POSTER: Enabling End-Users to Protect their Privacy

Mahmoud Barhamgi[1,2], Mu Yang[1], Chia-Mu Yu[3], Yijun Yu[1],
Arosha K. Bandara[1], Djamal Benslimane[2], Bashar Nuseibeh[1,4]
[1]The Open University, United Kingdom
firstname.lastname@open.ac.uk
[2]Claude Bernard University, France
firstname.lastname@univ-lyon1.fr
[3]National Chung Hsing University, Taiwan
chiamuyu@nchu.edu.tw
[4]The Irish Software Research Centre, Irland

ABSTRACT

In this paper we present our ongoing work to build an approach to empower users of IoT-based cyber physical systems to protect their privacy by themselves. Our approach allows users to identify the privacy risks involved in sharing private data with a data consumer, assess the value of their private data based on identified risks and take a pragmatic data sharing decision balancing the risks with the benefits generated by the sharing. Our approach features a knowledgebase, called the *Privacy Oracle*, that exploits the power of the Semantic Web to determine how raw metadata can be combined by data consumers to infer privacy-sensitive information as well as the privacy risks associated with the disclosure of inferred information.

CCS Concepts

•**Security and privacy** → *Privacy protections;*

Keywords

Privacy; IoT based smart environments; Privacy ontologies

1. INTRODUCTION

In today's world, we progressively find ourselves surrounded by new IoT-based cyber-physical systems that silently track our activities and collect sensitive information about us. Among the most prominent examples, we cite smart environments (e.g., smart homes and cities), quantified self technologies, smart energy meters, etc. While such systems promise to ease our lives, they raise major privacy concerns for their users, as collected data is often privacy-sensitive, such as location of individuals, patients' vital signs. In this work, we address these privacy concerns by proposing a solution that allows users to play a central role in protecting their privacy.

Figure 1: User-centred Data Sharing Methodology

This work is motivated by two key observations. First, existing approaches for privacy protection in IoT-based cyber-physical systems give users only a passive role in protecting their privacy [3]. Typically, users are prompted by the system to supply their privacy preferences as to who can access their data and for what purposes, and to accept a privacy policy referring to their preferences. However, users may not be aware of the direct and indirect risks associated with the disclosure of their data to a given entity to correctly specify their privacy preferences, which can also change depending on the user's context. Second, recent studies [2, 1] show that users are becoming more conscious of their privacy, and tend to take a pragmatic stance on sharing their private data, i.e. they would accept to release some of their private data in exchange for some incentives or services. However, making a tradeoff between the benefits generated by sharing private data and associated privacy concerns remains a challenging task to users [4]. This means that users can play a bigger role in protecting their privacy, but need assistance to play that role and to derive value out of their private data.

This work adopts the methodology depicted in Figure 1 to empower users to protect their privacy by themselves. That is, users should be enabled, before sharing a private data item (or a combination thereof) with a data consumer to:

1. Understand the privacy risks involved in that sharing;
2. Assess the value of the data to be shared, based on the identified privacy risks, and compare it to the potential benefits generated by the sharing;
3. Negotiate with data consumers to attain a (trade-off)

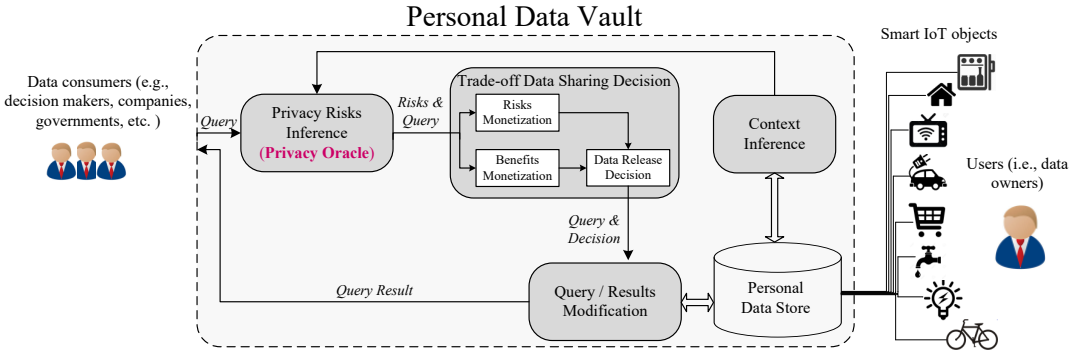

Figure 2: Overview of the proposed solution

data sharing decision satisfying both parties when conflicts happen;

4. Control the data release by applying the necessary data modification techniques (e.g., anonymization, data perturbation, modification, etc.) to implement the desired sharing decision.

We present in the following section an approach for empowering users to protect their privacy.

2. OUR APPROACH

In this work, we use the terms "*data owners*" to designate the users of cyber-physical systems that generate data by interacting with the systems (e.g., occupants of smart homes, monitored patients, etc.), and "*data consumers*" to designate the stakeholders that are interested in collecting and exploiting the data generated, such as electricity companies in smart grids, healthcare providers in intelligent healthcare networks, government agencies, etc.

We also use the term a "*privacy risk*" to designate a potentially harmful use of a disclosed personal data item that can be made against the individual, or a harmful effect of that disclosure. Examples include, loss of the individual's job or reputation, unfair discrimination, etc.

We give in Figure 2 an overview of our solution, which we detail in the following. The `Personal Data Vault` (`PDV`) is a secure private data container that falls under the control of data owners. Raw data generated by connected things (IoT objects) are stored within the `Personal Data Store` of a `PDV` before being released to any data consumers. All data flows between data owners and consumers pass by the `PDV`s of data owners.

When a data consumer queries some personal private data from a data owner (either directly, or by getting the data owner to use a service/application provided by the data consumer), the `PDV` processes the received query by carrying out the following steps (detailed in the following subsections):

2.1 Privacy risks inference

The `PDV` assesses, through the `Privacy Risks Inference Component` (called the `Privacy Oracle` onwards), the privacy risks associated with releasing the requested data to a given data consumer along with their probabilities. Knowing these privacy risks can help data owners better understand and evaluate the value and sensitivity of requested data. For example, granular readings of smart electricity meters can be analyzed to infer information about the occupants such as their presence/absence and wake/sleep cycles, the possession of specific devices (e.g., a medical device), etc. The disclosure of such information could lead to privacy risks such as being subject to discrimination, surveillance, burglaries, etc. Furthermore, the fact that data consumers may also have side information about data owners, or are able to combine multiple pieces of IoT collected data, can increase their inference capabilities, thus the sphere of possible privacy risks. For example, a malicious consumer could combine location data with a user's vital signs (e.g., heart-rate), to determine whether the user has an extramarital affair by determining if the user has a sexual activity outside his dwelling.

The `Privacy Oracle (PO)` is a knowledgebase that uses the Semantic Web technologies to determine the implicit information that can be inferred out of IoT collected data, along with their associated privacy risks (e.g., discrimination, surveillance, loss of job, etc.). To do so, the `PO` models the following aspects:

- The metadata collected by IoT objects in the considered smart environment (e.g., location, energy consumption, vital signs, etc.), user's context as well as all privacy-sensitive information that can be inferred along with their associated risks. These data features are modelled using a domain ontology;
- The inference relationships that exist between collected metadata and user's privacy-sensitive information. These relationships are represented as inference rules (expressed in a language such as the Semantic Web Rule Language SWRL).

We present in Figure 3 simplified[1] examples of inference rules. `Rule-1` states that the use of a device can be inferred from the readings of an energy smart meter (EMR). The terms `Person`, `EMR`, `Device` are ontological concepts in the domain ontology, whereas `hasEMR`, `isShared`, `useDevice`, `isInferrable` are properties. `Rule-2` states that the use of a medical device reveals the health conditions for which the device is used. `Rule-1` and `Rule-2` can be combined together to infer the fact that the health conditions could be inferred from the readings of an energy smart meter. `Rule-3` simply states that *location* and *heart-rate* metadata could be combined to infer if data owner has an extramarital affair (i.e. *heart-rate* can be used to infer if data owner is engaged in a sexual activity, and *location* can be exploited to infer whether data owner is in a suspicious location (e.g., outside home)). The rule uses contextual information such as whether the data owner is married.

[1]For clarity, we omit details such as the certainty of the inference and the associated inference algorithm.

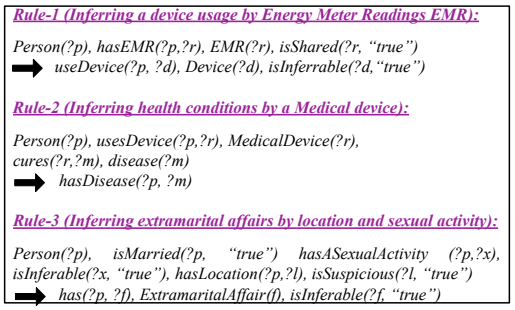

Figure 3: Sample Inference Rules

Privacy experts can supplement the `Privacy Oracle` with inference rules representing the capabilities of new data mining algorithms and the possible combinations of metadata.

2.2 Making a pragmatic data sharing decision

The `PDV` monetizes (i.e. quantifies) the identified privacy risks and the potential benefits using a numerical model, through the `Trade-off Data Sharing Component (TDS)`, and helps data owner take an informed and pragmatic data sharing decision balancing the two. The `TDS` builds on an extension of a numerical model that is proposed in [4] to quantify the privacy risks. Figure 4 shows a simplified overview of our extended model. The value of a private data item is computed based on factors such as the probability (i.e. the certainty) of the information that can be inferred out of the data item, the probability of the consumer misusing the inferred information (or the trust in data consumer), the user's preference (i.e. how sensitive the user considers their private data to be) and the value of the data item in a free data market (a reference value). Red pluses (+) and minuses (−) represent how a factor affects the computation of its dependent factors. A decision denotes whether the data item can be shared with the consumer, along with its accuracy (i.e. precision). Privacy decisions change the precision of released data to a level that would avoid the privacy risks that cannot be taken at the provided benefits.

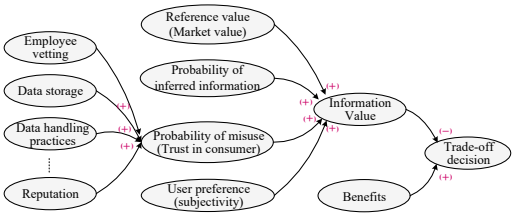

Figure 4: Trade-off Decision Model

2.3 Query / Result modification

Finally, the `PDV` modifies, through the `Query/Result Modification component`, the query before being applied to the `personal data store` to discard the data items to which the data consumer is not entitled, and the query's result to tune its accuracy before its submission to data consumer.

3. IMPLEMENTATION

We applied our solution to a smart environment for monitoring elderly people. The environment involves wearable sensors for monitoring several vital signs such as heart rates, blood pressure, ECG as well as smart objects and sensors installed in fixed positions of the monitored environment

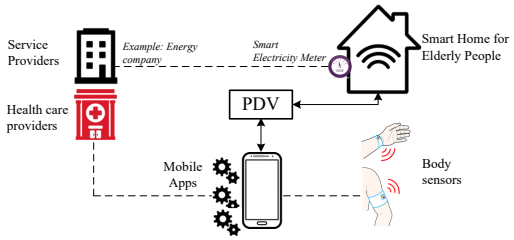

Figure 5: Evaluation Architecture

(i.e. a House). Figure 5 shows the implementation architecture. Healthcare providers provide users with personalized healthcare services by consuming collected data (through their mobile apps). Data flows between users and data consumers go through the `PDV`. Figure 6 (Window-1) shows the user interface of the `PDV`. Upon the reception of a new data sharing request, the `PDV` takes into account the user's context and her shared data (Window-2) to provide her with a description of associated privacy risks (Window-3) as well as a set of recommended actions (Window-4).

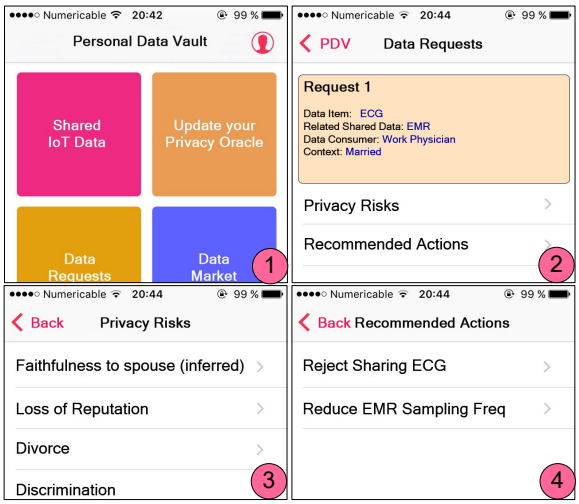

Figure 6: Demonstration Scenario

4. CONCLUSION AND FUTURE WORK

In this paper, we presented an ongoing work to enable the users of smart cyber-physical systems to protect their privacy by themselves. We are currently refining our approach with pricing models that would fit for various forms of benefits, e.g., financial, social benefits, etc.

5. ACKNOWLEDGMENTS

This work is supported, in part, by SFI grant 13/RC/2094, ERC Advanced Grant 291652 (ASAP), QNRF grant NPRP 05-079-1-018 and a MOST grant 105-2218-E-155-010.

6. REFERENCES

[1] B. Knijnenburg. Simplifying privacy decisions. In *ACM RecSys*, pages 40–41, 2013.
[2] C. Li, D. Li, and G. Miklau. A theory of pricing private data. *ACM Trans. Database Syst.*, 39(4):34, 2014.
[3] S. Sicari and A. Rizzardi. Security, privacy and trust in internet of things. *Computer Networks*, 76:46–64, 2015.
[4] M. Yang, Y. Yu, A. K. Bandara, and B. Nuseibeh. Adaptive sharing for online social networks: A trade-off between privacy risk and social benefit. In *IEEE TrustCom 2014*, pages 45–52, 2014.