Pervasive and Mobile Computing 16 (2015) 171–183

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Fast track article

Averting the privacy risks of smart metering by local data preprocessing

Andreas Reinhardt [a,*], Frank Englert [b], Delphine Christin [c]

[a] The University of New South Wales, Sydney, Australia
[b] Technische Universität Darmstadt, Darmstadt, Germany
[c] University of Bonn and Fraunhofer FKIE, Bonn, Germany

ARTICLE INFO

Article history:
Available online 12 October 2014

Keywords:
Power metering
Privacy protection
Data preprocessing

ABSTRACT

More and more renewable sources are integrated into electric power grids worldwide. Their high generation dynamics, however, require power grid operators to monitor electricity generation and demand at a fine temporal resolution. Even small mismatches between supply and demand can impact the power grid's stability, and thus ultimately lead to blackouts. As a result, smart metering equipment has been widely deployed to collect real-time information about the current grid load and forward it to utilities in a timely manner. Numerous research works have shown that power consumption data can, however, reveal the nature of used appliances and their mode of operation at high accuracy. This effectively puts user privacy at risk. In this manuscript, we investigate to which extent the local preprocessing of power data can mitigate this risk. We thus compare the efficacy of different preprocessing steps to eliminate characteristic consumption patterns from the data. Our evaluation shows that a combination of these preprocessing steps can provide a balanced trade-off that is in the interests of both users (privacy protection) and utilities (accurate and timely reporting).

1. Introduction

One of the key elements of future smart power grids is their integration of renewable sources [1]. The volatile nature of renewables, however, introduces previously unseen uncertainties in the electricity generation. Utility companies hence need to constantly maintain up-to-date knowledge about generation and load in order to avert the risk of power outages. Smart electricity meters have been deployed to this end in many countries [2], as they enable to capture both the distributed generation and the demand of a dwelling. While of immediate benefit to the utilities, the transmission of precise information about the current electric activity in households is often perceived as a threat to user privacy. This concern is underpinned by research results that have shown that information about the current user activities and even the television content can be inferred based solely on smart meter data (e.g., [3,4]). So while users may be reluctant to provide high-resolution data because of the possible privacy implications, utilities require exactly this consumption data at a fine temporal resolution in order to adapt the power generation of their non-renewable plants to the dynamically changing demand.

A common way to encounter this problem without forming a trust relationship between customer and utility is the removal of typical characteristics from the data before their transmission. This technique, called *privacy-aware data preprocessing*, has received significant attention in orthogonal domains like participatory sensing [5,6]. However, the applicability

* Correspondence to: School of Computer Science and Engineering, The University of New South Wales, UNSW Sydney NSW 2052, Australia. Tel.: +61 2 9385 7679; fax: +61 2 9385 5995.
E-mail addresses: andreasr@cse.unsw.edu.au (A. Reinhardt), frank.englert@kom.tu-darmstadt.de (F. Englert), christin@cs.uni-bonn.de (D. Christin).

http://dx.doi.org/10.1016/j.pmcj.2014.10.002
1574-1192/

of mechanisms from these domains is very limited due to the different nature of the data collected by smart meters (e.g., the absence of location information). Nonetheless, local preprocessing of sensed data represents a promising way to protect users from potential breaches to their privacy when their consumption data is received by untrusted third parties. In this manuscript we hence investigate to which extent the local preprocessing of power readings can eliminate possibilities to infer appliance types based on their consumption data. To this end, we apply different mechanisms to obfuscate the data and subsequently analyze to which degree appliance types can still be identified after this preprocessing step. The analyzed preprocessing algorithms solely rely on the reporting of slightly altered power consumption readings and do not leverage additional means (e.g., storage batteries [7] or controllable local renewable generation [8]) to physically alter the power demand. It is hence still possible to infer that electrical appliances are operating based on the reported consumption readings. However, when successfully applied, data preprocessing will make it impossible to determine the actual type of an operating appliance or its mode of operation.

Instead of analyzing data that aggregates a complete household's consumption, we focus on distributed smart metering in this manuscript. In this scenario, individual metering devices (sometimes referred to as *smart plugs*) are installed between each appliance's mains plug and the wall outlet. The reasons for selecting this application scenario are twofold. Firstly, existing approaches to infer device activity from smart meter data have shown that the disaggregation of loads performs significantly better when less appliances are connected at the same time [9]. A more efficient privacy protection is thus needed when less appliances are being monitored simultaneously. Secondly, very few household-wide meter data sets (like REDD [10] or Smart* [11]) are freely available. Moreover, these existing data sets are generally neither annotated by the actual appliance activity in the underlying building nor accompanied by the implementation of a disaggregation system. As a result, the effects of local data preprocessing on these data sets cannot be easily determined. In contrast, the Tracebase data set [12] used in this paper contains more than 1500 appliance power consumption traces, and in combination with our previously presented appliance identification system [12] allows for a better generalization of our results.

This manuscript significantly extends our prior publication [13] by analyzing twice as many preprocessors over larger parameter ranges and assessing the introduced errors in a much more detailed manner. It is structured as follows. First, we provide an overview of related work from the domains of data privacy and smart metering in Section 2. Subsequently, we describe our designed software framework and the preprocessing steps in more detail in Section 3. Our evaluation settings are explained in Section 4, followed by the presentation and discussion of our evaluation results in Section 5. Finally, we conclude this paper in Section 6.

2. Related work

The rise of smart meters has led to the availability of energy consumption readings at an unprecedented time and amplitude resolution. To date, two major applications have emerged that rely on these data. Firstly, knowledge about past, current, and expected energy consumption is vital for the smart grid [1], as it allows utilities to take action in order to maintain the grid's stability. Secondly, value-added services can be based on energy consumption data and cater for the creation of smart buildings [14,15].

While smart building functionalities can be realized when accurate measurements are available (cf. [16–19]), the same methods can be applied by third parties (e.g., the utility or external attackers) to infer the current situation in a building. Many institutions like the CEN-CENELEC-ETSI Smart Grid Coordination Group, the National Institute of Standards and Technology, or the German Federal Office for Information Security (BSI) have thus defined information security requirements to the smart grid in [20–22], respectively. Likewise, many researchers have proposed the use of cryptographic means to ensure a secure transport of data between end users and utilities (e.g., [23–25]). Although proposing a separation of personal information and actual power consumption data, countermeasures to prevent inferring user-specific information from meter data are not described in these documents. Moreover, the generally proposed use of pseudonyms has been shown to be ineffective due to the insufficient number of stakeholders on the electricity market [26].

In order to protect users from such intrusions into their privacy, several solutions have thus been presented in the related work. For example, [27,28] show how data collected by multiple meters can be aggregated data before sending them to the utility. Similarly, [29] rely on a virtual ring topology, along which meter readings are relayed before being forwarded. While the users are protected against attacks by legitimate receivers of the data (i.e., utilities) in this case, however, they need to trust and cooperate with other household owners. Moreover, transmissions can experience large delays due to the exchanges between clients that precede the final upload to the data recipient, which may render the approaches inapplicable for the highly dynamic nature of smart power grids.

In comparison to collaborative processing approaches, the local privacy-preserving preprocessing of smart meter data has received significantly less attention in the past. Instead of artificially manipulating the collected readings, existing local approaches mainly rely on the use of external energy storage components. The use of batteries to smoothen the load curve and eliminate characteristic features from the data has been presented in [7,30]. By dynamically adapting the battery output power to a particular appliance's power demand, its existence can be completely hidden. While leading to a potential increase in privacy protection, however, it needs to be remarked that the extent of hiding consumption data this way is inherently limited by the battery capacity. Furthermore, state-of-the-art battery technology suffers from severe limitations, e.g., decreasing capacities over time [31]. Using storage components to protect user privacy may thus not be practical until energy storages become available in large numbers, e.g., as a result of electromobility [32].