Computer Communications 116 (2018) 200–211

Contents lists available at ScienceDirect

Computer Communications

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

Increasing user controllability on device specific privacy in the Internet of Things

Waqar Asif*[,a], Muttukrishnan Rajarajan[a], Marios Lestas[b]

[a] School of Engineering and Mathematical Sciences, City, University of London, UK
[b] Department of Electrical Engineering, Frederick University, Nicosia, Cyprus

ARTICLE INFO

Keywords:
Smart meter
Privacy
Utility-Privacy trade-off
Mutual information
Non-intrusive load monitoring

ABSTRACT

With recent advancements in information technology more and more devices are integrated in the Internet of Things. These devices gather significant amount of private information pertinent to a user and while, in some cases it helps in improving the life style of an individual, in others it raises major privacy concerns. This trade-off between utility and privacy is highly dependent upon the devices in consideration and as the utility of the generated data increases, the privacy of an individual decreases. In this paper, we formulate a utility-privacy trade-off that enables a user to make appliance specific decisions as to how much data can be shared. This is achieved by parametrizing the degree of privacy allowed for each device and enabling the user to configure the parameter of each device. We use the smart metering application as the test case scenario for the proposed approach. We evaluate its performance using simulations conducted on the ECO data set. Our results indicate that, the proposed approach is successful in identifying appliances with an accuracy of 81.8% and a precision of 70.1%. In addition, it is demonstrated that device specific changes of the configuration parameters allow the degree of privacy achieved for the particular device and the utility to be well controlled, thus demonstrating the effectiveness of the proposed approach. Moreover, it is shown that, as expected, devices with higher power consumption contribute more to the overall privacy and utility achieved. A comparative study is also conducted and the proposed approach is shown to outperform the existing ElecPrivacy approach by producing a trace that is harder to identify, as reported after testing the Weiss' and Baranski's algorithm, both of which are well known Non-Intrusive Load Monitoring algorithms. Finally, it is demonstrated that the addition of noise, which is an integral part of the propose approach, can greatly improve performance.

1. Introduction

Internet of Things (IoT) has been one of the key innovations of the last decade. It eliminates the requirement of having a homogeneous network for ensuring information exchange, thus, enabling multiple devices to communicate with each other. By 2020, it is expected that 24 billion devices will be connected using this innovative technology thus, generating an expected benefit of 1.2 trillion USD for the communication industry [1]. IoT objectives go beyond serving the communications industry, aiming to improve the quality of life. IoT devices are built on a common architecture of repetitive sensing and forwarding of information thus forming time-series data. This time series data keeps the client aware and updated about the changing parameters thus enabling him to react accordingly. A smart health monitoring device enables a doctor to constantly update the record of a patient's health and later use this to make a detailed analysis [2] Likewise, a smart baby monitor empowers parents with the capability to take pre-emptive measures against the reduced growth of their child with the aid of a constantly updated record of their child's activity pattern [3,4]. Similarly, smart meters enable consumers to have knowledge on their consumption pattern thus allowing them to adopt a more cost effective lifestyle [5,6]. The benefits of having these IoT devices is only worthwhile if the generated time-series data is kept secure from the access of an adversary.

An adversary can use this data to learn vital information about a person with information spanning from as small as his average heart rate for a certain period of time to as huge as a detailed analysis of a person's heart rate, location and estimated activity pattern, all using the same health monitor. Such highly private data collection may happen with or without the consent of the user and potentially without informing the users how this data can affect their privacy [7]. The entity collecting the data may also make this data available to third parties such as insurance companies for accessing one's daily lifestyle to identify if they are maintaining a healthy schedule and then adjusting the premium accordingly [8], marketing agencies for initiating targeted

* Corresponding author.
E-mail addresses: waqar.asif@city.ac.uk (W. Asif), r.muttukrishnan@city.ac.uk (M. Rajarajan), eng.lm@frederick.ac.cy (M. Lestas).

https://doi.org/10.1016/j.comcom.2017.11.009
Received 11 June 2017; Received in revised form 25 September 2017; Accepted 18 November 2017
Available online 06 December 2017

advertisements, law enforcement agencies for detecting illegal activities and burglars for finding out the habits of the occupants of a house [9]. Each IoT device poses a different set of privacy risks and this has led to different privacy preserving solutions. One common approach that exists in literature is that of completely hiding the generated time-series data, either by using data aggregating methods or randomization techniques [10,11]. These approaches perform well in ensuring privacy but they curtail the benefits of installing these IoT devices by excessively protecting an individual's data thus, depriving a user from all the possible set of benefits. The problem then is to ensure that the privacy issue does not limit the access to valuable information thus dampening the data economy and at the same time, the privacy door is not widely open for anyone to extract vital information. The need then is to devise a mechanism that empowers the user with complete control over their information thus leaving it at their disposal to decide on the kind of benefits they want in exchange for their private information. Each IoT device harvests a different level of information and while some devices are very clear in what information they gather, others, such as the smart meters, work in disguise.

Smart meters are advertised as devices that sense and forward the consumption pattern of a household but the time-series data that they generate reveals a lot more information. An adversary can extract vital information such as the living pattern of an individual along with thorough information regarding the appliances installed in the household, such as the time of use and the brand of the washing machines, dryer, kettle, stove, freezer and television [9,12]. It is key to highlight here that, smart meters merely store the power consumption for as small as a single second and then forward it to the utility provider on a predefined data forwarding rate. The accuracy and timeliness of this data is the key to all the benefits related to the smart meters and despite the concerns of a user, the utility provider would not be willing to let anyone alter the average power consumption reported between two data forwarding instances. This bounds the user from adding or subtracting any information thus, limiting the choices to merely distorting the available data. Data distortion can be done using various approaches. One approach is to add an external hardware, where the privacy concerned users install an extra battery that is charged and discharged at irregular intervals, thus generating a distorted consumption pattern [13]. Another approach involves amending the smart meter network structure and ensuring neighbourhood-level aggregation of data before relaying it back to the utility provider (Electric company) [12]. These approaches have been reported to perform well in their considered scenarios but the distortion of data at irregular intervals ensures privacy at the cost of the potential benefits of smart meters.

In this paper, we propose a novel data distortion approach that returns the data sharing authority back to the hands of the user. We formulate a utility-privacy trade-off mechanism that enables the user to decide "what percentage", of "what data", related to "which device" should be shared with the utility provider. The proposed approach uses the unique signature pattern of each appliance to identify its existence in a harvesting interval and then, based on the choice of the user, hides the signature. We use the term harvesting interval as the time between two data forwarding instances. The proposed approach is tested on the ECO (Electricity Consumptions and Occupancy) data set [14] that provides a unique combination of quality and quantity of electricity consumption. In particular, it contains aggregate electricity consumption data, including real and reactive power for each of the three phases and plug level measurements of selected household appliances. The data is being collected at 1 Hz granularity and over a period of 8 months. Despite the size of the data and the huge variation in power consumption pattern of different devices in a household, the proposed approach was successful in correctly identifying appliances with an accuracy of 81.8% and a precision of 70.1%. Simulation results also indicate that the proposed approach is successful in enabling the user to control the privacy of each individual appliance with the aid of a configurable parameter. Furthermore, it is established that the degree of change in the resulting total privacy and utility is proportional to the power consumption of an appliance. The proposed approach is also compared to the existing ElecPrivacy approach [13] where the resultant smart meter readings are tested using two well-known Non-Intrusive Load Monitoring (NILM) algorithms namely: Weiss' [15] algorithm and Baranski's algorithm [16]. The proposed approach has been shown to outperform the existing approach by generating traces that are harder to identify thus demonstrating its effectiveness.

The rest of the paper is organized as follows: in Section 2 we highlight the related work in this field, in Section 3, we formulate the considered problem and in Section 4 we present the proposed approach. In Section 5 we evaluate its performance using simulations and finally in Section 6 we offer our conclusions and future research directions.

2. Related work

The constantly increasing awareness regarding the relationship between smart meter data and privacy has led to some interesting research in this field of study. The research involves proposed approaches stretching between smart meter data aggregation for introduction of anonymity to fixing the problem at the origination point and altering smart meter data before it is read by the smart meter.

Authors in [12,17,18] propose privacy enhancing approaches using neighbourhood level aggregation and cryptographic protocols. The idea is to use cryptography to secure data from being read by neighbouring smart meters and then aggregating the data of multiple smart meters before it is relayed to the utility provider. These algorithms are complemented by the use of verifiable secret sharing algorithms [19,20] to ensure minimum access to the private data by an adversary. Moreover, authors in [21,22] propose the use of secure multi-party computation where, the aggregate smart meter data is computed and released while preserving the confidentiality of each households. Similarly, authors in [23] use data randomization before aggregation, thus completely hiding the energy signature of a single household and then relaying the aggregated data to the utility provider. The baseline assumption is that the utility provider only needs energy consumption reading for a substation. These approaches introduce privacy to an individual's data and offer some utility to the utility provider in understanding the average energy consumption pattern of a certain location but it mitigates the benefits that an individual can obtain from his own smart meter reading, as he cannot identify the appliances which contribute most to the total energy consumption thus undermining his ability to achieve energy efficiency.

To address this problem, authors in [24] focus on distorting the data at the origination point. This would minimize the chances of a malicious node's participation in the smart meter network for both a passive or an active attack and would also enable an individual to keep a constant check of his energy consumption pattern. They propose the use of a stationary Gaussian Markov model for the energy load measurements. They report that privacy-utility trade-off can be optimized through water-filling and for this the privacy mechanism distorts the time-series data off-line after obtaining the whole sequence, thus using memory that increases exponentially with the reduction of the energy harvesting interval. On the other hand, the authors in [25,26] introduce the notion of partial information hiding by introducing uncertainty about individual values in a time-series by perturbing them. A similar notion is presented in [13,27] where it is pointed out that simple data perturbation would be easy to identify by the utility provider as he can identify the actual consumption pattern by installing a similar smart meter at the power origination point. It is thus better to off-load some of the power consumption to batteries at random time intervals. The proposed approach performs well in hiding the consumption pattern at the source but it affects the average consumption per harvesting interval, thus mitigating some benefits of having a smart meter.

Authors in [28,29] highlight the risks of sharing private data and propose appropriate trade-off mechanisms where the user is informed