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

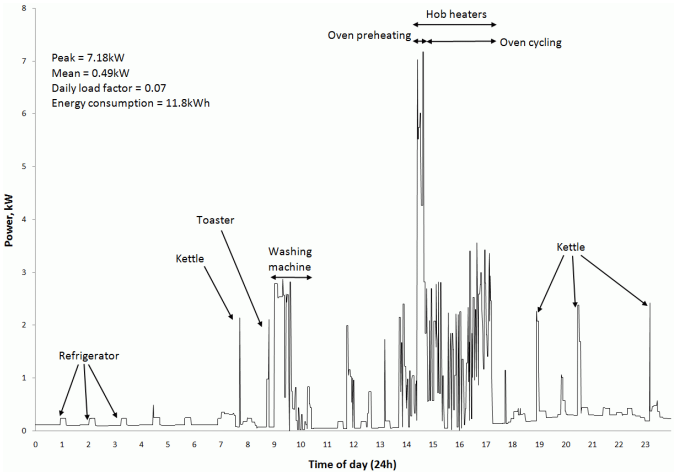

Fig. 1. Household electricity demand profile recorded on a one-minute time base, reproduced from [4]

III. Problem Statement and Assumptions

A. Problem Statement

The main dilemma and, at the same time, challenge this paper addresses is that high-frequency metering data which are required for efficient network operations, e.g. managing load, Demand Side Response and Management and load shedding or shifting, may expose private information. Assuming that metering data needed for utility purposes, e.g. billing and account management purposes, are communicated at lower frequencies, the main question that this paper addresses becomes: *How can high-frequency data be anonymized, i.e. not be attributable to a specific smart meter/home user, without negatively affecting network operations or the availability of high-frequency metering data?*

The authors recognise the advanced level of NALM algorithms and the method described here does not attempt to counter their operation directly. Instead, the method proposed in this paper can offer anonymity to the high-frequency energy measurement data, thus offering an increased level of privacy to the user.

B. Assumptions

We make the following assumptions:

1. Metering data needed for billing or account management purposes needs to be attributable – i.e. securely attached to a particular consumer and/or account holder with a utility.

2. Attributable metering data will typically be collected at low frequency, i.e. daily/weekly/monthly/quarterly.

3. Metering data needed for power generation and distribution network control doesn't need to be attributable. Anonymous data can suffice if it can be authenticated and if it can be securely tied with a particular distribution sub-station or other entity which controls a small sub-set of electrical energy consumers (for example, a sub-station that supplies and/or controls the electrical energy for a single street of houses or a block of apartments).

4. Anonymous data will be collected at high-frequency, i.e. every minute/five minutes to enable near real-time response to power quality or demand response issues within the electrical energy distribution network.

5. The smallest 'unit' of electrical energy consumers that needs to be known to an electrical distribution network is a distribution sub-station or any other entity which forms part of the electrical distribution network and which directly supplies energy consumers (e.g. home smart meters). The utility does not need to know which smart meter generates specific data – it is only interested in the aggregator and the substation these smart metering data are associated with. However, the utility should still be able to remotely micro-manage individual smart meters if needed (e.g. demand-side management), even if these smart meters remain anonymous.

6. An adequate trust relationship is present between 3rd party escrow service providers, utility companies and their customers. More specifically, escrow service providers should be trusted not to provide customer information to utility companies (unless this is approved by higher authorities and deemed necessary), and should also be trusted to securely provide anonymization credentials that cannot be repudiated.

The authors recognize that the distinction between attributable and anonymized data may either include or preclude metering data from microgeneration, e.g. a rooftop solar panel, depending on local policy and implementation.

IV. Escrow-Based Anonymization

A. Definitions

A distinction is made between two types of data that a smart meter needs to generate:

i. 'High-frequency' metering data, which are the meter readings a smart meter transmits to the utility often enough (e.g. every few minutes) to suggest information related with the electrical data user's private life (e.g. usage patterns of specific electrical appliances).

ii. 'Low-frequency' metering data, which are the meter readings a smart meter transmits to the utility scarcely enough (e.g. every week or month) to offer adequate privacy, e.g. meter readings for account management or billing purposes.

B. Architecture

The main structural difference to a smart meter introduced in this paper is that of having two separate IDs embedded in the smart meter, rather than a single ID as is the case with standard smart meters:

i. HFID, or High-Frequency ID (anonymous)

ii. LFID, or Low-Frequency ID (attributable).

These two IDs are attached to metering-related messages that are transmitted from the smart meter to the utility for the high-frequency and low-frequency metering data. In practice, this will mean that the vast majority of meter readings will be sent using the HFID.

The key concept here is the method for providing anonymity of the HFID, and by extension, to the HFID messages (metering data). The only way for the HFID to be anonymous from the start, and to remain so, is for it to never be known to the utility or the smart meter installer. The HFID

TABLE 1: ACRONYMS USED

Term	Definition
ADP	Anonymous Data Profile
AGG	Aggregation (Data aggregator process which chooses a suitable data aggregator or concentrator local to the smart meter. The aggregator collects metering data from its local area smart meters and sends them on to the utility, distribution network or other relevant parties)
ANSM	Anonymous SM Profile (Contains: Certificate, Private Key)
CDP	Client Data Profile
CERT	Certificate (Contains: ID, Public Key)
CL	Client (the 'user' of the smart meter, e.g. a home resident)
CLI	Client Information
ESC	Escrow (or Trusted Third Party - TTP)
HF	High Frequency (Metering data)
LF	Low Frequency (Metering data)
PDNet	Power Distribution Network
PISM	Personally Identifiable SM Profile (Contains: Certificate, Private Key)
SM	Smart Meter
U	Utility (company that gathers smart metering data for billing, account management, demand side response, demand side management activities, etc.)

can be 'hidden' inside the smart meter, hard-coded to be used for all HFID-related messages. There arises the problem of how a utility can be sure that messages being received from a specific HFID can be authenticated, i.e. verified to be legitimate, as the utility will not know which HFIDs are valid.

This is where the idea of the 3rd party escrow service is introduced – this 3rd party can be the manufacturer of the smart meter itself or some other trusted 3rd party which has been given access to this information. The manufacturer can assign two unique IDs (just like MAC addresses for IEEE 802.x devices are unique, for example) to each smart meter that is produced, only one of which (LFID) is visible to the utility, both during the procurement and deployment procedures. Essentially, the manufacturer (and, if different, the escrow service) is the only party which is aware (and has a record) of the connection between a valid HFID/LFID pair. It is important to note here that the escrow is required to comply with a strong data privacy policy. For example, the escrow may not be expected to access, process or store smart metering data – it will only know about the relationship between a valid HFID and LFID.

Fig. 2 shows the distinction between low-frequency metering data being sent directly to the utility, with the high-frequency metering data being sent to the distribution sub-station or other relevant entity. It also shows the internal structure of a smart meter which utilises this method, from a hardware/software ID perspective. The LFID and HFID values that have been mentioned here are hardcoded into the smart meter, with only the manufacturer and/or the 3rd party escrow service being aware of their link within a single smart meter. They are held in the smart meter as part of ID profiles:

- **PISM** or Personally Identifiable SM (Smart Meter) Profile, containing:
  - PISM Certificate (CERT), containing LFID, PISM Public Key and PISM Certifying Authority information
  - PISM Private Key
- **ANSM** or Anonymous SM (Smart Meter) Profile, containing:
  - ANSM Certificate (CERT), containing HFID, ANSM Public Key and ANSM Certifying Authority information
  - ANSM Private Key

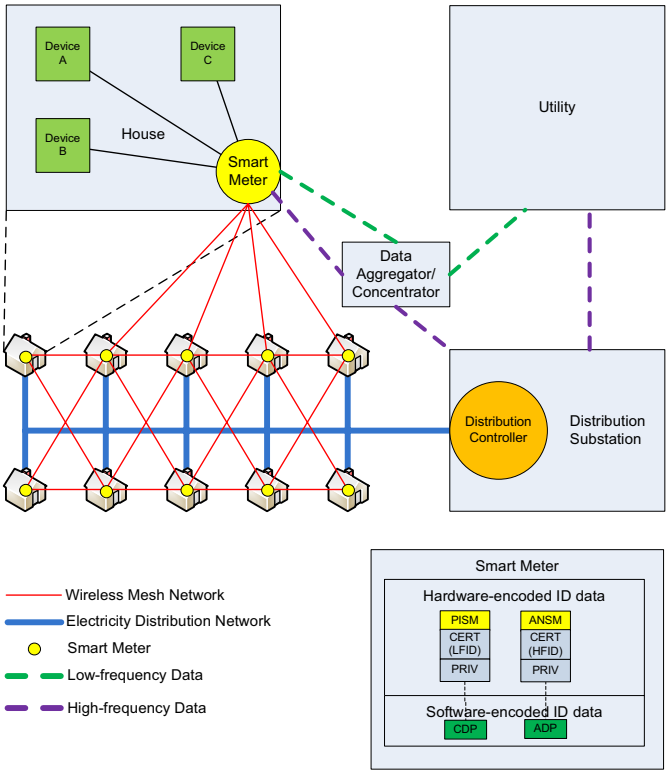

Fig. 2. Distribution network and smart metering data structure

Through a secure protocol setup mechanism described in the next sub-section, the PISM and ANSM are used to create a *Client Data Profile (CDP)* and *Anonymous Data Profile (ADP)* respectively.

It should be noted here that the escrow service provided through the unique HFID is needed for the following reason: suppose that we simply allow a block of smart meters to aggregate their data in a local area. This would apparently hide the parts of the sum attributed to individual smart meters, and the aggregate message can be considered to be anonymized. This, however, would require a separate mechanism for authenticating each smart meter, and this authentication mechanism should not carried out through the utility in order to maintain user privacy. This leads back to an escrow setup as proposed in this paper.

Furthermore, allowing each smart meter to anonymize its data through an aggregator would result in the aggregator receiving a number of anonymous meter readings every few minutes, without any way of correlating these readings with what went on before (in terms of energy usage). This would destroy some of the value of having a unique anonymous ID per smart meter. Similarly, simply aggregating data as such would not allow the utility to apply demand-side or load

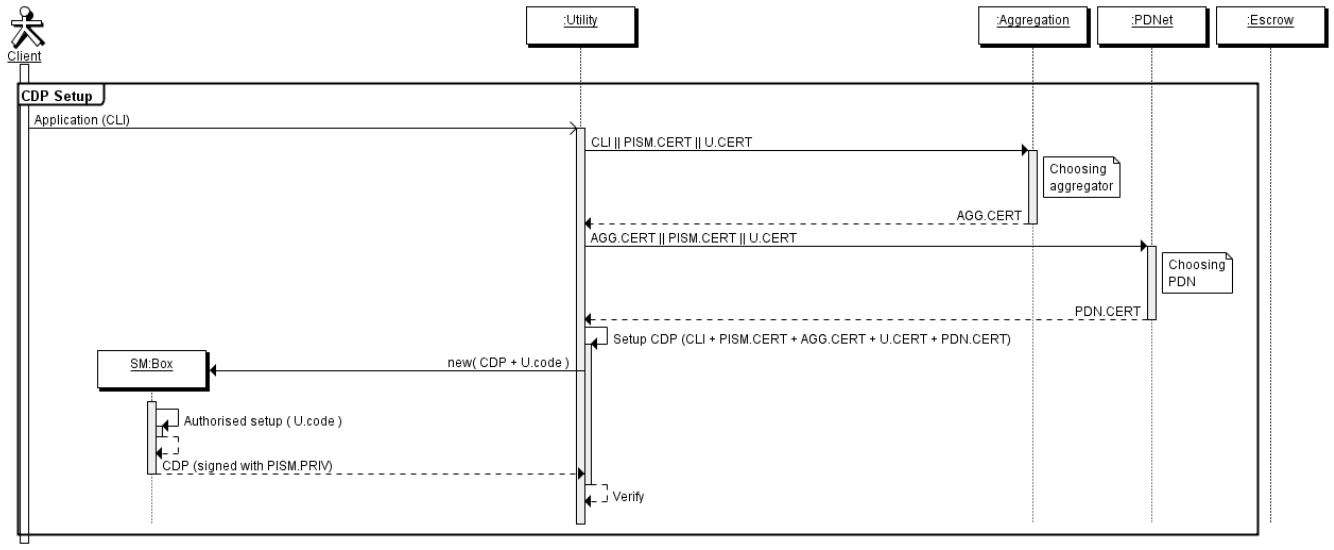

Fig. 3. CDP Setup Process

shedding management at the smart meter level (which, the escrow service will still allow). Unique anonymous IDs should allow the substation or aggregator to detect electricity theft or similar situations where its own meter readings do not match up with the summation of readings from the individual smart meters that are connected to it.

C. CDP and ADP Setup Procedures

Figures 3 and 4 depict examples of setup processes for the Client Data Profile (CDP) and the Anonymous Data Profile (ADP) respectively. The authors recognize that in a real implementation the given protocols should be enhanced to offer security services. For example, the protocol messages below should be digitally signed to reassure integrity. Additionally, they should be encrypted to reassure communications confidentiality. Also, they should include timestamps and/or a random nonce (number used once) message for added integrity and reliability, such as protection against reply attacks. Finally, all actors may be required to share a Certification Authority (CA), which may be used for key management and cross verification purposes. Such detailed protocol exchanges are not described in this paper, for simplicity. Standard security controls and protocols can be found in [8].

1) CDP Setup Process

A client (CL) initially makes an application to a utility for an escrowed smart meter. Alternatively, this process may be initiated by the utility itself in cases where an existing 'dumb' meter is to be upgraded to a smart meter (SM). This application includes the CLI (Client Information), which is all the data necessary to identify a client, such as name, address, etc. In the notation below, S_X denotes a signature operation with signature key X and E_K denotes an encryption operation with encryption key K.

CL → U: CL.CLI

The utility (U) then sends a message to the appropriate metering data aggregator (AGG) the smart meter will be connected to, which includes the CLI and the Personally Identifiable SM Profile Certificate (PISM.CERT). The PISM.CERT contains the LFID as discussed above. The AGG responds with its own certificate (AGG.CERT) as a means of verification that the correct aggregator has been contacted for the SM in question. The aggregator and PISM certificates are sent to the power distribution network (PDNet), which responds with its own certificate for verification.

U → AGG: CL.CLI || PISM.CERT || U.CERT

AGG → U: AGG.CERT

U → PDNet: AGG.CERT || PISM.CERT || U.CERT

PDNet → U: PDN.CERT

Having all the information needed to successfully register the SM in question, the utility now responds to the SM with a Client Data Profile (CDP), consisting of the CLI, the PISM, aggregator, utility and PDNet certificates and U.code. The SM installer (person) needs to input a setup code at this point, which has to match the U.code that the utility has sent to the smart meter, in order to verify genuine installation. As a final step, the smart meter sends back the CDP to the utility, signed with the PISM private key (PISM.PRIV).

CDP = CLI || PISM.CERT || AGG.CERT || U.CERT || PDN.CERT

U → SM: CDP || U.code

SM → U: CDP || $S_{PISM.PRIV}$(CDP)

The completed CDP is what will be attached to each low-frequency metering message from the smart meter to the utility. This CDP setup process is analogous to what would happen with a 'normal' smart meter. Although there could be slight differences between the process we have described and other processes to be used during the installation of a smart meter, the way the CDP setup process has been described here is required for similarity and smooth integration with the ADP setup process described below.

SM begins sending CDP data (infrequently)

SM → U: CDP || Data.LF || $S_{PISM.PRIV}$((CDP || Data.LF))

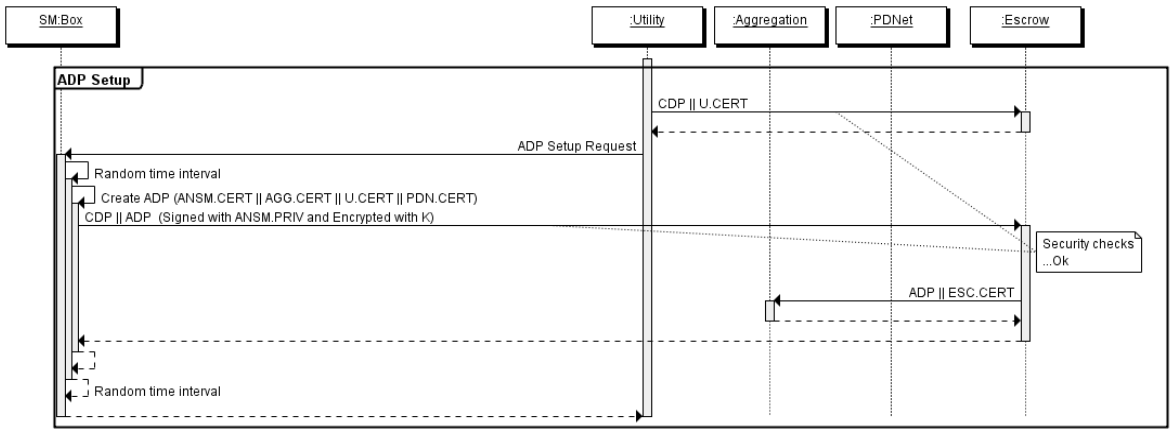

Fig. 4. ADP Setup Process

2) ADP Setup Process

After the CDP setup process has been completed, the CDP is sent to the 3rd party escrow service provider (ESC). The SM is then instructed to setup its ADP.

U → ESC: CDP || U.CERT

ESC → U: OK

U → SM: ADP setup request

A random time interval before the SM responds is introduced here, to allow sufficient time between the setup of the CDP and the ADP so that they cannot be easily correlated by the utility. The target of this approach is to make it difficult to associate a CDP and with a specific ADP. The random process which determines this interval must be set with a mean interval time which is sufficiently large compared to the mean time between successive smart meter setups by the utility, for this interval to be effective. For example, the smart meter could wait for a few days or weeks.

After the random time interval has elapsed, the SM creates its ADP which is comprised of its ANSM certificate, together with the aggregator, utility and PDNet certificates, which were obtained during the CDP setup process. The ADP and CDP are then securely sent to the escrow – it is noted that confidentiality is an important requirement here.

ADP = ANSM.CERT || AGG.CERT || U.CERT || PDN.CERT

SM → ESC: E_K(CDP || ADP) || $S_{ANSM.PRIV}$(E_K(CDP || ADP)),

where K is a secret key. After performing its own security checks, to validate the CDP/ADP combination with its expectation of a LFID/HFID pair, the escrow sends a copy of the ADP together with its certificate to the aggregator. The aggregator now knows that a securely verified and authorised SM will at some point start sending high-frequency metering data to it, which it will then forward on to the utility, distribution network or other authorised entity as appropriate.

ESC → AGG: ADP || ESC.CERT

AGG → ESC: OK

ESC → SM: OK

Finally, another random interval is introduced, analogous to the one used by the SM prior to creating its ADP, so that the aggregator cannot easily correlate the ADP with a specific smart meter. After this second random interval has elapsed, the SM can then commence sending high-frequency metering data. Each of these high-frequency metering messages from the smart meter will include this ADP.

SM begins sending ADP data (frequently)

SM → AGG: ADP || Data.HF || $S_{ANSM.PRIV}$(ADP || Data.HF)

D. Normal Operation

Once the CDP has been set up for the attributable low-frequency metering data, the smart meter responds to LF update requests as needed. Once the ADP has been set up and the appropriate random time intervals have passed, the SM chooses a random number as an initial meter reading and then proceeds to send frequent updates to the data aggregator. This random number is chosen to remove the correlation between LF and HF meter readings that are sent out from the SM.

If any 'micro-management' is necessary, i.e. the utility controlling the smart meter, this can still be achieved through managing what is happening in the area of a single sub-station, as per the assumptions made in this paper. Therefore, the utility could send a control message to the relevant aggregator which could be forwarded on to the anonymous ID.

E. Operation in 'Abnormal' Situations

There may be situations where temporary lifting of the anonymity provided by this solution is required and may be sanctioned. For example, this may be the case when power theft is detected, e.g. when there is a disparity between the reported meter readings in an area and that of the distribution substation serving that area. Anonymity may be reinstated by triggering a 'refresh cycle', effectively forcing each of the smart meters connected to a certain aggregator to re-setup their ADPs, as discussed above. This approach may also be used when a smart meter fails or when new homes are built in an area covered by an existing aggregator.

V. Security Analysis

This section analyzes data communications, CDP setup and ADP setup separately to assess the security properties of our proposed escrow service.

Data communications security controls (i.e. cryptographic functions such as encryption, message authentication codes, and digital signatures) should provide standard security services such as confidentiality, integrity, and accountability of the communicated messages and their origin. For example, in our proposed protocols we assume that all logical or physical communication entities are equipped with digital certificates, issued by some trusted party or CA (e.g. as in Public Key Infrastructure – PKI – systems).

The security of CDP setup needs to be discussed further, as the authenticity of both the client and the SM cannot be guaranteed when the CDP process begins. The client may be verified by the utility engineer during installation. The SM authenticity can be verified after administering the secure code U.code. If U.code remains unique (for a required amount of time) and it is not jointly misused (by the engineer and the client), the utility can safely verify that a genuine SM was installed at a genuine location.

The security of ADP setup depends primarily on the security of the CDP process as well as the trustworthiness of the escrow entity. Secondarily, ADP security depends on the confidentiality of the association between ADP and CDP, or else the level of anonymity achieved through the setup and use of ADP. Assuming that this secret will be kept confidential by the escrow service and the smart meter, we need to quantify how hard it is for the utility to deduce this secret.

The anonymity of ADP depends on the size of the 'anonymity set' (see related definitions in [9]). Suppose that the utility mounts passive attacks based on non-intrusive analysis of received messages: the harder it is to link ADP with CDP, the better the anonymity achieved. The anonymity set comprises all the ADP finalization responses (sent from different smart meters) the utility receives during the period between one smart meter sending a CDP finalization response and sending the associated ADP finalization response. For example, if the utility never receives any other ADP finalization responses between any CDP finalization response and any ADP finalization response (i.e. the anonymity set comprises only one ADP finalization response) then our ADP protocol offers no anonymity at all: the utility can trivially associate the first received ADP finalization response following a CDP finalization response. In our proposed protocol, the degree of anonymity (or conversely the weakness of our escrow protocol) depends on the random time interval: its average value needs to be large enough to allow a large enough anonymity set to be created (on average).

We can illustrate the importance of the random time interval by assuming that there is (on average) a rate of x CDP installations per unit of time for a certain aggregator (and a certain PDNet). Suppose that we require an anonymity set of size y; clearly, the time interval should be on average larger than $T=y/x$. Hence, our anonymizing time should be randomly chosen within $[0, 2T]$. On a practical note, we should mention that this procedure may not be applicable when the rate of CDP is low for a certain region characterised by a {AGG,PDNet} pair (consider, for example, the extreme case where y=50 and x=1 CDP/year). Also, the protocol assumes that the rate of new installations is somewhat constant: if, for example, there is a burst of CLI applications followed by a long period of no further applications, the utility could deliberately spread delay in some CDP setup procedures in order to reduce the anonymity set. This problem could, however, be reduced by allowing the escrow service to control the random time interval (and the anonymity set) as the escrow service will be in a position to know the total number of ADP setup requests (assuming that all ADP setup requests are sent to a single escrow service, or, equivalently, different escrow services collaborate for that purpose).

VI. Conclusion

Privacy concerns are very important to the future deployment of smart metering and smart grid networks, as the amount of data collected from future smart meters will be orders of magnitude more than data collected from current meters. This data can easily be mined, as demonstrated by NALM and other techniques. In this paper we have attempted to address the smart metering privacy issue by anonymizing the identity of high-frequency metering data through an escrow service. Our security analysis shows that the key to the security offered is the trust level of such an escrow service, together with the random time intervals between the setup of attributable and anonymous data profiles at the smart meter.

We note that our proposed method may not offer sufficient smart metering privacy protection. However it contributes an additional layer of security towards that direction.

Future work will involve defining how this anonymization process can be extended to address a number of practical scenarios, such as when anonymity needs to be lifted (e.g. for forensic reasons), when a faulty smart meter needs to be replaced or when the anonymity set determined by new smart meter installations is small, such as, for example, the case where only one new smart meter needs to be (re)-installed.

References

[1] European Commission, "European SmartGrids Technology Platform – Vision and Strategy for Europe's Electricity Networks of the Future", Directorate-General for Research – Sustainable Energy Systems, 2006, available from http://www.smartgrids.eu

[2] http://www.nist.gov/smartgrid, accessed 30 July 2010

[3] A. H. Rosenfeld, D. A. Bulleit, R. A. Peddie, "Smart Meters and Spot Pricing: Experiments and Potential," IEEE Technology and Society Magazine, vol.5, no.1, pp.23-28, March 1986

[4] E. L. Quinn, "Privacy and the New Energy Infrastructure", Social Science Research Network (SSRN), February 2009

[5] R. Stallman, "Is digital inclusion a good thing? How can we make sure it is?," IEEE Communications Magazine, vol. 48, pp. 112-118, February 2010

[6] G. W. Hart, "Nonintrusive appliance load monitoring", Proceedings of the IEEE, vol.80, no.12, pp.1870-1891, December 1992

[7] H. Y. Lam, G. S. K. Fung, W. K. Lee, "A Novel Method to Construct Taxonomy of Electrical Appliances Based on Load Signatures", IEEE Transactions on Consumer Electronics, vol.53, no.2, pp.653-660, May 2007

[8] A. W. Dent and C. J. Mitchell, "User's guide to cryptography and standards", Artech House, 2005

[9] A. Pfitzmann and M. Hansen, "Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management – A consolidated proposal for terminology", http://dud.inf.tu-dresden.de/Anon_Terminology.shtml, 2008, version 0.31e