Computer Communications 65 (2015) 43–54

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

A platform for privacy protection of data requesters and data providers in mobile sensing

Ioannis Krontiris [a], Tassos Dimitriou [b,*]

[a] Goethe University Frankfurt, Grueneburgplatz 1, 60323 Frankfurt, Germany
[b] Computer Engineering Dept., Kuwait University, Kuwait

ARTICLE INFO

Article history:
Available online 21 February 2015

Keywords:
Mobile ubiquitous sensing
Privacy
Anonymity

ABSTRACT

In typical mobile sensing architectures, sensing data are collected from users and stored in centralized servers at third parties, making it difficult to effectively protect users' privacy. A better way to protect privacy is to upload sensing data on personal data stores, which are owned and controlled by the users, enabling them to supervise and limit personal data disclosure and exercise access control to their data. The problem however remains how data requesters can discover the users who can offer them the data they need. In this paper we suggest a mobile sensing platform that enables data requesters to discover data producers within a specific geographic region and acquire their data. Our platform protects the anonymity of both requesters and producers, while at the same time it enables the incorporation of trust frameworks, incentive mechanisms and privacy-respecting reputation schemes. We also present extensive experimental results that demonstrate the efficiency of our approach in terms of scalability, load balancing and performance.

1. Introduction

The increasing availability of sensors on today's smartphones and other everyday devices, carried around by millions of people, has opened up diverse kinds of information gathering by people and their devices. Eventually, researchers envision the creation of a unified data-sharing infrastructure, where people and their mobile devices provide their collected data streams in accessible ways to third parties interested in integrating and remixing the data for a specific purpose. This trend is often named *mobile crowd sensing* [1].

Several of the works on mobile crowd sensing systems started differentiating very early between two data collection models [2]. In the participatory model, users are actively involved in the collection process by deciding on the spot when to report data, while in the opportunistic model, sensor sampling occurs whenever the state of the device (e.g. geographic location) matches the application's requirements described in a sensing task, without the knowledge of the individual phone user.

Independently from the collection model however, what is common in the majority of existing architectures is that the sensing data collected from the mobile phones are stored in centralized servers at third parties, creating massive databases of individuals' location, movements, images, and even health data. After collecting the data, the entity controlling the database aggregates, processes and releases them through various interfaces (e.g. statistical data on a map).

This paradigm raises several challenges concerning information access and reciprocity. Who controls data collection and who owns the data or benefits from them? In most cases, the data are collected, stored, and analyzed by data processors typically out of view of the individual whose life they describe. The collection of the data is not always restricted to the purpose for which they were collected. Also, individuals cannot pose restrictions on the collection and processing of their data and the data collected from them are not made available back to them through proper interfaces.

To deal with the power imbalance created in such paradigms, architectures taking a more user-centric approach started to appear [3]. What these architectures try to do is to enable individuals to supervise and limit personal data disclosure and exercise access control to their data by third parties. Several existing solutions in crowd sensing applications suggest a vault-like entity to provide an online trusted storage and processing. Mobile phones sense and upload data to this vault, which is owned and controlled by the individual. The process of storing personal data streams is decoupled from the sharing of that information. After the collection and archival of data, the users can define their own privacy policies and review/control who can see which kind of data.

* Corresponding author.
E-mail address: tassos.dimitriou@ieee.org (T. Dimitriou).

http://dx.doi.org/10.1016/j.comcom.2015.02.005
0140-3664/

The problem that remains largely unexplored in these architectures is that of *information discovery* from data consumers. We refer specifically to the case where data requesters, either being applications or physical persons, are interested in retrieving information according to some requirements (location area, time frame, sensor type, etc.) from multiple data contributors that satisfy these requirements. That means they need to search data from data contributor's individual data stores, since there is not a central place where all data is gathered. Given the distributed nature of data stores controlled by the corresponding data producers, this is not trivial to do.

But discovering data providers in a specific geographic area is not sufficient for a requester, because not all of them can provide the same quality of data. Data providers can inadvertently position their device in the wrong place while collecting sensor readings or they could deliberately contribute bad data. So, requesters want to select specific data contributors based on criteria like their reputation gathered in previous participations, in order to guarantee some quality in the sensed data. It is also usual that the requestor offers some micro-payments to the producers for motivating them to contribute. Then an additional selection criterion could be based on the amount that the producers require for their data.

What makes the problem more challenging is the growing requirement of protecting requester's data access privacy; a user may want to keep confidential whether (and when) she accessed the sensed data, the data types she was interested in, or from which nodes she obtained the data, as the disclosure of such information may be used to infer additional context about the user and used potentially against her interest. This means that data requests cannot be linked to the real identity of the requester. However, some access control mechanism is needed, so that not anybody can take benefit of the platform's services without demonstrating some sort of permission. How this permission can be obtained depends on the business model of the platform provider. For example, it could be that the requester has to pay for each "*sensing quantum*". To preserve the requester's privacy, the process of acquiring such a sensing quantum and the process of demonstrating it to the data providers for enabling the processing of the request should be unlinkable to each other.

Our Contribution. In this paper, we suggest a platform that enables data requesters to discover data providers in a specific geographic region of interest and retrieve data from them, while protecting the privacy of both against each other and the platform providers. Data providers maintain location privacy according to their own preferences and access policies to the data they own, while requesters may contact directly the mobile users in the area of interest and select the ones to get sensing data from, based on their own criteria.

Organization. In Section 3, we discuss the system and threat model and highlight the major components of the platform. In this work cloud agents represent the mobile users to the outer world according to the user's location privacy preferences. They interconnect with each other in such a structure that enables data requesters to discover mobile users in a specific geographic area. This structure (Section 4) is not stored by a central entity, but it is maintained by the agents in a *distributed* fashion, thus avoiding the bottlenecks and the privacy implications of centralized approaches. Requesters obtain *tokens* from the service provider in order to have access to the data provided by the selected data providers. Using appropriate cryptographic mechanisms that we describe in Section 5, the validity of the token can be verified without leaking the identity of the requester to the node or to the application owner. Sections 6 and 7 describe how the platform can incorporate incentives and reputation management mechanisms. Finally, in Section 8 we conduct extensive experiments demonstrating the efficiency of our approach in terms of scalability, load balancing and performance.

2. Related work

In the participatory sensing domain, various *centralized* solutions for distributing tasks or queries to sensor nodes have been proposed. In PRISM [4], participating nodes (i.e. mobile phones) register with the server and the server tracks the nodes and pushes only matching tasks to them, based on their context (e.g. location). For example, Alice may be assigned the task "measure temperature in area X", when she is entering this area. However, this solution does not consider privacy for any of the involved entities, queriers or mobile nodes.

A solution that offers a privacy-friendly way of task distribution is AnonySense [5]. Sensing tasks are posted on a server and the nodes download the tasks and match them to their context to decide which one to execute. This approach has the advantage that the nodes do not reveal anything about their context to the service provider, in order to receive the sensing task. Still, AnonySense does not consider privacy for the entities posting the tasks.

Recently, PEPSI [6] was suggested as a system designed with the privacy of the queriers in mind, queriers being entities external to the platform, who are interested in some specific sensing information. PEPSI is based on a *centralized* solution and to protect the privacy of the queriers, it introduces a Registration Authority, a trusted third party which collects queries from the queriers and provides back the corresponding cryptographic material. In that sense, the queries reach the platform in an encrypted form. However, the problem is shifted to the Registration Authority, where essentially all queries are known in advance, along with the identities of the queriers, leading to the assumption that this entity must be trusted.

To protect the privacy of data providers, we employ the concept of user-owned proxies. The use of stationary proxies has been already suggested in some participatory sensing systems so far, where they are used as data vaults or brokers for the user [7]. For example, Mun et al. proposed Personal Data Vault (PDV) [8], which functions as individual data storage with fine-grained access control mechanism, privacy rule recommender, and trace audit. Choi et al. also presented SensorSafe [9], an architecture that consists of multiple remote data stores and a broker enforcing a fine-grained access control by supporting privacy rules with context/behavior conditions and control for levels of inferences. What is common in all of the above systems is that the mobile phones sense and upload their data to their corresponding proxies *proactively*. Then the proxy is responsible to help the user manage this data and make it available to third parties, functioning as an access control mechanism. In our work, mobile nodes perform sensing operations only *reactively*, when there is a specific query to which they can respond to.

Drosatos et. al. [10] presented recently a privacy-respecting solution following the same reactive model, where mobile agents on the cloud store the data encrypted and execute a cryptographic protocol based on a homomorphic encryption scheme in order to aggregate the data and make them available. However, this setting does not consider the privacy of the data requesters, neither does it enable them to apply selection criteria on specific mobile nodes.

3. System and threat model

3.1. Main entities of the system

The Data Provider Users carry personal sensors, either embedded in their mobile phones or part of wearable devices, and collect contextual data from their immediate environment. We will use the term *Data Provider* to refer to these entities in the system who are actively sharing their sensing data with others. Part of