbf{p(c)} = \frac{\sum_{i=1}^{c}(i^c - (i-1)^c)\begin{pmatrix} r \\ i \end{pmatrix}\begin{pmatrix} c_1 \\ c \end{pmatrix}\begin{pmatrix} r-i \\ c_1-c \end{pmatrix}}{r^{c_1}}$$

where the denominator is the total number of ways to spread the c_1 collision over the r readers, and the numerator is obtained by considering that, for any i readers including the c collisions, we have to add the collision assignments not already considered in previous step (this explains the term $(i^c-(i-1)^c)$), and the number of possible ways to have c collisions among c_1 collisions. Finally, we have to consider all the possible ways in which the remaining c_1-c collisions are spread over the remaining $r-i$ readers, that is $\begin{pmatrix} r-i \\ c_1-c \end{pmatrix}$.

In reaction to local collisions, readers use the function R (recall step *Response* in Section III) to spread all (but the first one) colliding $QIDs$ over other readers. At this point, it could happen that a reader x receives a QID colliding with a QID generated by a resident located under the coverage area of x. This kind of collision is called *reader*-collision. Whereas QID-collisions increase privacy, reader-collisions are disadvantageous because more residents are associated with the same reader, thus reducing the number of possible locations for these residents.

To study the second probability, we introduce the random variable $\mathcal{C}_2$ defined as the total number of reader-collisions occurring after the execution of the function R. Clearly, $C_2 \leq C$. Therefore, we study the conditional probability $\mathbf{p_2}(\mathbf{c_2}, \mathbf{c_1}) = \mathcal{P}\big(\mathcal{C}_2 = c_2 \mid \mathcal{C}_1 = c_1\big)$, which is the probability of having c_2 reader-collisions given that there are c_1 residents with the same QID. Denoting by R_A the set of readers associated with residents generating the same QID and by R_B the remaining ones, we have: $\mathbf{p_2}(\mathbf{c_2}, \mathbf{c_1}) = \sum_{i=0}^{c_2}\Big(\sum_{c=0}^{c_1}\mathbf{p(c)}\big(\mathbf{p_{A2}}(\mathbf{c_1}, \mathbf{c}, \mathbf{i}) + \mathbf{p_{2B}}(\mathbf{c2}, \mathbf{c1}, \mathbf{c}, \mathbf{i})\big)\Big)$, where $\mathbf{p_{2A}}(\mathbf{c_1}, \mathbf{c}, \mathbf{i}) = \mathcal{P}\big(\mathcal{C}_2 = i \mid (\mathcal{C}_1 = c_1 \wedge \mathcal{C} = c)\big)$ is the probability to have i cases in which the function R returns a reader in R_A (thus generating reader-collisions), whereas $\mathbf{p_{2B}}(\mathbf{c2}, \mathbf{c1}, \mathbf{c}, \mathbf{i}) = \mathcal{P}\big(\mathcal{C}_2 = c_2 - i \mid (\mathcal{C}_1 = c_1 \wedge \mathcal{C} = c)\big)$ is the probability to have $c_2 - i$ cases in which a reader-collision is obtained because the function R returns the same reader among those in R_B. The sum considers all the possible cases generating c_2 reader-collisions. $\mathbf{p_{2A}}(\mathbf{c_1}, \mathbf{c}, \mathbf{i}) = i \cdot \frac{c_1-c+1}{r-1}$ because we run i random generations, $|R_A| = c_1 - c + 1$, and the remaining readers are $r-1$. By following the same reasoning as $\mathbf{p_1}$ in the case $a=0$, we obtain:

$$\mathbf{p_{2B}}(\mathbf{c_2}, \mathbf{c_1}, \mathbf{c}, \mathbf{i}) = \frac{\Big(r-(c_1-c+1)-1\Big)^{c-(c_2-1)}\frac{c!}{(c-(c_2-i))!}}{r-(c_1-c+1)}.$$

The problem we have to solve is to find d and a such that, for a given (small) ϵ, the following inequality holds:

$$\sum_{c_1=k}^{p}\Big(\mathbf{p_1}(\mathbf{c_1})\sum_{c_2=0}^{c_1-k}\mathbf{p_2}(\mathbf{c_2}, \mathbf{c_1})\Big) \geq 1-\epsilon$$

In words, we require at least k residents generating the same QID (P_1). Then, if we have $k+x$ colliding residents, we accept also cases in which at most x are in the same place (the sum on P_2 deals with this possibility). By satisfying the above equation, we guarantee with probability lower bounded by $1-\epsilon$ that at least k residents with the same QID in k different places occur.

To conclude the proof, we show that, if a solution satisfying the privacy requirements exists (as stated in the hypothesis of the theorem), then at least one setting of d and a satisfying the above inequality exists. Indeed, the trivial solution $d=1, a=1$ solves the above problem because all residents have the same QID. However, in this case, no meaningful information about the location of any resident is provided. As a consequence, we need to find the greatest d satisfying the above property. This number can be found by a guess-and-check iterative method, which starts from the minimum value $d=2$ and at each iteration increases by 1 the value of d (as we will see below, the choice of a is related to some security considerations that suggest us to set $a = (1-x^{-1})$, where x

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

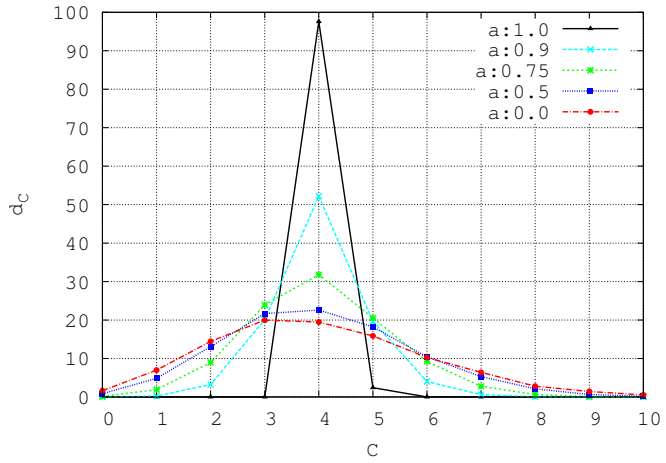

Fig. 2. Number of residents with the same QID versus d_C for different values of a.

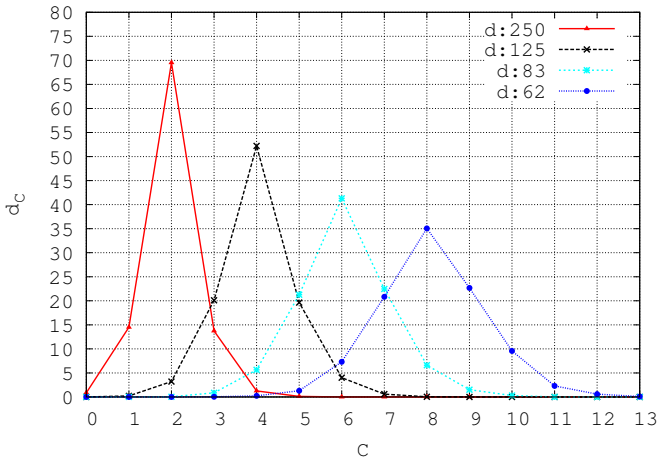

Fig. 3. Number of residents with the same QID versus d_C for different values of d.

is a suitable positive integer). □

Compliance with SP2 *(i.e., resistance to client-side attacks).* In this class of attacks, we have to consider the case in which a client can sniff messages exchanged in the environment. In this way, an attacker may knows the QID generated by a resident (for example, by eavesdropping the communication between tag and reader) and wants to guess who is the resident located near that reader at that time. However, residents' QIDs are dynamic and change at each reading in a pseudo-random way, so that no correlation between two subsequent readings should be guessed. Observe that, setting $a = 1$ means that the QID generation of each tag depends only on the random permutation function that guarantees a uniform distribution of QIDs and a number of collisions deterministically fixed. However, a side effect of setting $a = 1$ is that the sequence of QIDs generated by the same tag is periodic with period equal to the number of patients. An attacker may exploit this periodicity to know the next QIDs that a patient will generate and, thus, may try to track patient movements. The parameter a is used to prevent such an attack by randomly changing some elements of the deterministic QID sequence in such a way to break this periodicity. The parameter a measures the probability that an element of the deterministic QID sequence is not changed.

It is, then, obvious that the possibility to successfully carry out this attack depends on the values of the parameters of the system; therefore, by tuning such parameters, we can reduce the probability of success for this attack to any desired (low) value.

For this reason, we carried out several experiments to study the behavior of our technique for different values of system parameters. We start by studying the parameter a. In this first experiment, we choose five different values for a (i.e., 0.0, 0.1, 0.25, 0.5 and 1) and we measure the number C of colliding residents against the percentage d_C of the QIDs for which we observe C collisions. We set the parameter d (i.e., the number of possible QIDs) to $p/4$. In Fig. 2, we report the results of this experiment.

From the analysis of this figure, we observe that the peak of each curve decreases if a assumes low values. In particular, if a is equal to 0, then our approach assigns QIDs in a pseudo-random way, according to the generator PRNG described in Section III. For this reason, the trend of d_C is very smooth and the standard deviation of the number of collisions is high. By contrast, if a is equal to 1, then the QID generation is deterministic and, hence, we observe a very peaked trend of d_C for $C = 4$. Therefore, we should set a to the highest possible value allowing our approach to resist attacks based on the periodicity of the QIDs generation. If we assume that changing one element every x elements of the deterministic sequence is sufficient to avoid a periodicity-based attack, we should set $a = (1 - x^{-1})$.

In the second experiment, we test the behavior of our approach for different values of the parameter d (i.e., $p/2$, $p/4$, $p/6$, and $p/8$), which identifies the number of different QIDs generated by residents. We set the value of the parameter a to 0.9. The result of this experiment is reported in Fig. 3.

The discrete-Gaussian-like curves associated with the different values of d show decreasing height of the peak as d decreases, whereas the width of the bell (and, hence, the standard deviation) behaves the opposite. This can be explained by considering the fact that a lower value of d implies a greater number of users who generate the same QID. However, due to the presence of the parameter a, which represents the probability of assigning a QID different from that generated deterministically, the lower value of d also implies a greater number of residents that could have assigned a different QID w.r.t that assigned deterministically. These residents will decrease the number of collisions for the QID they should have assigned deterministically and will increase the average number of collisions for the other QIDs. This causes the reduction of the height of the curve's peak and the increment of its standard deviation.

In the next experiment, we consider another issue related to the dimensioning of the solution: indeed, a system parameter is related to the maximum expected number of residents (which has been considered equal to 500 in the previous experiments). In a real-life scenario, it is possible that the actual number

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

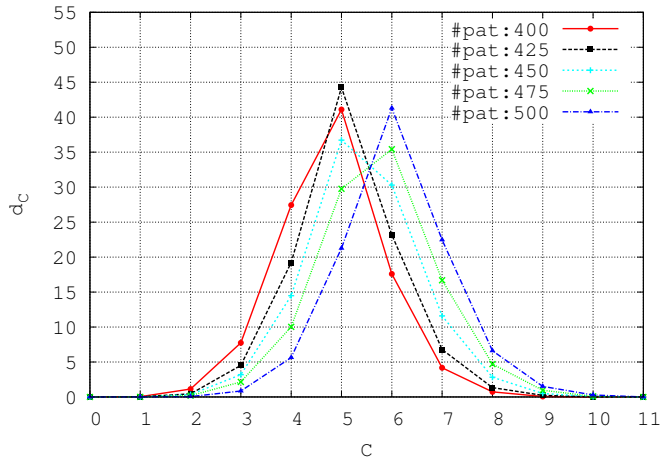

Fig. 4. Number of residents with the same QID versus d_C for different number of residents ($d = 83$).

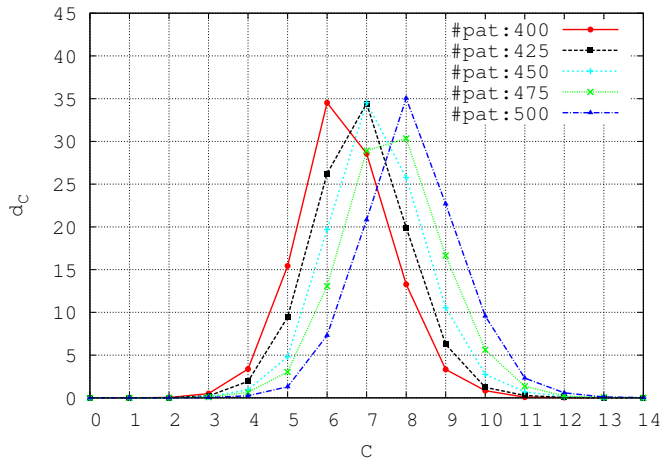

Fig. 5. Number of residents with the same QID versus d_C for different number of residents ($d = 62$).

of residents in the environment is less than the maximum expected one, because of the necessary oversizing during the system design. However, once the RFID tags have been configured, the tag reconfiguration every time a resident arrives or leaves the ALF could be too expensive. For this reason, we study the system performances when the number of residents is less than the expected one.

Fig. 4 shows again the number C of colliding residents when the overall number of residents is decreased up to 20%. In this run, we set $a = 0.9$ and $d = 83$. Now, if we assume that the anonymity requirement is $k = 4$, then the anonymity requirement is not guaranteed in the worst case. This problem can be solved by reducing the value of d. A value of d reaching this goal can be found by applying a simple guess-and-check iterative method. For example, by changing the value d to 62, we obtain the results reported in Fig. 5. With this new configuration, the anonymity requirement $k = 4$ is guaranteed with all the possible number of residents considered in this experiment.

Now, we compare the effectiveness of our technique in

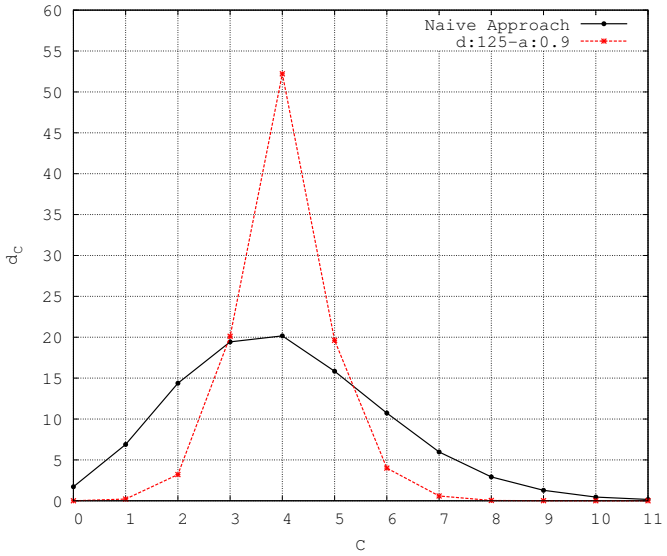

Fig. 6. Number of residents with the same QID versus d_C for the two techniques.

localizing a resident w.r.t. another approach that we call *naive*. In this approach, the tag of each resident is identified by a unique static ID. When a request for the resident with $ID = x$ is sent by readers, all tags act as follows. The tag with $ID = x$ replies to that request. The other tags generate a random r in the interval $[1, p]$, where p is the number of residents, and reply to that request if $r \leq k - 1$. Clearly, this approach expects that k tags reply to a location request on average, to satisfy the privacy requirement. In this experiment, we keep fixed the number of residents to 500 and we set $a = 0.9$ and $d = 125$ for our approach.

In Fig. 6, we report the number C of residents' collisions measured for the two approaches. The result shows that our approach always outperforms the naive approach. Specifically, it is worth noting that the curve associated with the latter has a trend similar to that obtained by our approach when $a = 1$.

Finally observe that in principle, an attacker may rely on both eavesdropped information and collusion with residents. By assuming that the greatest number of colluding residents is a given parameter τ (common assumption for collusion-based attacks), to contrast the above attack, we just should increase the privacy requirement k by τ.

Compliance with SP3 *(i.e., resistance to server-side attacks).* Here, we assume that an attacker has the rights to access the information managed by the server so that he can merge current and past information coming from different readers.

First, we observe that the knowledge of current information does not give an attacker any advantage: indeed, the mapping between a resident and its QID is never managed in a centralized way so that never the server can know such a mapping. This mapping is generated in a decentralized way depending on the real position of residents and random numbers generated by readers. Consequently, without the knowledge of such a mapping, privacy is guaranteed by Theorem 1.

As for merging current and past knowledge, the main benefit obtained by an attacker is related to the possibility of exploit-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

ing resident movements for an attack. Observe that, differently from previous cases, in server-side attacks an adversary may know consecutive reading of tag QIDs.

In this new scenario, the frequency of tags reading assumes a great importance, as shown through the following example. If we consider an average speed of 1.6 m/s for each resident, we have that a resident takes about 11 minutes to move from any point to any point, because the maximum distance between two points inside the considered environment is about 1 kilometer. As a consequence, if all readers query tags with a frequency higher than 1 reading every 11 minutes, then the attacker may take advantage from the reduced set of possible paths to guess residents' habits or, even worse, to identify them.

Therefore, to guarantee the compliance with **SP3**, we discuss the problem of setting a maximum reading frequency to solve the trade-off between privacy protection and localization accuracy. In the first experiment, we consider that each resident moves at a constant speed of 1.6 m/s and can choose a different movement direction (i.e., random shifts). We vary the reading period from 2.5 to 10.66 minutes because, as observed above, periods higher than 11 minutes do not give any advantage to the attacker.

Now, given a resident u, we measure the number C_{cvr} of residents with the same QID who are inside the *coverage area* of u, where as coverage area we mean the area that u can cross in the reading period. In this experiment, we assume that the anonymity requirement is $k = 2$. In Fig. 7, we report C_{cvr} versus tags reading period (in seconds) for different values of d. Looking at the value $d = 125$, we observe that if we set the RFID-reader query time to 300 seconds (i.e., 5 minutes), then we guarantee that there are about two users inside a coverage area (i.e., the privacy requirement $k = 2$ is satisfied). Clearly, if the value chosen for d is lower than 125, then the query time can be reduced. For instance, by looking at the same figure, if $d = 50$, we find that a query time of 150 seconds is sufficient to satisfy the privacy requirement, because it assures at least two collisions (i.e., $C_{cvr} \geq 2$).

Concerning the influence of the average speed of residents on these results, we observe that the higher this speed, the wider the coverage area. As a consequence, an increase in the speed results in an improvement of the privacy degree obtained because a wider coverage area implies a higher probability of collisions of residents inside it. In practice, the effect of speed increasing is that all the curves of Fig. 7 reach the final value (e.g., 4 for the curve $d = 125$) for a lower tag reading period.

Now, we study possible attacks when residents' movements are not fully random, as actually happens in real world. For this purpose, we use as mobility model the Random Waypoint Model [5], which has been widely adopted in the literature [53], [26], [15]. This mobility model generates resident shifts as described below. At the beginning of the simulation, each resident randomly selects one destination point inside the considered area. Then, he moves towards this end-point for a walking interval (W_{min}, W_{max}) with a speed selected uniformly at random in the interval (S_{min}, S_{max}). Speed and movement direction of each resident are selected independently of other residents. Once a resident reaches the end-point or ends his walking interval, he stops for a duration

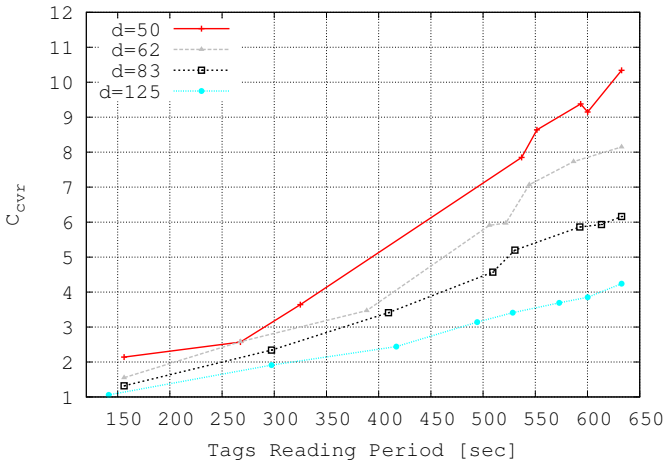

Fig. 7. Number of residents with the same QID inside a coverage area versus time (seconds).

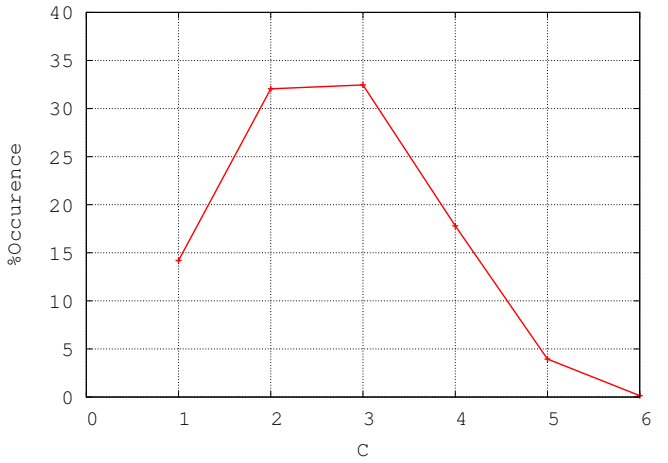

Fig. 8. Number of residents with the same QID inside the coverage area versus the percentage of involved users.

defined by the "pause" parameter that varies in the interval (P_{min}, P_{max}). After this interval, he continues his previous path or chooses another end-point and starts moving towards it. In our simulation, we set $W_{min} = 3.0$ seconds, $W_{max} = 10.0$ seconds, $S_{min} = 0.1$ m/s, $S_{max} = 1.6$ m/s, $P_{min} = 0.0$ seconds and $P_{max} = 6.0$ seconds.

In this experiment, we suppose the attacker is able to estimate the parameters of the residents' mobility model and to reduce the space of the possible target points. We set $a = 0.9$, $d = 125$, and the reading period to the value 300 seconds. For each resident u, we compute his coverage area as the maximum space he can cross in the tags reading period and count how many residents with the same QID as u fall in u's coverage area.

In Fig. 8, we report the number of collisions measured inside the coverage area versus the percentage of the involved users.

This figure shows that 85% of residents share their QID with more than 2 other residents inside the coverage area. Only 14% of them are alone. Thus, the probability that a resident is alone inside the coverage area is 0.14, but the probability of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSC.2016.2646363, IEEE Transactions on Services Computing

being alone again at the next step is $0.14 \cdot 0.14 = 0.019$ and becomes 0.003 if a third step is considered. Denoting by q the probability that a resident is alone inside a coverage area after one shift, an attacker guesses the i-th shift of a resident with a probability $\overline{q} = \prod_1^i p$. In practice, this experiment allows us to conclude that by suitably setting d, a, and the minimum tag reading period, we can reduce the probability q to satisfy any admissible privacy requirement.

Compliance with SP4 *(i.e., resistance to attacks based on fake tags or fake queries).* In attacks based on fake tags, forged tags are added to the environment to compromise the monitoring system or to allow a resident to leave the monitored environment.

In attacks based on fake queries, an adversary, by simulating to be a reader, sends queries to some tags with the purpose of *desynchronizing* the whole system and make impossible to recover the location of patients in case of emergency (i.e., DoS attack).

Both such attacks are disarmed by the authentication procedure run tags and readers included in the first step of the protocol: consequently, readers can detect fake tags and tags can detect fake queries coming from unauthorized readers and discard them. Thus, no advantage can be achieved by this type of attack.

VI. Discussion and Conclusion

As a final discussion, we provide some considerations on the effectiveness of our service.

We may state the following:

- Concerning the use of the RFID technology, we briefly recall that RFID tags can be classified into passive, semi-active and active on the basis of computational power, presence of power supply, transmission range and cost. Moreover, it is possible to distinguish between two kinds of tags operating at different frequencies and, hence, having different and adjustable transmission range (from few meters to hundreds meters): UHF tags and HF tags. In our design, we adopt RFID active tags and readers operating at UHF frequency of 433 MHz, which is allowed for healthcare applications. It is worth noting that our solution is feasible from an economic point of view, as witnessed by the fact that commercial RFID-based solutions specifically oriented to track residents in assistive environments exist, even though they do not address at all privacy issues. In our case, the cost of the solution can be estimated by considering plausible prices of about $\$2$ per tag and $\$200 - \$300$ per reader. In addition, we have to consider the cost of the network infrastructure connecting the readers, even though also non-RFID solutions require (possibly more expensive) network infrastructures. On the other hand, it is known that RFID technology is more cost effective than other technologies for localization such as Bluetooth, WLAN or UWB (Ultra-wideband) [64].
- The results obtained about the relationship between speed of residents (we consider 1.6 m/s) and parameter setting show that we obtain a good privacy level yet allowing quick residents' localization by means of a limited number of attempts and, thus, of human resources employed to manage emergency. This conclusion arises from the results of the experiments carried out to prove the compliance of our approach with **SP3** described in Section V-B.
- The attack model is realistic because the environment cannot be considered closed as a relevant number of external persons (visitors, medical representatives, maintainers, cleaners, etc.) are often present in the ALF. Thus, the environment can be considered hostile.
- The need of free movement in the environment is particularly realistic as assisted living facility are used for people with disabilities in which resident activities are monitored to help to ensure their health, safety, and well-being.
- A limitation of our approach is that it requires a hardware infrastructure to cover the whole ALF area both indoor and outdoor. Another limitation is that the setting of parameters, besides privacy requirements, is dependent on a number of features related to the considered ALF (e.g., number of residents, geometry, size).

In conclusion, the approach presented in this paper allows us to localize residents in an assisted living facility by preserving their privacy. As a matter of fact, even though the party that performs data elaboration and administration (e.g., medical staff, nursery managers, IT staff, etc.) can be assumed trusted, it is not true, in general, that the utility of having precise information about residents' location is stronger than the right of keeping private the exact movements of residents. The solution of the above trade-off is the main added value of this paper w.r.t. the existing literature, in which no protection against administrator-side attacks is provided. Our study, shows that the proposed service is enough flexible and precise to be an effective tool for residents' localization. Moreover, we show that the method is robust against several possible attacks on privacy.

References

[1] Method and apparatus for monitoring movements of an individual. http://www.google.com/patents/US6049281, 2014.

[2] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 376–385. Ieee, 2008.

[3] O. Abul, F. Bonchi, and M. Nanni. Anonymization of moving objects databases by clustering and perturbation. *Information Systems*, 35(8):884–910, 2010.

[4] ALFAA. Assisted living facilities association of america. *An overview of the assisted living industry*, 1993.

[5] F. Bai and A. Helmy. A survey of mobility models. *Wireless Adhoc Networks. University of Southern California, USA*, 206, 2004.

[6] A. Bekkali, S. Zou, A. Kadri, M. Crisp, and R. V. Penty. Performance analysis of passive uhf rfid systems under cascaded fading channels and interference effects. *IEEE Transactions on Wireless Communications*, 14(3):1421–1433, 2015.

[7] K. Bhaduri, M. D. Stefanski, and A. N. Srivastava. Privacy-preserving outlier detection through random nonlinear data distortion. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(1):260–272, 2011.

[8] F. Buccafurri, G. Lax, S. Nicolazzo, and A. Nocera. Comparing twitter and facebook user behavior: Privacy and other aspects. *Computers in Human Behavior*, 52:87–95, 2015.

[9] F. Buccafurri, G. Lax, S. Nicolazzo, and A. Nocera. A threat to friendship privacy in facebook. In *International Cross Domain Conference and Workshop (CD-ARES 2016)*, pages 96–105. Springer, 2016.

[10] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar. Preserving user location privacy in mobile data management infrastructures. In *Privacy Enhancing Technologies*, pages 393–412, 2006.

[11] C. Chow, M. F. Mokbel, and X. Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 171–178, New York, NY, USA, 2006. ACM.

[12] R. Chow and P. Golle. Faking contextual data for fun, profit, and privacy. In *Proceedings of the 8th ACM workshop on Privacy in the electronic society*, pages 105–108. ACM, 2009.

[13] H.-W. Chung. A novel time-obfuscated algorithm for trajectory privacy protection. *Services Computing, IEEE Transactions on*, 7(2):126–139, 2014.

[14] W. L. Croft, W. Shi, J.-R. Sack, and J.-P. Corriveau. Geographic partitioning techniques for the anonymization of health care data. *arXiv preprint arXiv:1505.06786*, 2015.

[15] C. Foh, G. Liu, B. Lee, B. Seet, K. Wong, and C. Fu. Network connectivity of one-dimensional manets with random waypoint movement. *IEEE Communications Letters*, 9(1):31–33, 2005.

[16] M. N. Gasson, E. Kosta, D. Royer, M. Meints, and K. Warwick. Normality mining: Privacy implications of behavioral profiles drawn from gps enabled mobile phones. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(2):251–261, 2011.

[17] B. Gedik and L. Liu. Location privacy in mobile systems: A personalized anonymization model. *icdcs*, 00:620–629, 2005.

[18] H. Ghayvat, J. Liu, S. C. Mukhopadhyay, and X. Gui. Wellness sensor networks: A proposal and implementation for smart home for assisted living. *IEEE Sensors Journal*, 15(12):7341–7348, 2015.

[19] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K. Tan. Private queries in location based services: anonymizers are not necessary. In J. T.-L. Wang, editor, *SIGMOD Conference*, pages 121–132. ACM, 2008.

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

[21] C. Hawes, C. Phillips, M. Rose, S. Holan, and M. Sherman. A national survey of assisted living facilities. *Gerontologist*, 43(6):875–882, 2003.

[22] C. Hawes, C. D. Phillips, M. Rose, et al. *High service or high privacy assisted living facilities, their residents and staff: Results from a national survey*. US Department of Health and Human Services Washington, DC, 2000.

[23] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.

[24] A. Hossain, A.-A. Hossain, S.-J. Jang, and J.-W. Chang. Privacy-aware cloaking technique in location-based services. In *Mobile Services (MS), 2012 IEEE First International Conference on*, pages 9–16. IEEE, 2012.

[25] C. Huang, P. Chung, M. Tsai, Y. Yang, and Y. Hsu. Reliability improvement for an RFID-based psychiatric patient localization system. *Computer Communications*, 31(10):2039–2048, 2008.

[26] E. Hyytiä and J. Virtamo. Random waypoint mobility model in cellular networks. *Wireless Networks*, 13(2):177–188, 2007.

[27] M. Jang, M. Yoon, H.-i. Kim, and J.-W. Chang. A privacy-aware location cloaking technique reducing bandwidth consumption in location-based services. In *Proceedings of the Third ACM SIGSPATIAL International Workshop on Querying and Mining Uncertain Spatio-Temporal Data*, pages 2–9. ACM, 2012.

[28] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias. Preventing location-based identity inference in anonymous spatial queries. *IEEE Trans. on Knowl. and Data Eng.*, 19(12):1719–1733, 2007.

[29] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis. Privacy-preserving indoor localization on smartphones. *Knowledge and Data Engineering, IEEE Transactions on*, 27(11):3042–3055, 2015.

[30] J. Krumm. Inference attacks on location tracks. In *International Conference Pervasive on Pervasive Computing*, Lecture Notes in Computer Science, pages 127–143. Springer, 2007.

[31] H. Kuijs, C. Reich, M. Knahl, and N. Clarke. Towards privacy for ambient assisted living in a hybrid cloud environment. *BW-CAR—SINCOM*, page 41, 2015.

[32] H. Kuijs, C. Rosencrantz, and C. Reich. A context-aware, intelligent and flexible ambient assisted living platform architecture. *Cloud Computing*, 2015.

[33] C. Laoudias, G. Constantinou, M. Constantinides, S. Nicolaou, D. Zeinalipour-Yazti, and C. G. Panayiotou. The airplace indoor positioning platform for android smartphones. In *2012 IEEE 13th International Conference on Mobile Data Management*, pages 312–315. IEEE, 2012.

[34] A. Lazaro, D. Girbau, and R. Villarino. Effects of interferences in uhf rfid systems. *Progress In Electromagnetics Research*, 98:425–443, 2009.

[35] B. Li, J. Salter, A. G. Dempster, C. Rizos, et al. Indoor positioning techniques based on wireless lan. In *First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, Sydney, Australia*, pages 13–16, 2006.

[36] M. Li, Z. Qin, and C. Wang. Query-privacy-aware location cloaking for mobile p2p system. *International Journal of Future Generation Communication & Networking*, 6(4), 2013.

[37] M. Li, K. Sampigethaya, L. Huang, and R. Poovendran. Swing & swap: user-centric approaches towards maximizing location privacy. In *WPES '06: Proceedings of the 5th ACM workshop on Privacy in electronic society*, pages 19–28, New York, NY, USA, 2006. ACM.

[38] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115. IEEE, 2007.

[39] N. Li, T. Li, and S. Venkatasubramanian. Closeness: A new privacy measure for data publishing. *IEEE Transactions on Knowledge and Data Engineering*, 22(7):943–956, 2010.

[40] X. Lin, R. Lu, D. Kwan, and X. Shen. REACT: An RFID-based privacy-preserving children tracking scheme for large amusement parks. *Computer Networks*, 54(15):2744–2755, 2010.

[41] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen. A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned. In *Proceedings of the 14th international conference on information processing in sensor networks*, pages 178–189. ACM, 2015.

[42] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[43] B. A. Malin, K. El Emam, and C. M. O'Keefe. Biomedical data privacy: problems, perspectives, and recent advances. *Journal of the American Medical Informatics Association*, 20(1):2–6, 2013.

[44] B. Manard, R. Cameron, and S. Kaplan. National study of assisted living for the frail elderly: Literature review update, 1996.

[45] D. J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J. Y. Halpern. Worst-case background knowledge for privacy-preserving data publishing. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 126–135. IEEE, 2007.

[46] Y. Ouyang, Y. Xu, Z. Le, G. Chen, and F. Makedon. Providing location privacy in assisted living environments. In *PETRA '08: Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, pages 1–8, New York, NY, USA, 2008. ACM.

[47] P. Peris-Lopez, J. C. Hernandez-Castro, J. M. Estevez-Tapiador, and A. Ribagorda. M2ap: A minimalist mutual-authentication protocol for low-cost rfid tags. In *Ubiquitous intelligence and computing*, pages 912–923. Springer, 2006.

[48] L. Petrou, G. Larkou, C. Laoudias, D. Zeinalipour-Yazti, and C. G. Panayiotou. Demonstration abstract: Crowdsourced indoor localization and navigation with anyplace. In *Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on*, pages 331–332. IEEE, 2014.

[49] S. Piramuthu. Protocols for rfid tag/reader authentication. *Decision Support Systems*, 43(3):897–914, 2007.

[50] D. Quercia, I. Leontiadis, L. McNamara, C. Mascolo, and J. Crowcroft. Spotme if you can: Randomized responses for location obfuscation on mobile phones. In *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, pages 363–372. IEEE, 2011.

[51] P. Rashidi and A. Mihailidis. A survey on ambient-assisted living tools for older adults. *IEEE journal of biomedical and health informatics*, 17(3):579–590, 2013.

[52] D. Rebollo-Monedero, J. Forne, and J. Domingo-Ferrer. From t-closeness-like privacy to postrandomization via information theory. *IEEE Transactions on Knowledge and Data Engineering*, 22(11):1623–1636, 2010.

[53] A. Saha and D. Johnson. Modeling mobility for vehicular ad-hoc networks. In *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 91–92. ACM, 2004.

[54] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Trans. on Knowl. and Data Eng.*, 13(6):1010–1027, 2001.

[55] P. Samarati and L. Sweeney. Generalizing data to provide anonymity when disclosing information. In *PODS*, volume 98, page 188, 1998.

[56] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki. Caravan: Providing location privacy for vanet. In *Embedded Security in Cars (ESCAR)*, 2005.

[57] A. Solanas, J. Domingo-Ferrer, A. Martínez-Ballesté, and V. Daza. A distributed architecture for scalable private rfid tag identification. *Computer Networks*, 51(9):2268–2279, 2007.

[58] A. Solanas and A. Martínez-Ballesté. Privacy protection in location-based services through a public-key privacy homomorphism. In *Public Key Infrastructure*, pages 362–368. Springer, 2007.

[59] A. Solanas and A. Martínez-Ballesté. A ttp-free protocol for location privacy in location-based services. *Computer Communications*, 31(6):1181–1191, 2008.

[60] D. Song, J. Sim, B. Kim, K. Park, J. Kang, D. Sin, I. Lee, and M. Song. Tl-diversity: Type of l-diversity for privacy protection of the clients within the cloaking area. *Testing and Measurement: Techniques and Applications*, page 403, 2015.

[61] J.-H. Song, V. W. S. Wong, and V. C. M. Leung. Wireless location privacy protection in vehicular ad-hoc networks. *Mobile Networks and Applications*, 15:160 – 171, 2010.

[62] J. Soria-Comas and J. Domingo-Ferrer. Probabilistic k-anonymity through microaggregation and data swapping. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–8. IEEE, 2012.

[63] A. Stubblefield, J. Ioannidis, and A. D. Rubin. A key recovery attack on the 802.11b wired equivalent privacy protocol (wep). *ACM Trans. Inf. Syst. Secur.*, 7(2):319–332, 2004.

[64] S. P. Subramanian, J. Sommer, S. Schmitt, and W. Rosenstiel. Ril-reliable rfid based indoor localization for pedestrians. In *Software, Telecommunications and Computer Networks, 2008. SoftCOM 2008. 16th International Conference on*, pages 218–222. IEEE, 2008.

[65] S. S. Suryawanshi and V. S. Wadne. Securing health care data in collaborative data publishing using mapreduce framework. 2015.

[66] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[67] R. Trujillo-Rasua and J. Domingo-Ferrer. On the privacy offered by *(k, δ)*-anonymity. *Information Systems*, 38(4):491–494, 2013.

[68] R. Trujillo-Rasua and A. Solanas. Efficient probabilistic communication protocol for the private identification of rfid tags by means of collaborative readers. *Computer Networks*, 55(15):3211–3223, 2011.

[69] A. Vilamovska, E. Hatziandreu, H. R. Schindler, C. V. Oranje-Nassau, H. D. Vries, and J. Krapels. Study on the requirements and options for rfid application in healthcare, 2009.

[70] Q. Wang, W. Shin, X. Liu, Z. Zeng, C. Oh, B. K. AlShebli, M. Caccamo, C. A. Gunter, E. Gunter, J. Hou, et al. I-living: An open system architecture for assisted living. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 4268–4275. IEEE, 2006.

[71] Y. Wang and K. N. Plataniotis. An analysis of random projection for changeable and privacy-preserving biometric verification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 40(5):1280–1293, 2010.

[72] L. C. Watson, J. M. Garrett, P. D. Sloane, A. L. Gruber-Baldini, and S. Zimmerman. Depression in assisted living: Results from a four-state study. *The American journal of geriatric psychiatry*, 11(5):534–542, 2003.

[73] A. D. Wood, J. A. Stankovic, G. Virone, L. Selavo, Z. He, Q. Cao, T. Doan, Y. Wu, L. Fang, and R. Stoleru. Context-aware wireless sensor networks for assisted living and residential monitoring. *IEEE network*, 22(4):26–33, 2008.

[74] T. Xu and Y. Cai. Location anonymity in continuous location-based services. In *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, pages 1–8, New York, NY, USA, 2007. ACM.

[75] Z. M. Yahia. Radio frequency identification: the big role player in health care management. *Journal of Health Organization and Management*, 25(5):490–505, 2011.

[76] L. Yang, F. Hao, S. Li, G. Min, H. Kim, and S. Yau. An efficient approach to generating location-sensitive recommendations in ad-hoc social network environments. *IEEE Transactions on Services Computing*, 8(3):520–533, 2015.

[77] Z.-t. Yu, Q. Qian, C.-Y. Lin, and C.-L. Hung. High performance datafly based anonymity algorithm and its l-diversity. *International Journal of Grid and High Performance Computing (IJGHPC)*, 7(3):85–100, 2015.

[78] D. Zhang, X. Wang, X. Song, and D. Zhao. A novel approach to mapped correlation of id for rfid anti-collision. *Services Computing, IEEE Transactions on*, 7(4):741–748, 2014.

[79] J. Zhang, C.-Y. Chow, and Y. Li. igeorec: A personalized and efficient geographical location recommendation framework. *IEEE Transactions on Services Computing*, 8(5):701–714, 2015.

[80] H. Zhu, S. Tian, and K. Lü. Privacy-preserving data publication with features of independent l-diversity. *The Computer Journal*, 58(4):549–571, 2015.

[81] S. Zimmerman. *Assisted living: Needs, practices, and policies in residential care for the elderly*. JHU Press, 2001.

Francesco Buccafurri. Full professor of computer science at the University Mediterranea of Reggio Calabria, Italy. In 1995 he took the PhD degree in computer science at the University of Calabria. In 1996, he was visiting researcher at the database and knowledge representation group of Vienna University of Technology. His research interests include deductive-databases, knowledge-representation and non-monotonic reasoning, model checking, information security, data compression, data streams, agents, P2P systems. He has published several papers in top-level international journals and conference proceedings. He serves as a referee for international journals and he is a member of a number of conference PCs. He is Associate Editor of Information Sciences (Elsevier), is also included in the editorial board of a number of international journals and played the role of PC chair in some international conferences.

Gianluca Lax. Assistant professor of computer science at the University Mediterranea of Reggio Calabria, Italy. In 2005, he took the PhD degree in computer science at the University of Calabria. His research interests include information security and social network analysis. He is author of more than 100 papers published in top-level international journals and conference proceedings.

Serena Nicolazzo. PhD student at the University Mediterranea of Reggio Calabria. She received his MsC Degree in Telecommunication Engineering from the University Mediterranea of Reggio Calabria in July 2013. Her research interests include security, privacy, social network and multi-social network analysis.

Antonino Nocera. Research assistant at the University Mediterranea of Reggio Calabria. He received his MsC Degree in Telecommunication Engineering from the University Mediterranea of Reggio Calabria in July 2009 and his PhD in Information Engineering from the University Mediterranea of Reggio Calabria in March 2013. His research interests include social network and social internetworking analysis, privacy, security, trust and reputation, schema integration, XML, intelligent agents and folksonomies.