An Anonymized Authentication Framework for Smart Metering Data Privacy

Sabrina Afrin
B. Thomas Golisano College of
Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY 14623
Email: sa3213@rit.edu

Sumita Mishra, PhD
Senior member, IEEE
B. Thomas Golisano College of
Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY 14623
Email: sumita.mishra@rit.edu

***Abstract*—Emergence of the smart grid necessitates data collection from various entities connected to the power grid. By collecting fine-grained smart metering data from consumer household, smart grid is envisioned to offer innovative features, such as load-demand balance, dynamic pricing and demand-response. However, the advent of sophisticated data analysis tools makes it possible to extract detailed electricity usage patterns, which can reveal personal information of consumer household. The privacy concern is one of the reasons jeopardizing general deployment of smart meters, despite their effectiveness in cost savings for both electricity suppliers and consumers. Anonymous transmission of smart metering data is a potential solution but it requires mechanisms to authenticate anonymous data. Moreover, enabling the service provider's access to consumer information for grid functionality while preserving consumer privacy, is a challenging task. The proposed anonymized authentication framework consists of an authentication scheme to protect data from unauthorized entities and an anonymization scheme to achieve privacy against authorized entities. The framework is designed to prevent service providers from correlating different types of data from a smart meter and avoid single point of failure.**

I. Introduction

The 20^{th} century's radial, one-way, generation-centric traditional electric grid is gradually being replaced by the distributed, two-way, information-centric grid known as smart grid. Smart Meter (SM) with its bidirectional communication capabilities, works as the interface between the consumer domain and the rest of the grid. As electric power grid interconnects an entire country, it is considered as one of the most vulnerable infrastructures to attackers, as a compromised grid can cripple the entire nation. However, protecting the grid from external attacks is not the only concern with the smart grid. Authors in [1], [2], have shown that, by analyzing electricity consumption pattern of any household, it is possible to identify information, such as the absence or presence of a person inside the house, type of appliance being used, or even the movies or channels being shown on TV. National Institute of Standards and Technology (NIST) has identified consumer privacy concern as one of the seven major challenges to be addressed before the full roll-out of smart grid [3].

Anonymization of smart metering data is one of the recommended methods by NIST [3]. In smart grid communication, smart meters need to interact with multiple entities which may or may not be part of the same organization. When multiple service providers share information, detailed information of a consumer may be revealed. The main goal of the proposed method is to enable SMs to communicate anonymously with different service providers in smart grid by using different pseudonyms. However, as service providers are interested in authenticated smart metering communication, this framework offers an authentication mechanism integrated with the anonymization method. Simultaneously, it aims to prevent collusion of multiple service providers and avoid chances of single point of failure. In current literature, escrow-based anonymization method [4] has been proposed for smart metering privacy. However, it is susceptible to vulnerabilities such as single point of failure, unguaranteed anonymity in some cases etc. In this paper, an adaptation of anonymized authentication framework with escrow-based method is analyzed to show how these vulnerabilities can be improved.

The rest of the paper is organized as follows. Section II discusses different proposals in current literature for smart metering data privacy followed by problem statement and description of escrow-based method in section III. The proposed anonymized authentication framework is described in section IV. Section V discusses the combined protocol of anonymized authentication framework and escrow-based method. The analysis of the combined protocol is given in section VI, followed by conclusion in section VII.

II. Related Work

The increasing concern of smart metering data privacy has led to numerous attempts by researchers to address the issue. The method proposed in [4] introduces anonymous transmission of fine-grained metering data but is susceptible to the compromise of a trusted party. The distributed anonymization method in [5] requires the absence of bidirectional metering communication and computationally expensive public key operation for each round of data transmission. To address privacy of fine-grained metering data, Kalogridis et al. have proposed a method to draw energy within a home from a rechargeable battery, instead of drawing directly from the main electricity supply [6]. This approach prevents the detection of load

978-1-5090-5167-0/16/$31.00 ©2016 IEEE

signatures of appliances but requires large battery power for complete privacy. In [7], authors proposed the use of a differential private algorithm for hiding distinguishing consumption patterns by adding noise. In-network data aggregation methods using homomorphic encryption with collaboration of smart meters have been proposed in [8], [9]. However, the use of homomorphic encryption induces computational overhead on the resource constrained metering hardware.

In the proposed framework, the segmentation of protocol execution over multiple service providers prevents any single entity from having full control over the anonymization method. This mechanism reduces chances of single point of failure. The authentication mechanism is designed not only to thwart unauthorized entities but also preventing authorized entity's from deducing consumer information. As a result, this framework is expected to be resilient to collusion of service providers having access to different pieces of consumer information.

III. Background

A. Problem Statement

The operation of smart grid relies on data collected from entities connected to the grid and managed by different Service Providers (SP). For example, Electricity Supplier (ES) (also referred to as Utility) provides electricity to the consumers and requires aggregated electricity usage data of the households in a certain area for dynamic load-demand management. A Grid Operator (GO) may require real-time data from grid monitoring sensors. Data Collector (DC) or Data Aggregator collects data generated by different devices connected to the grid, processes (aggregates) the data and forwards to other entities (i.e., ES or GO). However, collection of smart metering data poses privacy threats to the consumers. Anonymous data transmission enables a user to transmit data using pseudonyms without revealing its original identity. Nevertheless, when multiple SPs from the same or different organizations share information, there is a possibility that the combined data may reveal additional information about a user, thus reducing privacy. As a result, anonymous metering data transmission requires an effective method to make the correlation of consumer information difficult for multiple SPs working together. On the other hand, to prevent malicious external entities from injecting false data by taking advantage of the anonymous method, an authentication process is required. Hence, a method enabling the SPs to check the authenticity of anonymous data and at the same time preventing the SPs from extracting consumer information, is desired. An entity named ANonymizer (AN) is introduced in this paper, which participates in the proposed anonymization method to prevent service providers (single or multiple) from correlating pseudonyms of a meter. The interactions between different entities of the grid are explained further in section IV.

B. Escrow-based Method

Efthymiou and Kalogridis are the first to propose escrow-based anonymization method for SMs by differentiating two types of metering data [4]. In this method, aggregated monthly (or, weekly) electricity consumption data sent to the electricity supplier/utilities for billing purposes is referred to as low frequency data. On the other hand, fine-grained electricity consumption data generated every few minutes and required for accurate load-demand prediction, is referred to as high frequency data. Every SM is hardcoded with two pseudonyms- Low Frequency ID (LFID) and High Frequency ID (HFID) along with other credentials (e.g., certificates, and public and private key associated with these two IDs). By following similar setup procedures and using the credentials, LFID and HFID are used to create two profiles which are employed in transmission of two different types of metering data. For billing management, the LFID associated with a particular house is known to ES/utility. On the other hand, HFID is related to high frequency data and should not be related to a consumer for privacy issues. To obscure the relationship between any LFID-HFID pair, a random time interval is introduced between the two profile setup procedures. However, for authentication purposes, this method requires the presence of a Trusted Third Party (TTP) which has the knowledge about the association of any LFID-HFID pair.

The anonymity achieved by an SM in escrow-based method depends on number of meters setting up profiles during the random time interval of a meter. If no other meter sets up profile during this time interval, this procedure does not provide anonymity [4]. Since the setup procedures require interaction between the ES and the TTP, ES has the capability to reduce the anonymity set and may be able to find the owner of an LFID. Moreover, the security of the whole system depends on the trustworthiness of the TTP. As a result, compromise of the TTP can jeopardize the scheme causing single point of failure. Additionally, hardcoding the HFID in the meter weakens the anoymization method if it is accidentally leaked. In section V, we have analyzed how these challenges can be overcome with anonymized authentication framework.

IV. Anonymized Authentication Framework

Suppose, an SM is authenticated with Service Provider B (SP_B) with consumer information of its owner and needs to communicate with Service Provider C (SP_C). SP_B and SP_C are entities having access to consumer information. The goals of the mechanism are as follows,

- The SM and any SP should be able to authenticate each other as valid entities.
- SPs should not be able to correlate multiple pseudonyms used by an SM with the consumer's identity (except ES in a particular case as explained in next section).

To achieve these goals, the authentication mechanism is carried out by each SM whenever it communicates with an SP, and the anonymization process is carried out by the AN between an SM's communication with two other SPs. The authentication process is derived from Kerberos authentication protocol [10]. The entities participating in the proposed protocol use public key cryptography only for the initial communication before authentication process. After two entities check each other's

978-1-5090-5167-0/16/$31.00 ©2016 IEEE

TABLE I: Notations used in this work.

Notation	Meaning
$E_K(m)$, $D_K(m)$	Encryption and decryption of text m with key K
$K_{R,pb}$ and $K_{R,pr}$	Public and private key of entity R
K_{PQ}	Symmetric key between two entities with P and Q as the first letters of their names
N_{PQ}	Nonce transmitted between two entities with P and Q as the first letters of their names
$EncMsg_{PQ}$	Message encrypted with the symmetric key K_{PQ}
$x\|\|y$	Concatenation of texts x and y

authenticity, transmitted messages are encrypted using symmetric keys, thus avoids computational complexity of public key cryptography. It is assumed that, every pair of SPs have pre-deployed pairwise mutual keys. Additionally, the mutual key between two entities is not known to any third entity (Perfect Secrecy Assumption (PSA)). The notations used in this paper are summarized in Table I.

Authentication: The authentication mechanism includes collecting required credentials from one SP and using them for authentication with next SP. SP_B generates a pseudonym $Psnym_S$ for the SM, and a mutual key K_{SA} for the SM and the AN. Next, it encrypts them using its mutual key with the AN to create an encrypted message ($EncMsg_{BA}$) and transmits it with other credentials required for next stage, to the SM.

$$EncMsg_{BA} = E_{K_{BA}}(Psnym_S||K_{SA})$$
$$SP_B \rightarrow SM : E_{K_{SB}}(Psnym_S||AN||EncMsg_{BA}||K_{SA})$$

where, K_{SB} is the mutual key between SP_B and the SM. From this message, the SM knows it has to communicate with AN next. SM generates a nonce N_{SA} and encrypts it with K_{SA}. Next, it sends the following message,

$$SM \rightarrow AN :$$
$$E_{K_{AN,pb}}(Psnym_S||SP_B||EncMsg_{BA}||E_{K_{SA}}(N_{SA}))$$

The AN checks the authenticity of the SM from this message as shown in Fig. 1 and the steps are described below.

Step 1: The received message indicates that it is an authentication request from an SM with identity $Psnym_S$ and this request is forwarded from SP_B. AN decrypts $EncMsg_{BA}$ using its mutual key with the SP_B (K_{BA}).

Step 2: The decrypted $EncMsg_{BA}$ should yield the pseudonym and key provided by SP_B.

Step 3: The AN checks if this pseudonym is same as the requesting pseudonym.

Step 4: If both the pseudonyms are same, AN considers the request as valid since no one else should be able to forge $EncMsg_{BA}$ (PSA). Next, AN decrypts the nonce N_{SA} using the K_{SA} found in $EncMsg_{BA}$.

The AN next forwards the nonce to the SM after encrypting with the SM's public key, so that the ES cannot tamper this message.

$$AN \rightarrow SM : E_{K_{SM,pb}}(N_{SA})$$

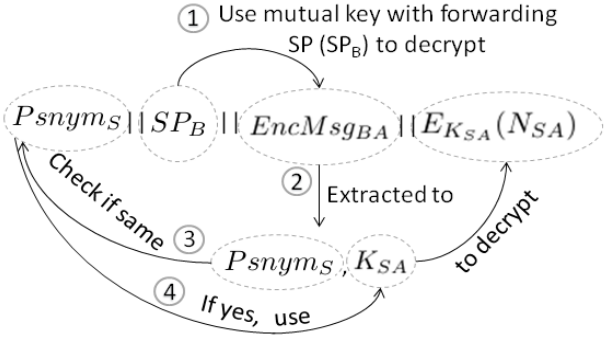

Fig. 1: SM's authentication request to the AN

If the SM receives the correct nonce, then it is communicating with the right entity. The SM and AN updates their mutual key using any reliable key exchange protocol (e.g., Diffie-Hellman Key Exchange (DHKE)). As a result, the ES cannot eavesdrop on their future communication.

Anonymization: The SMs in an area participate in the authentication with the AN before communicating with another SP. The AN keeps the pseudonyms of these SMs (e.g., $Psnym_S$ for this particular SM) in an anonymization table. Every pseudonym is assigned a maximum waiting time (T_{Max}) to ensure that no pseudonym has been waiting for an indefinitely long time. At a predefined time interval (selection round of time T_{Select}), the AN randomly selects one (or more) pseudonym from the anonymization table by following the conditions given below,

1) *Selection of pseudonym:* AN randomly selects $N_{selected}$ pseudonyms from $N_{waiting}$ number of waiting pseudonyms, with equal probability (i.e., $\frac{N_{selected}}{N_{waiting}}$). If there is any pseudonym waiting longer than T_{Max}, it should be selected with probability 1.
2) *Selection of T_{Max}:* T_{Max} can be same for all the pseudonyms, or variable for different pseudonyms. In both cases, the following conditions should be met,
   a) If one pseudonym has been waiting longer than T_{Max} when $N_{waiting} = 0$, this pseudonym has anonymity set of size zero. In this case, the AN waits for multiple selection rounds until more pseudonyms arrive and then selection is done according to step 1. If more than one pseudonym is waiting longer than T_{Max}, pseudonyms can be selected according to step 1 without waiting for multiple rounds.
   b) The anonymity achieved by a pseudonym depends on the number of pseudonyms waiting in the table. Maximum anonymity is achieved when number of pseudonyms in an anonymity set is equal to the number of SMs in an area. To ensure standard minimum anonymity, a minimum threshold ($N_{threshold}$) value for $N_{waiting}$ can be set. In this case, another selection condition is,
   $$N_{selected} \geq max(N_{threshold}, N_{MaxWait})$$
   where, $N_{MaxWait}$ is the number of pseudonyms wait-

978-1-5090-5167-0/16/$31.00 ©2016 IEEE

ing for more than T_{Max}.

After a pseudonym is selected, AN generates a new pseudonym and key, and transacts the credentials for the SM's authentication with next SP (SP_C in this case). The authentication process between the SM and SP_C is same as described before.

V. Adaptation of Anonymized Authentication Framework with Escrow-based Method

In this section, the proposed anonymized authentication framework is applied to the escrow-based method to improve the vulnerabilities associated with it. This combined approach is referred to as hybrid protocol. In the hybrid protocol, it is assumed that, only the LFID is hardcoded to every SM and ES has the knowledge of which LFID is associated with which household (i.e., consumer). Moreover, the authentication method does not require the presence of TTP. For simplicity, the profiles associated with low and high frequency metering data are referred to as $Profile_{LF}$ (Low Frequency Profile) and $Profile_{HF}$ (High Frequency Profile), respectively. In this hybrid protocol, profiles are generated using escrow-based method and SMs participate in anonymized authentication procedure for anonymity and authentication. The combined procedure is split into three parts- 1) Low Frequency Profile Setup with ES, 2) High Frequency Profile Setup and Anonymization with AN, and 3) High Frequency (HF) Data Transmission with DC.

A. Low Frequency Profile Setup with ES

The authentication between the SM and the ES is slightly different than the process described in section IV. This is because, the ES is the first SP an SM communicates in the hybrid protocol. As there is no mutual key between the SM and the ES at this stage, the SM generates the initial key K_{SE} by calculating a hashed value of the LFID with any collision resistant, one-way hash function (e.g., SHA-2). Next, it generates a nonce N_{SE} and creates $EncMsg_{SE}$. A profile setup request consisting of Consumer Information $ConInfo$ (e.g., name, address etc.) and $EncMsg_{SE}$ is sent to the ES.

$$K_{SE} = hash(LFID)$$
$$EncMsg_{SE} = E_{K_{SE}}(LFID||N_{SE})$$
$$SM \rightarrow ES : E_{K_{ES,pb}}(ConInfo||EncMsg_{SE})$$

After receiving the request, the ES finds the LFID associated with the $ConInfo$ and calculates a hashed value of the LFID to decrypt $EncMsg_{SE}$. If the decrypted $EncMsg_{SE}$ yields the correct LFID, that means K_{SE} and the calculated hashed value are same and this request is valid. The ES sends back N_{SE} after encrypting with the SM's public key. If the SM receives the correct nonce, both entities update their mutual key. By following the process in [4], the ES sets up the $Profile_{LF}$. After this part, the SM's authentication process with the AN is same as described in section IV. Suppose, the ES generates pseudonym $tempID$ and key K_{SA} as the credentials, creates $EncMsg_{EA}$, and sends them to the SM.

$$EncMsg_{EA} = E_{K_{EA}}(tempID||K_{SA})$$
$$ES \rightarrow SM : E_{K_{SE}}(tempID||AN||EncMsg_{EA}||K_{SA})$$

B. High Frequency Profile Setup and Anonymization with AN

After receiving the required credentials, the SM generates a nonce N_{SA} and transmits the following message to the AN.

$$SM \rightarrow AN :$$
$$E_{K_{AN,pb}}(tempID||ES||EncMsg_{EA}||E_{K_{SA}}(N_{SA}))$$

From the received message, the AN knows an SM with $tempID$ is requesting for a profile setup and this request is forwarded from the ES. Next, following the authentication process, the AN checks the $tempID$ and the SM checks the correctness of nonce forwarded by AN and both update their mutual key. If the temporary IDs don't match the request is discarded. After the authentication, the AN generates an HFID for the SM and creates $Profile_{HF}$. Next, by performing the proposed anonymization process, the AN selects a $Profile_{HF}$ for transaction of credentials for communication with the DC.

The anonymization process makes the order of $Profile_{HF}$ setup and HF data transmission different. Suppose, an SM starts transmitting high frequency data right after it has set up its $Profile_{LF}$. Meanwhile, there is no other SM who has set up its profile. If ES and DC share information, the ES can trivially find out that the most recent data is transmitted by the SM who just set up $Profile_{HF}$. In [4], a random time interval is used for this purpose which does not guarantee anonymity even for indefinitely long random time. However, AN works as the entity ensuring a minimum anonymity set. If the ES and the DC are parts of different organizations and the DC provides metering data to the ES without revealing information of $Profile_{HF}$, no anonymization method is required. However, considering that the ES and the DC share information makes the adversarial model stronger.

C. High Frequency (HF) Data Transmission with DC

Before a meter starts transmitting high frequency data, it participates in the authentication process with the DC. After both the SM and the DC ensure each other's authenticity and set up a mutual key, the SM starts transmitting high frequency data with the $Profile_{HF}$.

VI. Comparative Analysis of Hybrid Approach and Escrow-based method

In this section, the anonymity achieved by SMs is analyzed with respect to SPs ability of correlating a $Profile_{LF}$ to a $Profile_{HF}$. The main purpose of the hybrid and escrow-based method is to prevent ES, or DC, or both from identifying High Frequency (HF) data transmitted by a particular SM (consumer) at a particular time. Since $Profile_{LF}$ is associated with consumer information and $Profile_{HF}$ to HF data, an adversary should not be able to relate a particular $Profile_{HF}$ to a $Profile_{LF}$.

In the $Profile_{LF}$ setup, $ConInfo$ is used as pseudonym instead of LFID for authentication, to prevent accidental

978-1-5090-5167-0/16/$31.00 ©2016 IEEE

compromise of LFID. This is required as LFID is hardcoded in each meter and compromise of LFID may have long-term effect on privacy. For the same reason, HFID is not hardcoded in meters as $Profile_{HF}$ is more frequently transmitted and more exposed to external attacks. However, both $Profile_{LF}$ and $Profile_{HF}$ should be reset periodically by adding nonce or other methods to achieve perfect forward secrecy.

The proposed authentication mechanism removes the need of presence of TTP. As a result, there is no single entity in the protocol which has the knowledge of the association of $Profile_{LF}$ (or consumer) and $Profile_{HF}$ (or HF data). This reduces the chances of single point of failure. In the escrow-based method, ES (or utility) has the ability to let one SM setting up its profile while preventing others. This leads to an anonymity set of 1 meter and reveals the HF data sent by this SM. Even if the ES does not deliberately refrain other meters from participating, this situation may occur if the time interval between two successive profile's setup is longer than the random time. However, the conditions of multiple selection rounds and minimum threshold number of selection in the proposed anonymization method ensures a standard minimum anonymity set for all the meters. Suppose, in the hybrid approach, the ES refrains from providing $tempID$ to other meters after forwarding an SM to the AN. After T_{Max} time interval, the AN will see that, $N_{waiting} = 0$ and so it will wait for multiple selection rounds until $N_{threshold}$ number of pseudonyms arrive. If $Profile_{HF}$ of every SM is reset periodically, the chance of anonymity set being decreased to 1 is also reduced.

In the hybrid approach, the ES knows the $Profile_{LF}$ and $tempID$ associated with every consumer and the aggregated consumption data of a group of meters. The ES also knows which consumers have collected their $tempID$, but not which consumer's profile is waiting in the anonymization table or which consumer is transmitting a particular HF data. The DC on the other hand, has access to the HF data transmitted by every $Profile_{HF}$ though not the consumer information associated with these profiles. So, the ES, or the DC alone cannot correlate a $Profile_{HF}$ to a $Profile_{LF}$. For the DC, the number of possible senders (anonymity set) is $S = N$, where N is the total number of meters in an area.

If ES and DC share information, they know which consumers have collected $tempID$ and how many consumers are transmitting HF data (N_{HF}) at a particular time. Since, total number of meters in an area (N) should be known to all the SPs, they can find number of $Profile_{HF}$ waiting in the anonymization table as,

$$N = N_{HF} + N_{temp} + N_{waiting}$$

where, N_{temp} is the number of SMs which collected $tempID$. The combined adversary of ES and DC has access to fine-grained data transmitted by any $Profile_{HF}$, but a $Profile_{HF}$ can be associated to any of the $(N_{temp}+N_{waiting})$ consumers and the identity of the $N_{waiting}$ meters is hidden by the anonymization process. As a result, the anonymity set consists of $S = N_{temp}+N_{waiting}$ meters in this case. This is the worst case scenario, when both the ES and DC are colluding. In the escrow-based method, the anonymity set can be reduced to 1 in the worst case scenario.

It should be noted that the proposed protocol does not consider privacy attacks using external information. For example, if an SP knows a particular appliance being used by a certain consumer, it can try to find the HF data sent by this consumer by detecting load signature of this appliance. Moreover, the proposed scheme does not achieve anonymity when the ES, the DC and the AN collude. Nevertheless, since AN does not have access to consumer data, it is considered as neutral entity.

VII. Conclusion

Addressing smart metering data privacy requires urgent attention due to the rapid development of the smart grid. Efficient grid operation is desirable not only to the electricity suppliers, but also to the consumers. However, the benefits of the smart grid should not be gained at the cost of consumer privacy. The information-centric operation of smart grid creates a dilemma by introducing a trade-off between consumer privacy and grid performance. In this paper, an authentication and anonymization framework is proposed to address smart metering data privacy. The effectiveness of this framework is shown by applying it to improve vulnerabilities of the escrow-based method. In comparison with escrow-based method, the proposed scheme is more resilient against privacy threats from service providers. Future work will focus on designing a more robust anonymization scheme addressing collusion of all service providers as well as different external attack models.

References

[1] HY Lam, GSK Fung, and WK Lee. A novel method to construct taxonomy electrical appliances based on load signatures. *Consumer Electronics, IEEE Transactions on*, 2007.

[2] Ulrich Greveler, Benjamin Justus, and Dennis Loehr. Multimedia content identification through smart meter power usage profiles. *Computers, Privacy and Data Protection*, 2012.

[3] NIST. Guidelines for smart grid cyber security. *Draft NISTIR*, 7628, 2010.

[4] Costas Efthymiou and Georgios Kalogridis. Smart grid privacy via anonymization of smart metering data. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010.

[5] Sören Finster and Ingmar Baumgart. Pseudonymous smart metering without a trusted third party. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on*. IEEE, 2013.

[6] Georgios Kalogridis, Costas Efthymiou, Stojan Z Denic, Tim A Lewis, and Rafael Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010.

[7] Klaus Kursawe, George Danezis, and Markulf Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *Privacy Enhancing Technologies*. Springer, 2011.

[8] Fenjun Li, Bo Luo, and Peng Liu. Secure information aggregation for smart grids using homomorphic encryption. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010.

[9] Flavio D Garcia and Bart Jacobs. Privacy-friendly energy-metering via homomorphic encryption. In *Security and Trust Management*. Springer, 2011.

[10] B Clifford Neuman and Theodore Ts'o. Kerberos: An authentication service for computer networks. *Communications Magazine, IEEE*, 1994.

978-1-5090-5167-0/16/$31.00 ©2016 IEEE