Explanation-based Tool for Helping Data Producers to Reduce Privacy Risks

Hira Asghar, Christophe Bobineau, Marie-Christine Rousset

Explanation-based Tool for Helping Data Producers to Reduce Privacy Risks *

Hira Asghar, Christophe Bobineau, and Marie-Christine Rousset

Université Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France
firstname.lastname@univ-grenoble-alpes.fr

Abstract. This paper demonstrates the interactive user interface *PrivEx* that helps data producers to reduce privacy risks raised by data collection from service providers in the Semantic Web of Things. *PrivEx* provides several types of support to data producers in their management of the tension between the privacy risks and the utility of the data they accept to publish. *PrivEx* is grounded on the formal framework presented in [1] for detecting automatically privacy risks raised by utility queries. In the demonstration, we will illustrate the functionalities of *PrivEx*, on a smart meter scenario inspired by a real-world use case providing time series of electrical consumptions of different customers associated with some metadata on their demographics, home sizes and equipment.

Keywords: Privacy risks · Data exchange · Semantic Web of Things

1 Introduction

Personal data are increasingly disseminated over the Web through mobile devices and smart environments. They are exploited for developing more and more sophisticated services and applications. All these advances come with serious risks for privacy breaches that may reveal private information wanted to remain undisclosed by data producers. It is therefore of utmost importance to help them to identify privacy risks raised by requests of service providers for utility purposes. In [1], we have presented a formal framework supporting utility-aware privacy preservation in the setting of applications where service providers request collecting data from data producers to perform useful aggregate data analytics. The approach that we promote to face the privacy versus utility dilemma in this setting can be summarized as follows:

- Data producers specify by a set of *privacy queries* (kept secret) the (possibly aggregated) data they *do not want* to be disclosed.

- Data consumers make explicit by a set *utility queries* the data they request to each data producer for offering them services in return.

* Partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003), PERSYVAL-Lab (ANR-11-LABX-0025-01) and TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215

- The compatibility between privacy and utility queries is automatically verified, and in case of incompatibility data producers get an explanation that can be exploited later to help them find an acceptable privacy-utility trade-off.

In this paper, we demonstrate *PrivEx* an interactive user interface that we have built[1] on top of the implementation of the formal results presented in [1] for detecting automatically privacy risks raised by utility queries. The user interface *PrivEx* provides several types of support to data producers in their management of the tension between the privacy risks and the utility of the data they accept to publish. First, it presents in an interpretable form the requests of a service provider for utility purpose. Second, it provides a form-based interface for guiding data producers in construction of privacy queries. Third, it detects the privacy risks and provides a factual explanation for each detected privacy risk. Last, it provides several options for modifying the utility queries to reduce the detected privacy risks.

2 Smart Meter Use Case

We consider a smart meter scenario inspired by a real-world use case provided by the *Irish Social Science Data Archive (ISSDA) Commission for Energy Regulation (CER)* [2]. This dataset includes time series of electrical consumptions of different house owners. In addition, pseudonymized metadata are available on customers' demographics, home sizes and equipment associated to the electric consumption time series. For capturing the properties describing the smart meter data and the associated customers' metadata in a uniform way, we have designed a simple RDFS ontology [3].

This ontology provides a shared vocabulary used by service providers to express their utility queries (as illustrated in Figure 1) and by data producers to express their privacy queries (as shown in Figure 2).

In their most general form, the (privacy and utility) queries have 4 parts: *(i)* a *core pattern* that specifies the combinations between properties to be satisfied by the requested data; *(ii)* a *constraints* part on the values of some of the properties for filtering more precisely the requested data; *(iii)* a *result* defining the target properties the values of which must be returned by the query evaluation, and possibly an aggregate function to be computed on groups; and *(iv)* a *time window* part, if the aggregate function is computed on a dynamic property (such as *issda:consumption*), to specify the time intervals over which the aggregation must be computed.
Time windows are specified with two parameters: a *size* to express the duration of each time window, and a *step* to express the time interval separating consecutive time windows, which can thus be sliding (like in the UQ3 query in Figure 1 or tumbling (like in the PQ2 query in Figure 2).

[1] Code is available at https://github.com/repository-code/PrivEx
[2] https://www.ucd.ie/issda/data/commissionforenergyregulationcer/
[3] Available at https://raw.githubusercontent.com/fr-anonymous/puck/main/issda_schema.ttl

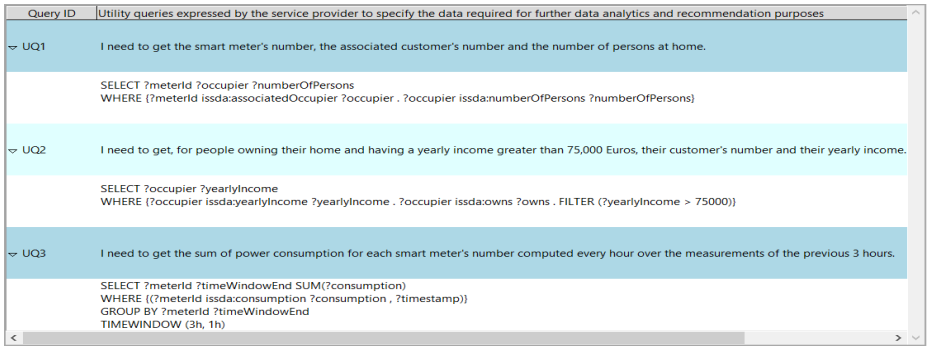

Fig. 1. Example of 3 utility queries expressed in their textual and SPARQL-like syntax

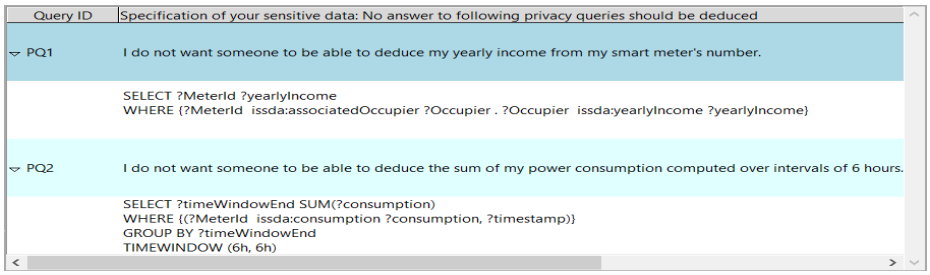

Fig. 2. Example of 2 privacy queries expressed in their textual and SPARQL-like syntax

3 Demonstration Scenario

In the demo, the following functionalities of *PrivEx* will be demonstrated.

1. *Guided construction of privacy queries:* The form-based interface facilitates the step by step construction of each privacy query as illustrated in Figure 3. In first step, the user enters the textual description of the query to be constructed. The second step (construction of the core pattern of each query) is guided by the display of the ontology to help the user choose properties. For the other parts, the user is guided by the interface to enter their choices easily. During the demo, the attendees can see the interactive construction of privacy query PQ2.
2. *Detection and explanation of privacy risks:* Each detected privacy risk comes with an explanation based on the proof produced by the incompatibility checking algorithm decribed in [1]. Each privacy risk is explained using two different levels as illustrated in Figure 4. The first level simply points out the privacy queries likely to be violated by some utility queries that are also shown to the user. The second level exhibits the corresponding privacy risk by providing a counter example in the form of synthetic data built from the ontology and the (utility and privacy) queries involved. In the demo, attendees can see the explanation for each detected privacy risk.

3. *Guided negotiation to reduce privacy risks:* As illustrated in Figure 5, the interface lists several options for negotiating the utility queries involved in privacy risks, either by refusing to answer them, or by modifying their result, or by generalizing their conditions, or by changing the aggregate function, or by changing the time window size or step. In the demo, attendees can observe how the interface guides users for interactively removing or reducing the privacy risks.

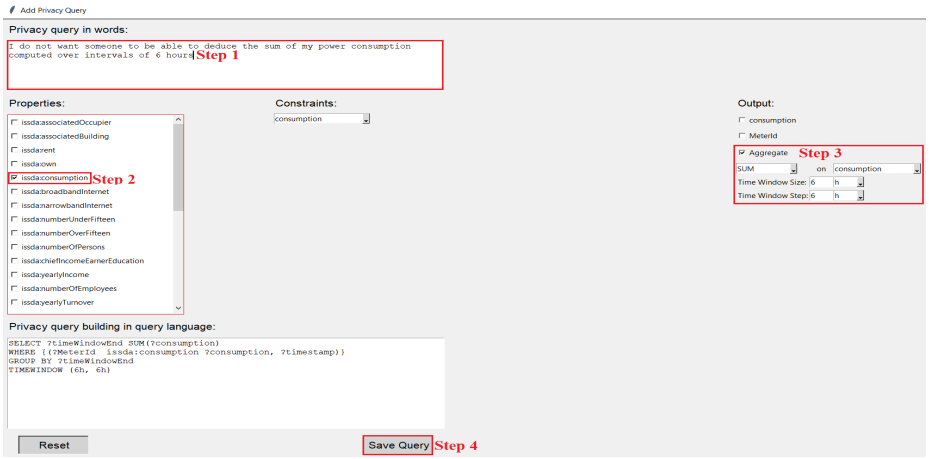

Fig. 3. Screenshot of the steps followed for the construction of privacy query PQ2

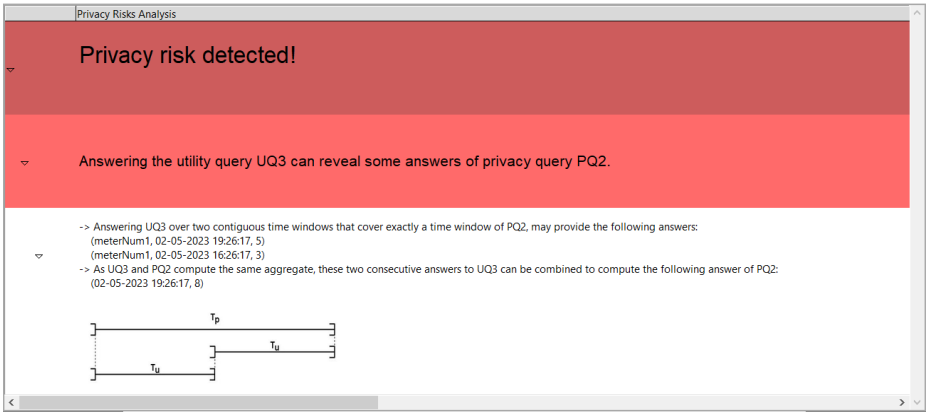

Fig. 4. Screenshot illustrating the explanation of detected privacy risk

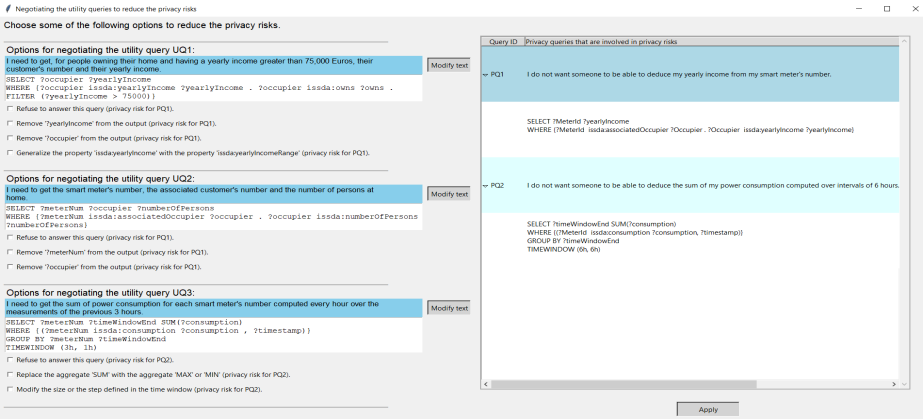

Fig. 5. Screenshot of the interface for guiding the negotiation of privacy risks

The functionalities of *PrivEx* are demonstrated in a demo video that is accessible via the following link: https://www.veed.io/view/3f1f8db3-0ca8-4ebc-b143-52bdc26f73de?panel=share

Reference

1. Asghar, H., Bobineau, C., Rousset, M.C.: Identifying privacy risks raised by utility queries. In: Web Information Systems Engineering – WISE 2022: 23rd International Conference, Biarritz, France, November 1–3, 2022, Proceedings. p. 309–324. Springer-Verlag, Berlin, Heidelberg (2022). https://doi.org/10.1007/978-3-031-20891-1_22, https://doi.org/10.1007/978-3-031-20891-1_22