Paper Presentation 2 - Privacy in the smart grid

2014-04-08 by Anders Nordin

http://www.eon.se/100koll

Content

- Part 1: Smart Grid Privacy via Anonymization of Smart Metering Data
    - Problem Description
    - Method

- Part 2: Analysis of the impact of data granularity on privacy for the smart grid
    - Problem Description
    - Method

- Part 3: Smart metering de-pseudonymization
    - Problem Description
    - Method

- Comparison / Summarize / Thoughts

Smart Grid Privacy via Anonymization of Smart Metering Data

Costas Efthymiou and Georgios Kalogridis
Toshiba Research Europe Ltd.
Telecommunications Research Laboratory, 32 Queen Square, Bristol, BS1 4ND, UK
{costas, george}@toshiba-trel.com

***Abstract*— The security and privacy of future smart grid and smart metering networks is important to their rollout and eventual acceptance by the public: research in this area is ongoing and smart meter users will need to be reassured that their data is secure. This paper describes a method for securely anonymizing frequent (for example, every few minutes) electrical metering data sent by a smart meter. Although such frequent metering data may be required by a utility or electrical energy distribution network for operational reasons, this data may not necessarily need to be attributable to a specific smart meter or consumer. It does, however, need to be securely attributable to a specific location (e.g. a group of houses or apartments) within the electricity distribution network. The method described in this paper provides a 3rd party escrow mechanism for authenticated anonymous meter readings which are difficult to associate with a particular smart meter or customer. This method does not preclude the provision of attributable metering data that is required for other purposes such as billing, account management or marketing research purposes.**

I. Introduction

Historically, the electrical grid of each country has been a 'broadcast' grid, where a few central power generators (i.e. power stations) produce electricity to cover demand in a country or region, and distribute this electricity to the end users via a large network of cables and transformers. While this model has served well for the last century or so, there is a growing need to reform the world's electrical grids, both from an aging infrastructure point of view and to address new environmental and societal challenges. In response to this need, national governments and relevant stakeholders are making significant efforts in the development of future electrical grids or "Smart Grids"; see examples in [1] and [2]. Development of this new grid will require significant efforts in technology development, standards, policy and regulatory activities because of its inherent complexity. Smart Metering [3] is a key component of the future vision of smart grids.

Security and privacy are considered to be of prime importance to smart grids, given how easily large networks, such as the Internet, can be hacked. This paper focuses on the privacy aspect of smart metering data, discussing its importance and vulnerabilities and proposes a solution for anonymizing high-frequency metering data through the use of a pseudonymous ID without compromising the operations of the utility and/or distribution network.

This paper is organised as follows: Section II briefly discusses the background to metering privacy issues, while Section III presents the problem that this paper addresses. Section IV discusses the escrow-based anonymization process that is proposed to address some of the privacy concerns, with a thorough security analysis of the proposed solution in Section V. Conclusions are drawn in Section VI.

II. Background

A smart meter is an advanced meter (usually an electrical meter, but could also integrate or work together with gas, water and heat meters) that measures energy consumption in much more detail than a conventional meter. Future smart meters will communicate information back to the local utility for monitoring and billing purposes. A smart meter may also potentially communicate with a number of appliances and devices within future 'smart-homes'.

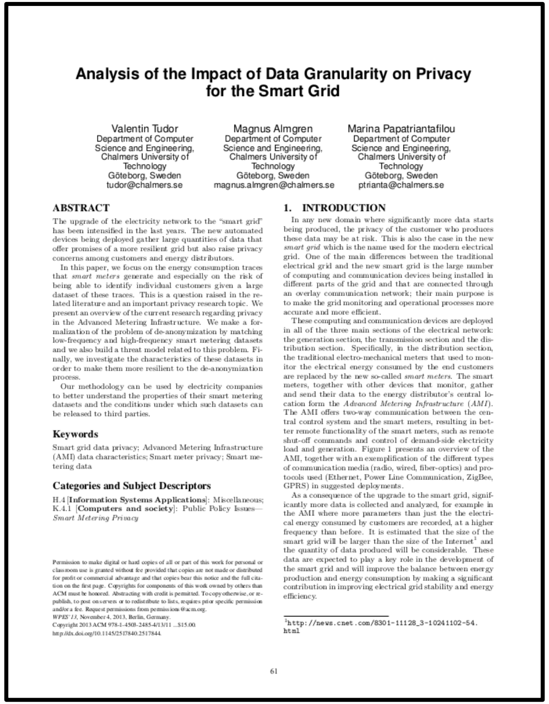
Smart meters are expected to provide accurate readings automatically at requested time intervals to the utility company, electricity distribution network or to the wider 'Smart Grid'. The expected frequency of such readings is yet to be defined; it has been speculated that this could be as high as every few (1-5) minutes, which raises important privacy issues regarding the availability and processing of such data [4]. Such detailed energy usage information could lay bare the daily energy usage patterns of a household and even go so far as to enable deduction of what kind of device or appliance was in use at any given time.

An example of this is given in Fig. 1, reproduced from [4], which discusses these privacy concerns at length with regards to expected and/or projected availability of high-frequency metering data. Another good argument for privacy is given in a recent paper on 'Digital Inclusion' and its ramifications [5].

There is a rich literature in load signature algorithms which use energy measurements to extract detailed information regarding domestic appliance usage. This research is typically termed NALM (Non-intrusive Appliance Load Monitoring), as originally discussed in [6]. There is an active line of research in the construction and upkeep of appliance libraries and detection algorithms; see, for example, [7].

In this paper we address the privacy problem by anonymizing smart metering data so that information gleaned from it cannot easily be associated with an identified person.

Costas Efthymiou and Georgios Kalogridis - Smart Grid Privacy via Anonymization of Smart Metering Data



Analysis of the Impact of Data Granularity on Privacy for the Smart Grid

Valentin Tudor
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
tudor@chalmers.se

Magnus Almgren
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
magnus.almgren@chalmers.se

Marina Papatriantafilou
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
ptrianta@chalmers.se

ABSTRACT

The upgrade of the electricity network to the "smart grid" has been intensified in the last years. The new automated devices being deployed gather large quantities of data that offer promises of a more resilient grid but also raise privacy concerns among customers and energy distributors.

In this paper, we focus on the energy consumption traces that *smart meters* generate and especially on the risk of being able to identify individual customers given a large dataset of these traces. This is a question raised in the related literature and an important privacy research topic. We present an overview of the current research regarding privacy in the Advanced Metering Infrastructure. We make a formalization of the problem of de-anonymization by matching low-frequency and high-frequency smart metering datasets and we also build a threat model related to this problem. Finally, we investigate the characteristics of these datasets in order to make them more resilient to the de-anonymization process.

Our methodology can be used by electricity companies to better understand the properties of their smart metering datasets and the conditions under which such datasets can be released to third parties.

Keywords

Smart grid data privacy; Advanced Metering Infrastructure (AMI) data characteristics; Smart meter privacy; Smart metering data

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; K.4.1 [**Computers and society**]: Public Policy Issues—*Smart Metering Privacy*

1. INTRODUCTION

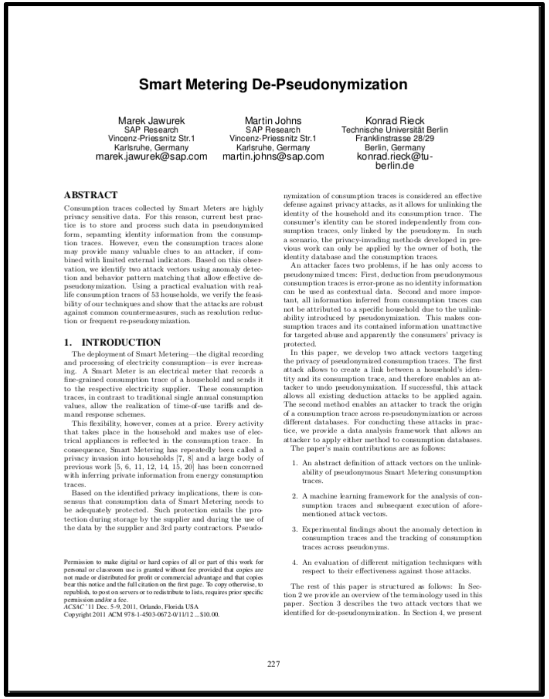
In any new domain where significantly more data starts being produced, the privacy of the customer who produces these data may be at risk. This is also the case in the new *smart grid* which is the name used for the modern electrical grid. One of the main differences between the traditional electrical grid and the new smart grid is the large number of computing and communication devices being installed in different parts of the grid and that are connected through an overlay communication network; their main purpose is to make the grid monitoring and operational processes more accurate and more efficient.

These computing and communication devices are deployed in all of the three main sections of the electrical network: the generation section, the transmission section and the distribution section. Specifically, in the distribution section, the traditional electro-mechanical meters that used to monitor the electrical energy consumed by the end customers are replaced by the new so-called *smart meters*. The smart meters, together with other devices that monitor, gather and send their data to the energy distributor's central location form the *Advanced Metering Infrastructure (AMI)*. The AMI offers two-way communication between the central control system and the smart meters, resulting in better remote functionality of the smart meters, such as remote shut-off commands and control of demand-side electricity load and generation. Figure 1 presents an overview of the AMI, together with an exemplification of the different types of communication media (radio, wired, fiber-optics) and protocols used (Ethernet, Power Line Communication, ZigBee, GPRS) in suggested deployments.

As a consequence of the upgrade to the smart grid, significantly more data is collected and analyzed, for example in the AMI where more parameters than just the the electrical energy consumed by customers are recorded, at a higher frequency than before. It is estimated that the size of the smart grid will be larger than the size of the Internet[1] and the quantity of data produced will be considerable. These data are expected to play a key role in the development of the smart grid and will improve the balance between energy production and energy consumption by making a significant contribution in improving electrical grid stability and energy efficiency.

[1] http://news.cnet.com/8301-11128_3-10241102-54.html

Tudor et al. - Analysis of the impact of data granularity on privacy for the smart grid



Smart Metering De-Pseudonymization

Marek Jawurek
SAP Research
Vincenz-Priessnitz Str.1
Karlsruhe, Germany
marek.jawurek@sap.com

Martin Johns
SAP Research
Vincenz-Priessnitz Str.1
Karlsruhe, Germany
martin.johns@sap.com

Konrad Rieck
Technische Universität Berlin
Franklinstrasse 28/29
Berlin, Germany
konrad.rieck@tu-berlin.de

ABSTRACT

Consumption traces collected by Smart Meters are highly privacy sensitive data. For this reason, current best practice is to store and process such data in pseudonymized form, separating identity information from the consumption traces. However, even the consumption traces alone may provide many valuable clues to an attacker, if combined with limited external indicators. Based on this observation, we identify two attack vectors using anomaly detection and behavior pattern matching that allow effective depseudonymization. Using a practical evaluation with real-life consumption traces of 53 households, we verify the feasibility of our techniques and show that the attacks are robust against common countermeasures, such as resolution reduction or frequent re-pseudonymization.

1. INTRODUCTION

The deployment of Smart Metering—the digital recording and processing of electricity consumption—is ever increasing. A Smart Meter is an electrical meter that records a fine-grained consumption trace of a household and sends it to the respective electricity supplier. These consumption traces, in contrast to traditional single annual consumption values, allow the realization of time-of-use tariffs and demand response schemes.

This flexibility, however, comes at a price. Every activity that takes place in the household and makes use of electrical appliances is reflected in the consumption trace. In consequence, Smart Metering has repeatedly been called a privacy invasion into households [7, 8] and a large body of previous work [5, 6, 11, 12, 14, 15, 20] has been concerned with inferring private information from energy consumption traces.

Based on the identified privacy implications, there is consensus that consumption data of Smart Metering needs to be adequately protected. Such protection entails the protection during storage by the supplier and during the use of the data by the supplier and 3rd party contractors. Pseudonymization of consumption traces is considered an effective defense against privacy attacks, as it allows for unlinking the identity of the household and its consumption trace. The consumer's identity can be stored independently from consumption traces, only linked by the pseudonym. In such a scenario, the privacy-invading methods developed in previous work can only be applied by the owner of both, the identity database and the consumption traces.

An attacker faces two problems, if he has only access to pseudonymized traces: First, deduction from pseudonymous consumption traces is error-prone as no identity information can be used as contextual data. Second and more important, all information inferred from consumption traces can not be attributed to a specific household due to the unlinkability introduced by pseudonymization. This makes consumption traces and its contained information unattractive for targeted abuse and apparently the consumers' privacy is protected.

In this paper, we develop two attack vectors targeting the privacy of pseudonymized consumption traces. The first attack allows to create a link between a household's identity and its consumption trace, and therefore enables an attacker to undo pseudonymization. If successful, this attack allows all existing deduction attacks to be applied again. The second method enables an attacker to track the origin of a consumption trace across re-pseudonymization or across different databases. For conducting these attacks in practice, we provide a data analysis framework that allows an attacker to apply either method to consumption databases.

The paper's main contributions are as follows:

1. An abstract definition of attack vectors on the unlinkability of pseudonymous Smart Metering consumption traces.
2. A machine learning framework for the analysis of consumption traces and subsequent execution of aforementioned attack vectors.
3. Experimental findings about the anomaly detection in consumption traces and the tracking of consumption traces across pseudonyms.
4. An evaluation of different mitigation techniques with respect to their effectiveness against those attacks.

The rest of this paper is structured as follows: In Section 2 we provide an overview of the terminology used in this paper. Section 3 describes the two attack vectors that we identified for de-pseudonymization. In Section 4, we present

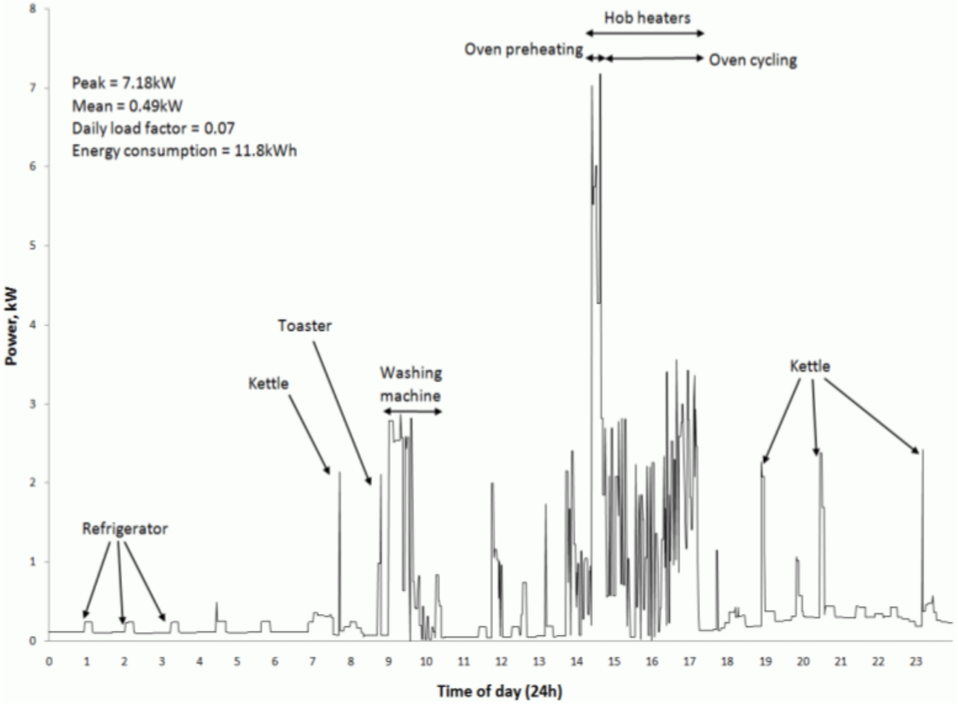
Jawurek et al. - Smart metering de-pseudonymization

Problem Description

- “High Frequency” metering data.
  - About every 5 minute
  - Electric data from home

- “Low Frequency” metering data.
  - Weekly/Monthly
  - Meter reading for billing

How can we anonymize high frequency data?



Picture: E. L. Quinn, “Privacy and the New Energy Infrastructure”, Social
Science Research Network (SSRN), February 2009

Method(1)

HFID = High Frequency ID

LFID = Low Frequency ID

- HFID should never be known to the power company or the smart meter installer
- HFID hardcoded by the manufacturer
  - 3rd party escrow
  - Manufacturer is not expected to manage any data
  - Manufacturer requires a strong data privacy policy to ensure the secret of the relation between LFID and HFID
- Secure protocol setup mechanism
- The protocol is not perfect w.r.t privacy protection but described as a step in the right direction

Method(2)

- Client Data Profile(CDP)
  - Initial process done to identify the client
  - Client <-> Power Company
  - LFID included
- Anonymous Data Profile(ADP)
  - Initiated after the CDP process.
  - Power Company <-> Escrow
  - Escrow <-> Client
  - HFID included

Smart Grid Privacy via Anonymization of Smart Metering Data

Costas Efthymiou and Georgios Kalogridis

Toshiba Research Europe Ltd.
Telecommunications Research Laboratory, 32 Queen Square, Bristol, BS1 4ND, UK
{costas, george}@toshiba-trel.com

***Abstract*— The security and privacy of future smart grid and smart metering networks is important to their rollout and eventual acceptance by the public: research in this area is ongoing and smart meter users will need to be reassured that their data is secure. This paper describes a method for securely anonymizing frequent (for example, every few minutes) electrical metering data sent by a smart meter. Although such frequent metering data may be required by a utility or electrical energy distribution network for operational reasons, this data may not necessarily need to be attributable to a specific smart meter or consumer. It does, however, need to be securely attributable to a specific location (e.g. a group of houses or apartments) within the electricity distribution network. The method described in this paper provides a 3rd party escrow mechanism for authenticated anonymous meter readings which are difficult to associate with a particular smart meter or customer. This method does not preclude the provision of attributable metering data that is required for other purposes such as billing, account management or marketing research purposes.**

I. Introduction

Historically, the electrical grid of each country has been a 'broadcast' grid, where a few central power generators (i.e. power stations) produce electricity to cover demand in a country or region, and distribute this electricity to the end users via a large network of cables and transformers. While this model has served well for the last century or so, there is a growing need to reform the world's electrical grids, both from an aging infrastructure point of view and to address new environmental and societal challenges. In response to this need, national governments and relevant stakeholders are making significant efforts in the development of future electrical grids or "Smart Grids"; see examples in [1] and [2]. Development of this new grid will require significant efforts in technology development, standards, policy and regulatory activities because of its inherent complexity. Smart Metering [3] is a key component of the future vision of smart grids.

Security and privacy are considered to be of prime importance to smart grids, given how easily large networks, such as the Internet, can be hacked. This paper focuses on the privacy aspect of smart metering data, discussing its importance and vulnerabilities and proposes a solution for anonymizing high-frequency metering data through the use of a pseudonymous ID without compromising the operations of the utility and/or distribution network.

This paper is organised as follows: Section II briefly discusses the background to metering privacy issues, while Section III presents the problem that this paper addresses. Section IV discusses the escrow-based anonymization process that is proposed to address some of the privacy concerns, with a thorough security analysis of the proposed solution in Section V. Conclusions are drawn in Section VI.

II. Background

A smart meter is an advanced meter (usually an electrical meter, but could also integrate or work together with gas, water and heat meters) that measures energy consumption in much more detail than a conventional meter. Future smart meters will communicate information back to the local utility for monitoring and billing purposes. A smart meter may also potentially communicate with a number of appliances and devices within future 'smart-homes'.

Smart meters are expected to provide accurate readings automatically at requested time intervals to the utility company, electricity distribution network or to the wider 'Smart Grid'. The expected frequency of such readings is yet to be defined; it has been speculated that this could be as high as every few (1-5) minutes, which raises important privacy issues regarding the availability and processing of such data [4]. Such detailed energy usage information could lay bare the daily energy usage patterns of a household and even go so far as to enable deduction of what kind of device or appliance was in use at any given time.

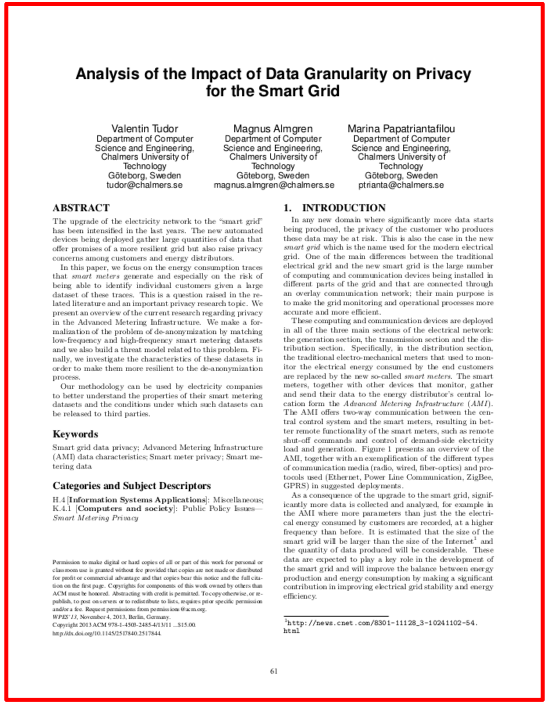
An example of this is given in Fig. 1, reproduced from [4], which discusses these privacy concerns at length with regards to expected and/or projected availability of high-frequency metering data. Another good argument for privacy is given in a recent paper on 'Digital Inclusion' and its ramifications [5].

There is a rich literature in load signature algorithms which use energy measurements to extract detailed information regarding domestic appliance usage. This research is typically termed NALM (Non-intrusive Appliance Load Monitoring), as originally discussed in [6]. There is an active line of research in the construction and upkeep of appliance libraries and detection algorithms; see, for example, [7].

In this paper we address the privacy problem by anonymizing smart metering data so that information gleaned from it cannot easily be associated with an identified person.

978-1-4244-6511-8/10/$26.00 ©2010 IEEE

Costas Efthymiou and Georgios Kalogridis - Smart Grid Privacy via Anonymization of Smart Metering Data



Analysis of the Impact of Data Granularity on Privacy for the Smart Grid

Valentin Tudor
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
tudor@chalmers.se

Magnus Almgren
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
magnus.almgren@chalmers.se

Marina Papatriantafilou
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
ptrianta@chalmers.se

ABSTRACT

The upgrade of the electricity network to the "smart grid" has been intensified in the last years. The new automated devices being deployed gather large quantities of data that offer promises of a more resilient grid but also raise privacy concerns among customers and energy distributors.

In this paper, we focus on the energy consumption traces that *smart meters* generate and especially on the risk of being able to identify individual customers given a large dataset of these traces. This is a question raised in the related literature and an important privacy research topic. We present an overview of the current research regarding privacy in the Advanced Metering Infrastructure. We make a formalization of the problem of de-anonymization by matching low-frequency and high-frequency smart metering datasets and we also build a threat model related to this problem. Finally, we investigate the characteristics of these datasets in order to make them more resilient to the de-anonymization process.

Our methodology can be used by electricity companies to better understand the properties of their smart metering datasets and the conditions under which such datasets can be released to third parties.

Keywords

Smart grid data privacy; Advanced Metering Infrastructure (AMI) data characteristics; Smart meter privacy; Smart metering data

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; K.4.1 [**Computers and society**]: Public Policy Issues—*Smart Metering Privacy*

WPES'13, November 4, 2013, Berlin, Germany.
Copyright 2013 ACM 978-1-4503-2485-4/13/11 ...$15.00.
http://dx.doi.org/10.1145/2517840.2517844.

1. INTRODUCTION

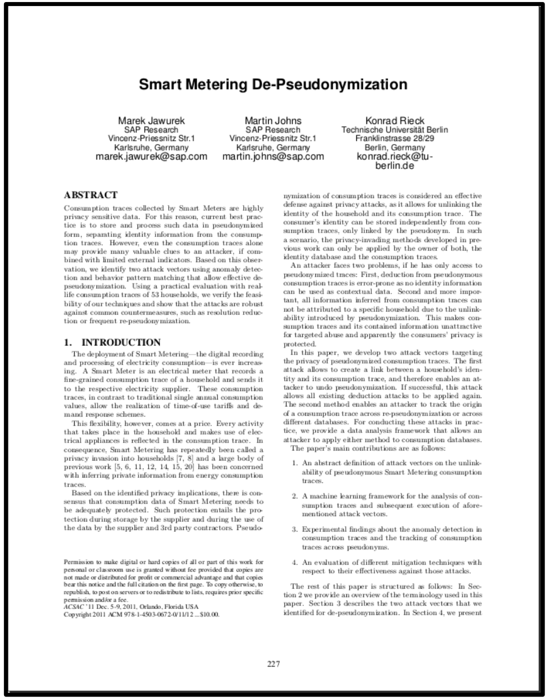
In any new domain where significantly more data starts being produced, the privacy of the customer who produces these data may be at risk. This is also the case in the new *smart grid* which is the name used for the modern electrical grid. One of the main differences between the traditional electrical grid and the new smart grid is the large number of computing and communication devices being installed in different parts of the grid and that are connected through an *overlay* communication network; their main purpose is to make the grid monitoring and operational processes more accurate and more efficient.

These computing and communication devices are deployed in all of the three main sections of the electrical network: the generation section, the transmission section and the distribution section. Specifically, in the distribution section, the traditional electro-mechanical meters that used to monitor the electrical energy consumed by the end customers are replaced by the new so-called *smart meters*. The smart meters, together with other devices that monitor, gather and send their data to the energy distributor's central location form the *Advanced Metering Infrastructure (AMI)*. The AMI offers two-way communication between the central control system and the smart meters, resulting in better remote functionality of the smart meters, such as remote shut-off commands and control of demand-side electricity load and generation. Figure 1 presents an overview of the AMI, together with an exemplification of the different types of communication media (radio, wired, fiber-optics) and protocols used (Ethernet, Power Line Communication, ZigBee, GPRS) in suggested deployments.

As a consequence of the upgrade to the smart grid, significantly more data is collected and analyzed, for example in the AMI where more parameters than just the the electrical energy consumed by customers are recorded, at a higher frequency than before. It is estimated that the size of the smart grid will be larger than the size of the Internet[1] and the quantity of data produced will be considerable. These data are expected to play a key role in the development of the smart grid and will improve the balance between energy production and energy consumption by making a significant contribution in improving electrical grid stability and energy efficiency.

[1] `http://news.cnet.com/8301-11128_3-10241102-54.html`

Tudor et al. - Analysis of the impact of data granularity on privacy for the smart grid



Smart Metering De-Pseudonymization

Marek Jawurek
SAP Research
Vincenz-Priessnitz Str.1
Karlsruhe, Germany
marek.jawurek@sap.com

Martin Johns
SAP Research
Vincenz-Priessnitz Str.1
Karlsruhe, Germany
martin.johns@sap.com

Konrad Rieck
Technische Universität Berlin
Franklinstrasse 28/29
Berlin, Germany
konrad.rieck@tu-berlin.de

ABSTRACT

Consumption traces collected by Smart Meters are highly privacy sensitive data. For this reason, current best practice is to store and process such data in pseudonymized form, separating identity information from the consumption traces. However, even the consumption traces alone may provide many valuable clues to an attacker, if combined with limited external indicators. Based on this observation, we identify two attack vectors using anomaly detection and behavior pattern matching that allow effective de-pseudonymization. Using a practical evaluation with real-life consumption traces of 53 households, we verify the feasibility of our techniques and show that the attacks are robust against common countermeasures, such as resolution reduction or frequent re-pseudonymization.

1. INTRODUCTION

The deployment of Smart Metering—the digital recording and processing of electricity consumption—is ever increasing. A Smart Meter is an electrical meter that records a fine-grained consumption trace of a household and sends it to the respective electricity supplier. These consumption traces, in contrast to traditional single annual consumption values, allow the realization of time-of-use tariffs and demand response schemes.

This flexibility, however, comes at a price. Every activity that takes place in the household and makes use of electrical appliances is reflected in the consumption trace. In consequence, Smart Metering has repeatedly been called a privacy invasion into households [7, 8] and a large body of previous work [5, 6, 11, 12, 14, 15, 20] has been concerned with inferring private information from energy consumption traces.

Based on the identified privacy implications, there is consensus that consumption data of Smart Metering needs to be adequately protected. Such protection entails the protection during storage by the supplier and during the use of the data by the supplier and 3rd party contractors. Pseudonymization of consumption traces is considered an effective defense against privacy attacks, as it allows for unlinking the identity of the household and its consumption trace. The consumer's identity can be stored independently from consumption traces, only linked by the pseudonym. In such a scenario, the privacy-invading methods developed in previous work can only be applied by the owner of both, the identity database and the consumption traces.

An attacker faces two problems, if he has only access to pseudonymized traces: First, deduction from pseudonymous consumption traces is error-prone as no identity information can be used as contextual data. Second and more important, all information inferred from consumption traces can not be attributed to a specific household due to the unlinkability introduced by pseudonymization. This makes consumption traces and its contained information unattractive for targeted abuse and apparently the consumers' privacy is protected.

In this paper, we develop two attack vectors targeting the privacy of pseudonymized consumption traces. The first attack allows to create a link between a household's identity and its consumption trace, and therefore enables an attacker to undo pseudonymization. If successful, this attack allows all existing deduction attacks to be applied again. The second method enables an attacker to track the origin of a consumption trace across re-pseudonymization or across different databases. For conducting these attacks in practice, we provide a data analysis framework that allows an attacker to apply either method to consumption databases.

The paper's main contributions are as follows:

1. An abstract definition of attack vectors on the unlinkability of pseudonymous Smart Metering consumption traces.
2. A machine learning framework for the analysis of consumption traces and subsequent execution of aforementioned attack vectors.
3. Experimental findings about the anomaly detection in consumption traces and the tracking of consumption traces across pseudonyms.
4. An evaluation of different mitigation techniques with respect to their effectiveness against those attacks.

The rest of this paper is structured as follows: In Section 2 we provide an overview of the terminology used in this paper. Section 3 describes the two attack vectors that we identified for de-pseudonymization. In Section 4, we present

ACSAC '11 Dec. 5-9, 2011, Orlando, Florida USA
Copyright 2011 ACM 978-1-4503-0672-0/11/12 ...$10.00.

Jawurek et al. - Smart metering de-pseudonymization

Problem Description

- Matching high-frequent data with low-frequent data => Customer Identity
- Sum(High Frequent Data for Time Period) = Low Frequent data

Method

- What if the granularity is rounded to every 10 kWh instead of 1 kWh

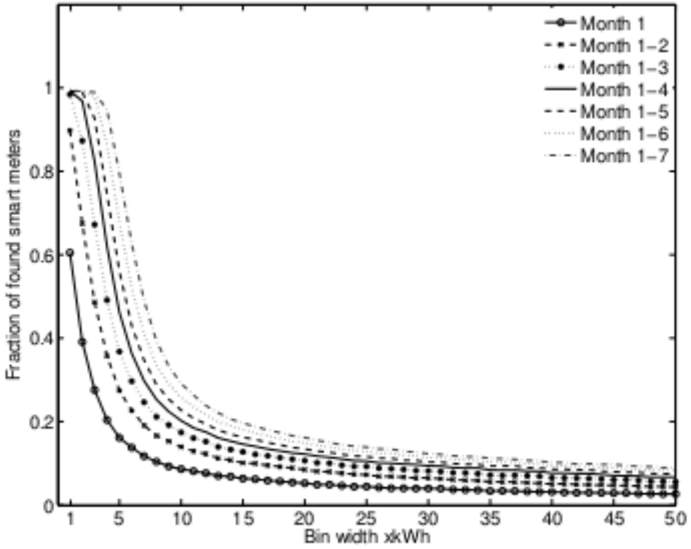
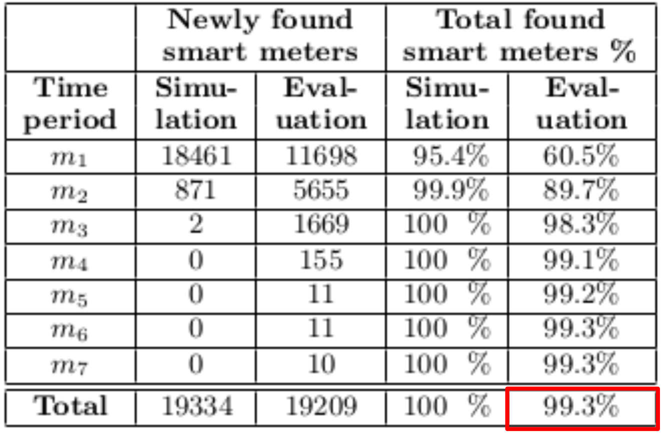
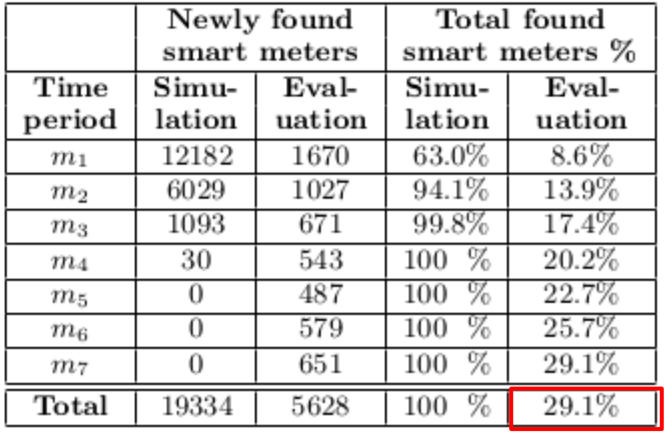


Figure 5: Fraction of unique smart meters - seven months of data - dataset case



	Newly found smart meters		**Total found smart meters %**	
Time period	**Simulation**	**Evaluation**	**Simulation**	**Evaluation**
m_1	18461	11698	95.4%	60.5%
m_2	871	5655	99.9%	89.7%
m_3	2	1669	100 %	98.3%
m_4	0	155	100 %	99.1%
m_5	0	11	100 %	99.2%
m_6	0	11	100 %	99.3%
m_7	0	10	100 %	99.3%
Total	19334	19209	100 %	99.3%

Table 3: Expected number of identified smart meters for a reporting granularity of 1 kWh



	Newly found smart meters		**Total found smart meters %**	
Time period	**Simulation**	**Evaluation**	**Simulation**	**Evaluation**
m_1	12182	1670	63.0%	8.6%
m_2	6029	1027	94.1%	13.9%
m_3	1093	671	99.8%	17.4%
m_4	30	543	100 %	20.2%
m_5	0	487	100 %	22.7%
m_6	0	579	100 %	25.7%
m_7	0	651	100 %	29.1%
Total	19334	5628	100 %	29.1%

Table 4: Expected number of identified smart meters for a reporting granularity of 10 kWh

Smart Grid Privacy via Anonymization of Smart Metering Data

Costas Efthymiou and Georgios Kalogridis

Toshiba Research Europe Ltd.
Telecommunications Research Laboratory, 32 Queen Square, Bristol, BS1 4ND, UK
{costas, george}@toshiba-trel.com

***Abstract*— The security and privacy of future smart grid and smart metering networks is important to their rollout and eventual acceptance by the public: research in this area is ongoing and smart meter users will need to be reassured that their data is secure. This paper describes a method for securely anonymizing frequent (for example, every few minutes) electrical metering data sent by a smart meter. Although such frequent metering data may be required by a utility or electrical energy distribution network for operational reasons, this data may not necessarily need to be attributable to a specific smart meter or consumer. It does, however, need to be securely attributable to a specific location (e.g. a group of houses or apartments) within the electricity distribution network. The method described in this paper provides a 3rd party escrow mechanism for authenticated anonymous meter readings which are difficult to associate with a particular smart meter or customer. This method does not preclude the provision of attributable metering data that is required for other purposes such as billing, account management or marketing research purposes.**

I. Introduction

Historically, the electrical grid of each country has been a 'broadcast' grid, where a few central power generators (i.e. power stations) produce electricity to cover demand in a country or region, and distribute this electricity to the end users via a large network of cables and transformers. While this model has served well for the last century or so, there is a growing need to reform the world's electrical grids, both from an aging infrastructure point of view and to address new environmental and societal challenges. In response to this need, national governments and relevant stakeholders are making significant efforts in the development of future electrical grids or "Smart Grids"; see examples in [1] and [2]. Development of this new grid will require significant efforts in technology development, standards, policy and regulatory activities because of its inherent complexity. Smart Metering [3] is a key component of the future vision of smart grids.

Security and privacy are considered to be of prime importance to smart grids, given how easily large networks, such as the Internet, can be hacked. This paper focuses on the privacy aspect of smart metering data, discussing its importance and vulnerabilities and proposes a solution for anonymizing high-frequency metering data through the use of a pseudonymous ID without compromising the operations of the utility and/or distribution network.

This paper is organised as follows: Section II briefly discusses the background to metering privacy issues, while Section III presents the problem that this paper addresses. Section IV discusses the escrow-based anonymization process that is proposed to address some of the privacy concerns, with a thorough security analysis of the proposed solution in Section V. Conclusions are drawn in Section VI.

II. Background

A smart meter is an advanced meter (usually an electrical meter, but could also integrate or work together with gas, water and heat meters) that measures energy consumption in much more detail than a conventional meter. Future smart meters will communicate information back to the local utility for monitoring and billing purposes. A smart meter may also potentially communicate with a number of appliances and devices within future 'smart-homes'.

Smart meters are expected to provide accurate readings automatically at requested time intervals to the utility company, electricity distribution network or to the wider 'Smart Grid'. The expected frequency of such readings is yet to be defined; it has been speculated that this could be as high as every few (1-5) minutes, which raises important privacy issues regarding the availability and processing of such data [4]. Such detailed energy usage information could lay bare the daily energy usage patterns of a household and even go so far as to enable deduction of what kind of device or appliance was in use at any given time.

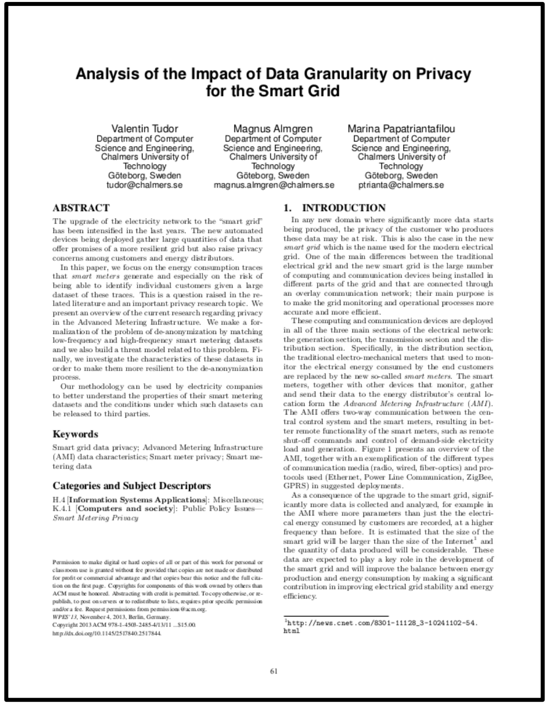
An example of this is given in Fig. 1, reproduced from [4], which discusses these privacy concerns at length with regards to expected and/or projected availability of high-frequency metering data. Another good argument for privacy is given in a recent paper on 'Digital Inclusion' and its ramifications [5].

There is a rich literature in load signature algorithms which use energy measurements to extract detailed information regarding domestic appliance usage. This research is typically termed NALM (Non-intrusive Appliance Load Monitoring), as originally discussed in [6]. There is an active line of research in the construction and upkeep of appliance libraries and detection algorithms; see, for example, [7].

In this paper we address the privacy problem by anonymizing smart metering data so that information gleaned from it cannot easily be associated with an identified person.

978-1-4244-6511-8/10/$26.00 ©2010 IEEE

Costas Efthymiou and Georgios Kalogridis - Smart Grid Privacy via Anonymization of Smart Metering Data



Analysis of the Impact of Data Granularity on Privacy for the Smart Grid

Valentin Tudor
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
tudor@chalmers.se

Magnus Almgren
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
magnus.almgren@chalmers.se

Marina Papatriantafilou
Department of Computer Science and Engineering, Chalmers University of Technology
Göteborg, Sweden
ptrianta@chalmers.se

ABSTRACT

The upgrade of the electricity network to the "smart grid" has been intensified in the last years. The new automated devices being deployed gather large quantities of data that offer promises of a more resilient grid but also raise privacy concerns among customers and energy distributors.

In this paper, we focus on the energy consumption traces that *smart meters* generate and especially on the risk of being able to identify individual customers given a large dataset of these traces. This is a question raised in the related literature and an important privacy research topic. We present an overview of the current research regarding privacy in the Advanced Metering Infrastructure. We make a formalization of the problem of de-anonymization by matching low-frequency and high-frequency smart metering datasets and we also build a threat model related to this problem. Finally, we investigate the characteristics of these datasets in order to make them more resilient to the de-anonymization process.

Our methodology can be used by electricity companies to better understand the properties of their smart metering datasets and the conditions under which such datasets can be released to third parties.

Keywords

Smart grid data privacy; Advanced Metering Infrastructure (AMI) data characteristics; Smart meter privacy; Smart metering data

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; K.4.1 [**Computers and society**]: Public Policy Issues—*Smart Metering Privacy*

WPES'13, November 4, 2013, Berlin, Germany.
Copyright 2013 ACM 978-1-4503-2485-4/13/11 ...$15.00.
http://dx.doi.org/10.1145/2517840.2517844.

1. INTRODUCTION

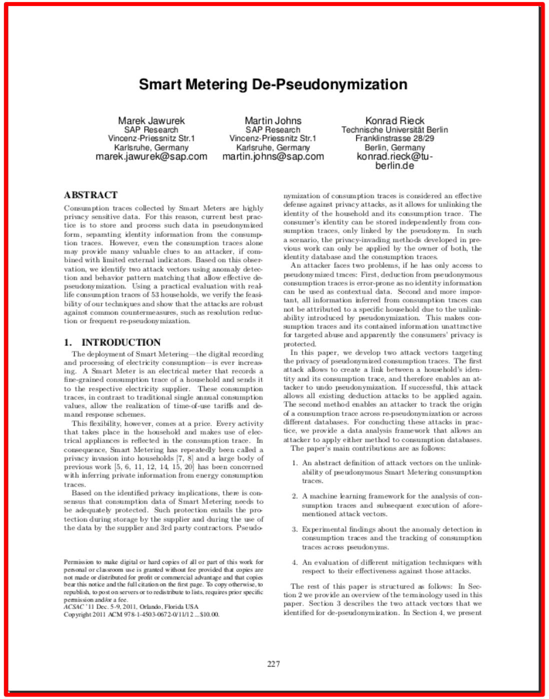
In any new domain where significantly more data starts being produced, the privacy of the customer who produces these data may be at risk. This is also the case in the new *smart grid* which is the name used for the modern electrical grid. One of the main differences between the traditional electrical grid and the new smart grid is the large number of computing and communication devices being installed in different parts of the grid and that are connected through an *overlay* communication network; their main purpose is to make the grid monitoring and operational processes more accurate and more efficient.

These computing and communication devices are deployed in all of the three main sections of the electrical network: the generation section, the transmission section and the distribution section. Specifically, in the distribution section, the traditional electro-mechanical meters that used to monitor the electrical energy consumed by the end customers are replaced by the new so-called *smart meters*. The smart meters, together with other devices that monitor, gather and send their data to the energy distributor's central location form the *Advanced Metering Infrastructure* (*AMI*). The AMI offers two-way communication between the central control system and the smart meters, resulting in better remote functionality of the smart meters, such as remote shut-off commands and control of demand-side electricity load and generation. Figure 1 presents an overview of the AMI, together with an exemplification of the different types of communication media (radio, wired, fiber-optics) and protocols used (Ethernet, Power Line Communication, ZigBee, GPRS) in suggested deployments.

As a consequence of the upgrade to the smart grid, significantly more data is collected and analyzed, for example in the AMI where more parameters than just the the electrical energy consumed by customers are recorded, at a higher frequency than before. It is estimated that the size of the smart grid will be larger than the size of the Internet[1] and the quantity of data produced will be considerable. These data are expected to play a key role in the development of the smart grid and will improve the balance between energy production and energy consumption by making a significant contribution in improving electrical grid stability and energy efficiency.

[1] `http://news.cnet.com/8301-11128_3-10241102-54.html`

Tudor et al. - Analysis of the impact of data granularity on privacy for the smart grid



Smart Metering De-Pseudonymization

Marek Jawurek
SAP Research
Vincenz-Priessnitz Str.1
Karlsruhe, Germany
marek.jawurek@sap.com

Martin Johns
SAP Research
Vincenz-Priessnitz Str.1
Karlsruhe, Germany
martin.johns@sap.com

Konrad Rieck
Technische Universität Berlin
Franklinstrasse 28/29
Berlin, Germany
konrad.rieck@tu-berlin.de

ABSTRACT

Consumption traces collected by Smart Meters are highly privacy sensitive data. For this reason, current best practice is to store and process such data in pseudonymized form, separating identity information from the consumption traces. However, even the consumption traces alone may provide many valuable clues to an attacker, if combined with limited external indicators. Based on this observation, we identify two attack vectors using anomaly detection and behavior pattern matching that allow effective de-pseudonymization. Using a practical evaluation with real-life consumption traces of 53 households, we verify the feasibility of our techniques and show that the attacks are robust against common countermeasures, such as resolution reduction or frequent re-pseudonymization.

1. INTRODUCTION

The deployment of Smart Metering—the digital recording and processing of electricity consumption—is ever increasing. A Smart Meter is an electrical meter that records a fine-grained consumption trace of a household and sends it to the respective electricity supplier. These consumption traces, in contrast to traditional single annual consumption values, allow the realization of time-of-use tariffs and demand response schemes.

This flexibility, however, comes at a price. Every activity that takes place in the household and makes use of electrical appliances is reflected in the consumption trace. In consequence, Smart Metering has repeatedly been called a privacy invasion into households [7, 8] and a large body of previous work [5, 6, 11, 12, 14, 15, 20] has been concerned with inferring private information from energy consumption traces.

Based on the identified privacy implications, there is consensus that consumption data of Smart Metering needs to be adequately protected. Such protection entails the protection during storage by the supplier and during the use of the data by the supplier and 3rd party contractors. Pseudonymization of consumption traces is considered an effective defense against privacy attacks, as it allows for unlinking the identity of the household and its consumption trace. The consumer's identity can be stored independently from consumption traces, only linked by the pseudonym. In such a scenario, the privacy-invading methods developed in previous work can only be applied by the owner of both, the identity database and the consumption traces.

An attacker faces two problems, if he has only access to pseudonymized traces: First, deduction from pseudonymous consumption traces is error-prone as no identity information can be used as contextual data. Second and more important, all information inferred from consumption traces can not be attributed to a specific household due to the unlinkability introduced by pseudonymization. This makes consumption traces and its contained information unattractive for targeted abuse and apparently the consumers' privacy is protected.

In this paper, we develop two attack vectors targeting the privacy of pseudonymized consumption traces. The first attack allows to create a link between a household's identity and its consumption trace, and therefore enables an attacker to undo pseudonymization. If successful, this attack allows all existing deduction attacks to be applied again. The second method enables an attacker to track the origin of a consumption trace across re-pseudonymization or across different databases. For conducting these attacks in practice, we provide a data analysis framework that allows an attacker to apply either method to consumption databases.

The paper's main contributions are as follows:

1. An abstract definition of attack vectors on the unlinkability of pseudonymous Smart Metering consumption traces.
2. A machine learning framework for the analysis of consumption traces and subsequent execution of aforementioned attack vectors.
3. Experimental findings about the anomaly detection in consumption traces and the tracking of consumption traces across pseudonyms.
4. An evaluation of different mitigation techniques with respect to their effectiveness against those attacks.

The rest of this paper is structured as follows: In Section 2 we provide an overview of the terminology used in this paper. Section 3 describes the two attack vectors that we identified for de-pseudonymization. In Section 4, we present

ACSAC '11 Dec. 5-9, 2011, Orlando, Florida USA
Copyright 2011 ACM 978-1-4503-0672-0/11/12 ...$10.00.

Jawurek et al. - Smart metering de-pseudonymization

Two types of attack

Linking by behaviour anomaly

Unique event creates a peak or valley in the consumption trace

Linking by Behavior Pattern

Tracks the origin of a consumption trace

- Multiple pseudonyms
- Multiple databases

Possible ways to protect against the attacks

- Create new pseudonyms more often to confuse the attacker and harder to track
  - Overhead for storage
  - Maybe the attacker can follow the trace anyway?

- Lower Resolution of Smart metering
  - Proved in the paper that the linking accuracy drops significantly

Not discussed in the papers

- Proper protection during storage of the data

- Cryptographic methods

- **Politics:** Under what circumstances should the identity be revealed?
  - Court order, police suspect something illegal
  - Employer spy on workers who called in sick
  - Power theft

Questions?