A Lightweight Lattice-Based Homomorphic Privacy-Preserving Data Aggregation Scheme for Smart Grid

Asmaa Abdallah and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—Consumer privacy and consumption confidentiality and integrity are the main security concerns for smart grid connection with the residential electricity consumers. This paper proposes a lightweight privacy-preserving electricity consumption aggregation scheme that exploits lightweight lattice-based homomorphic cryptosystem. In the proposed scheme, smart household appliances aggregate their readings without involving the smart meter. Although smart meters or the intermediate base station cannot decrypt this aggregated consumption, they can validate the message's authenticity. The proposed scheme also investigates the impact of different types of smart appliances on the home area network's overhead. The total communication and computation load for the proposed scheme is trivial and tolerable by different parties in the connection, i.e., smart appliances, smart meters, and the base station. In addition, the deployed cryptosystem, which depends on simple arithmetic operations, can further reduce the computation duty for smart appliances. Simulation results and security analysis show that our proposed scheme guarantees consumers privacy, and messages authenticity and integrity, with lightweight communication and computation complexity.**

***Index Terms*—Lattice-based homomorphic cryptosystem, smart appliances, smart meter.**

I. Introduction

SMART GRID utilizes communication technologies to improve reliability and efficiency of the power grid. The core of the smart grid's communication with consumers is home area networks (HANs). Each HAN consists of a smart meter (SM) that connects to certain smart household appliances (APs), i.e., consumer's electric vehicle (EV) is considered a smart appliance too. SM is primarily responsible for aggregating APs' readings and forwarding the aggregated value to the utility periodically. HANs send these periodic reports about houses' electricity consumption so that the utility can accurately compute customers' electricity bills and utilize this information to forecast future power demand and prices for the region [2], [3].

Manuscript received October 7, 2015; revised November 24, 2015 and February 1, 2016; accepted April 10, 2016. Date of publication April 13, 2016; date of current version December 21, 2017. A preliminary version of this paper was presented at the 2014 Wireless Communications and Signal Processing Conference [1]. Paper no. TSG-01287-2015.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: a3abdall@bbcr.uwaterloo.ca; xsheng@bbcr.uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2016.2553647

In addition, different types of APs may have impact on HAN's performance and security concerns, as APs vary in their electricity needs and consequently in their communication requirements. For instance, light bulb's information definitely is much less than air conditioner's (AC)'s data so that ACs require to exchange more messages with control center (CC) than bulbs. According to [3], APs are divided into four groups according to their communication needs:

Group 1 consists of small loads, such as light bulbs, phone chargers, and laptops. These loads do not have a significant impact on the total load and only need to inform CC when they are (dis)connected. *Group 2* consists of large uncontrollable loads, e.g., stoves, which are working according to consumer's needs and cannot be delayed to a later time. This type of APs needs minimal communication infrastructure to only send its power and expected duration of usage. *Group 3* consists of controllable large loads, such as ACs and clothes washers. These APs should send requests via SM to CC that include information, such as expected load, expected duration of usage, and duration of availability. However, they cannot operate until they receive acceptance acknowledge. CC can accept/reject the request; its decision depends on dynamic pricing, duration of availability, and customer's agreement with the utility. *Group 4* are EVs that communicate to schedule their (dis)charging processes. Due to their extensive communication needs, they are categorized as separate loads. Accordingly, different groups of APs have different communication needs and consequently different impact on the HAN's security. Groups 3 and 4 increase the security threats on HAN, as they require to exchange more messages with SM than groups 1 and 2.

The major security concerns of HANs are consumers' privacy and consumption readings' confidentiality and integrity. During APs' readings aggregating and sending to the utility, the householders' personal information and daily habits, e.g., the current used APs, when owners are in/out the house, can be revealed to any party, i.e., CC or eavesdroppers, and consequently threaten the customer's privacy. In addition, the confidentiality and integrity of electricity readings' information are significant concerns, because this information could be misused by outsiders to gain benefits or harm consumers. For instance, if the customer is a small business company or industrial institution, the adversary can extract information about the types and quantity of the company's products. Then, he/she can exploit this information to get financial benefits,

i.e., by blackmail or sell the information to competitors. Moreover, SMs and household APs are restricted computation-capability devices; the communication and computation burden on them is also a major concern. Consequently, customers' privacy and information security should be guaranteed with lightweight cryptosystems [2]–[5].

In this paper, we propose a lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for the residential electricity consumers in smart grid. The proposed scheme (unlike the related works) allows the household APs to aggregate their consumption among themselves without involving SM; utilizing a lightweight lattice-based homomorphic cryptosystem. SM does not know the reading for each individual AP; it receives the total encrypted aggregated consumption for all APs in the HAN. So, SMs, also the related base station (BS), work as relay nodes and just forward HAN's total consumption to CC. However, SM, i.e., BS as well, has the ability to check the authenticity of messages' senders without revealing their contents. Consequently, the proposed scheme guarantees the security and privacy demands, i.e., customers' privacy, data confidentiality and integrity, for the connection. It is also a lightweight and efficient in terms of communication and computation complexities so that it is suitable for limited-capability devices, i.e., SMs and APs.

The remaining of the paper is organized as follows: Section II introduces the related works in literature. Section III describes the system model and security requirements. Section IV reviews the lattice-based homomorphic encryption scheme. Section V presents our proposed scheme. The security analysis and performance evaluation are illustrated in Sections VI and VII respectively. Finally, Section VIII concludes the paper.

II. Related Work

Many proposed solutions in the literature attempt to satisfy security requirements for smart grid's connection with residential customers. Current research studies can be divided into three categories:

The first one is connecting SMs to hardware devices, e.g., temper-resistance devices or electrical batteries, to conceal the real electricity consumption of the place [6]–[8]. These procedures alleviate computation and communication burden, but it is expensive to connect a temper-resistance or battery to each SM in the grid besides the necessary maintenance operations.

The second category is distorting the consumption readings by adding noise to them at HAN side and removing it at CC [9]–[11]. These methods conserve the computational abilities for the HAN's smart devices but cannot accurately reconstruct the original message from the received one.

Third category employs cryptographic schemes, such as public key infrastructure [12] or key-policy attribute-based encryption (ABE) scheme [13]–[15], to guarantee the security and privacy requirements for the connection. This category can be further subdivided into:

Studies that use authentication schemes [16]–[19] to guarantee information integrity and confidentiality, but authentication operation consumes computation and communication abilities. Lightweight Diffe-Hellman authentication scheme [16], for example, causes an average delay varied from 1 to 10 s, as SMs' number increases. Also, each meter should previously have a secret value to create its authentication key.

Other researches employ anonymization techniques to conceal the relation between meter's real identity and its consumption. These techniques are mainly based on issuing two identities (real and pseudorandom) for each device, creating binding factors [20], [21], or attaching credentials to prove messages' validity [22]. These methods can guarantee users' privacy, but increase the total overhead, as they perform several processes especially during the setup phase. Moreover, they depend on the presence of a trusted party online most of the time.

Several studies exploit the homomorphic features for certain public key schemes to aggregate the electricity consumption for a specific region without revealing the individual consumption values. Reference [23] proposes a full framework to aggregate the electricity readings for the customers in a specific region and guarantees the privacy of customers using additive homomorphic encryption; at the same time, it provides an access control scheme based on ABE scheme. The homomorphic hash function [24] is utilized for mutual authentication between SMs and the main control server. Reference [25] proposes an efficient privacy-preserving demand response scheme to aggregate the electricity demand messages for users in the local area. In [26], the authors propose a new privacy-preserving scheme for aggregating multi-dimensional metering data using homomorphic encryption scheme (EPPA). However, the applied homomorphic schemes in the literature provide high computation and communication overhead. For instance, EPPA [26] scheme that is based on homomorphic Paillier encryption requires time from 100 to 220 ms as messages' number increases, which consider high load especially with the increase of SMs' number. Generally, the utilized homomorphic schemes are not scalable; their performance degrades, as the number of the involved SMs within the same cluster increases [5].

Attempting to overcome the aforementioned disadvantages, we have previously proposed an efficient lightweight security and privacy-preserving scheme that forecasts the future electricity demand for a cluster of HANs in a specific region and limits the cluster's connection with utility only when its total demand needs to be adjusted [27]. Only few research works attempt to conserve consumers' privacy during APs' readings collection, i.e., before SM sends house's consumption to the local CC. Reference [28] proposes a secure in-network data aggregation utilizing an orthogonal chip code and circuit shifting operation to guarantee the confidentiality and anonymity of APs' information. However, this scheme requires sharing the chip code with APs and performing mutual authentication. In addition, SM can reconstruct the original reading for each AP from the mixed data.

While our proposed scheme preserves customers' privacy and data confidentiality during readings aggregation inside HAN. APs are permitted to aggregate their consumptions (without involving SM) employing the lightweight

Lattice-based Homomorphic cryptosystem. Then, SM receives the total encrypted aggregated consumption for HAN and does not know the reading for each AP. Consequently, SMs and the related BS work as relay nodes to forward HAN's total consumption to CC. However, they, i.e., BS and SM, can authenticate the exchanged messages without revealing their contents. The proposed scheme guarantees the security and privacy requirements for the smart grid's connection with customers by efficient and lightweight communication and computation overhead especially on APs. A preliminary version of this work was presented in [1].

III. System Model

A. Network Model

Consider a residential area that consists of the main CC for service provider, i.e., utility company, that is connected to a number of base stations (BSs) located in different areas $BSs = \{BS_1, BS_2, \ldots, BS_h\}$. Each BS is responsible for a cluster of HANs in its local region $HANs = \{HAN_1, HAN_2, \ldots, HAN_m\}$. HAN could be a townhouse or a unit in a building; each HAN has a SM that connects to the house's $APs = \{AP_1, AP_2, \ldots, AP_n\}$. APs also can communicate to each other directly without involving SM. The communication inside HAN is through inexpensive short coverage distance technology, such as Bluetooth or ZigBee. While, the connection between HANs and the corresponding BS is through inexpensive WiFi technology. CC, BSs, and SMs have public keys provided by an independent trusted authority (TA). Each AP has a unique ID that issued to it by TA and stored in a secured memory. Fig 1 shows the system model.

B. Adversary Model and Security Requirements

We consider that CC, BSs, and SMs are honest but curious. However, an adversary $\mathscr{A}$ can eavesdrop the exchanged messages between different parties, i.e., the messages among APs and between them and SM, also the forwarded messages from SMs to BS, to extract consumers' personal information. $\mathscr{A}$ may establish some active attacks, e.g., falsify the captured messages or begin a replay attack; also, $\mathscr{A}$ may compromise SMs. Thus, we should thwart $\mathscr{A}$'s malicious actions by guaranteeing:

- *Consumers' Privacy:* assure that any attacker could not gain any knowledge about HAN's consumption. In addition, CC should not know the detailed consumption pattern for each customer in the region.

- *Authenticity and Data Integrity:* guarantee the confidentiality and authenticity of customers' consumption; even if $\mathscr{A}$ already intercepts a message, he/she cannot extract any knowledge. Likewise, we should ensure messages' integrity; suppose $\mathscr{A}$ attempts to resend/modify a message, we should detect these malicious actions.

C. Design Goals

The main objective of the proposed scheme is fulfilling the security requirements for the network; it should guarantee consumers' privacy. It also should prevent any illegal access/modification of messages/devices in addition to be efficient and lightweight in terms of communication and computation overhead.

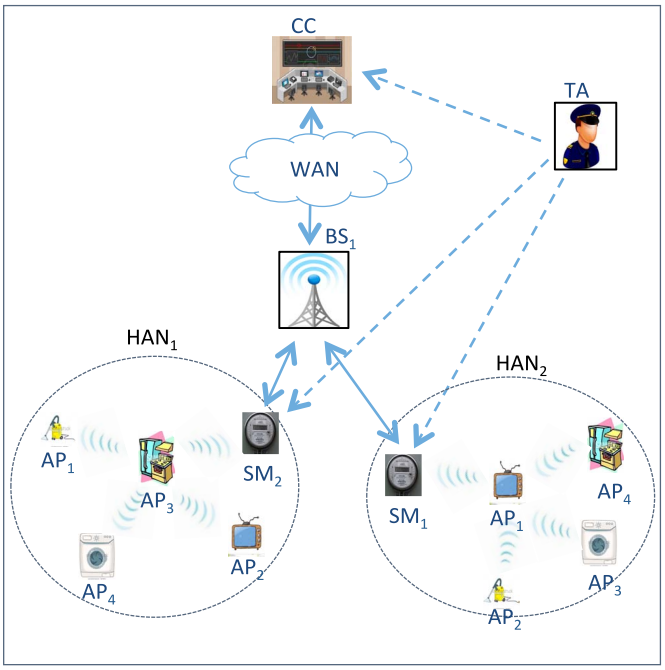

Fig. 1. System Model.

IV. Preliminaries

A. Lattice-Based Homomorphic Encryption Scheme

Our scheme exploits the lightweight lattice-based homomorphic encryption scheme [29], which utilizes the vector space structure to encrypt messages as noisy lattices. So, it guarantees messages' security with low computation complexity, as it mainly performs simple addition and multiplication operations in vector space.

B. Key Generation

The scheme defines five global integer parameters: N is the number of coordinates of plaintext vectors, r is the characteristic of the ring over which they are constructed, l is the maximum number of homomorphic operations that can be done, n is the number of softly disturbed matrices in the public key, and ε_{max} is an upper bound for the coordinates of random vectors used to insert noise.

Let $l_0 = n \times N \times \varepsilon_{max} + (N-1) \times r$, $q = 2 \times l_0 \times (2l+1)$, and $p = q \times r + \varepsilon$ is a prime number, $\varepsilon < l_0$. Then, generate two random $N \times N$ matrices over $GF(p)$: $\boldsymbol{A}$ and $\boldsymbol{B}$, where $\boldsymbol{A}$ is invertible and $\boldsymbol{M} = [\boldsymbol{A} \mid \boldsymbol{B}]$. Also, generate a random scrambling matrix Δ, which is an $\boldsymbol{N \times N}$ diagonal invertible matrix over $\boldsymbol{GF(p)}$. Compute $\ddot{\boldsymbol{M}}_{\boldsymbol{i}} = [\boldsymbol{A}_{\boldsymbol{i}} \mid \boldsymbol{B}_{\boldsymbol{i}}]$ by multiplying $\boldsymbol{M}$ to the left of a random invertible matrix $\boldsymbol{P}_{\boldsymbol{i}}$. Subsequently, Generate a soft noise matrix $\boldsymbol{D}_{\boldsymbol{i}}$, a random $\boldsymbol{N \times N}$ matrix over $\boldsymbol{\{-1, 1\}}$, for each $\boldsymbol{i \in \{1, 2, \ldots, n\}}$. Next compute softly distributed matrix $\dot{\boldsymbol{M}}_{\boldsymbol{i}} = [\boldsymbol{A}_{\boldsymbol{i}} \mid \boldsymbol{B}_{\boldsymbol{i}} + \boldsymbol{D}_{\boldsymbol{i}}\Delta]$. Similarly, compute a hard noise matrix $\boldsymbol{D_0}$ by generating a soft noise matrix then replacing the diagonal values by $\boldsymbol{q}$. Then compute the hardly distributed matrix

$\dot{M_0} = [A_0 \mid B_0 + D_0\Delta]$. Next, choose a permutation operation $\mathscr{P}(.)$, and compute $M_i = \mathscr{P}(\dot{M_i})$, $i \in \{1, 2, \ldots, n\}$.

Finally, the $n+1$ matrices $\{M_0, M_1, \ldots, M_n\}$ are the public key. While the private key consists of the permutation $\mathscr{P}(.)$, the hidden matrix M, and the scrambling matrix Δ.

C. Encryption

First, the plaintext message is constructed as a message vector m in Z_r^N. Then m multiplies by the hard noise matrix M_0. The result is disturbed by adding mM_0 to the summation of n soft noise vectors $\sum_i r_i * M_i$, where r_i are n random vectors with coordinates smaller than ε_{max}. Then, the ciphertext is

$$c = mM_0 + \sum_{i=1}^{n} r_i * M_i. \tag{1}$$

D. Decryption

The decryption operation is based on filtering the added noise. First, the permutation is reversed as

$$\dot{c} = \mathscr{P}^{-1}(c), \tag{2}$$

where $c \in GF(p)^{2N}$ is the ciphertext. Then, the receiver computes the scrambled noise

$$e = \dot{c_D} - \dot{c_U}A^{-1}B, \tag{3}$$

where $\dot{c_D}, \dot{c_U}$ are the undisturbed and disturbed halves of $\dot{c}$. Then the unscrambled noise is

$$\dot{e} = e\Delta^{-}1. \tag{4}$$

For each $\dot{e_j}$ in $\dot{e} = [\dot{e_1} \ldots \dot{e_N}]$, get

$$\ddot{e_j} = \dot{e_j} - \mu, \tag{5}$$

where

$$\mu = \begin{cases} \dot{e_j} \bmod q & \dot{e_j} \bmod q < \frac{q}{2} \\ (\dot{e_j} \bmod q) - q & \text{otherwise.} \end{cases}$$
$$m_j = \ddot{e_j}\, q^{-1}, \tag{6}$$

where $i \in \{1, 2, \ldots, N\}$. Lastly, return the original plaintext:

$$m = (m_1, \ldots, m_N) \tag{7}$$

This cryptosystem resists lattice-based and chosen plaintext attacks so that it assures data's security. The system is suitable for APs with limited capabilities because of its low computation complexity.

V. The Proposed Scheme

Our proposed scheme has two phases: initialization phase, which sets-up the secure connection between APs and CC via SMs and BSs. While reading aggregation phase organizes the aggregation operation of electricity consumption readings.

A. Initialization Phase

TA assigns a pair of public private keys for CC, each BS and each connected SM.

- For CC, its public key parameters are $\{M_{cc0}, M_{cc1}, \ldots, M_{ccn}\}$, where M_{cc0} is the hard noise matrix, and $\{M_{cc1}, \ldots, M_{ccn}\}$ are the n soft noise matrices. APs use this key to encrypt their readings. The CC's private key parameters are $\mathscr{P}_{cc}(.), M_{cc}, \Delta_{cc}$.
- For each BS, the public key is $\{M_{bs0}, M_{bs1}, \ldots, M_{bsa}\}$, where M_{bs0} is the hard noise matrix, and $\{M_{bs1}, \ldots, M_{bsa}\}$ are the a soft noise matrices. Its private key parameters are $\mathscr{P}_{bs}(.), M_{bs}, \Delta_{bs}$.
- According to each SM, its public key is $\{M_{sm0}, M_{sm1}, \ldots, M_{smi}\}$, and the private key is $\mathscr{P}_{sm}(.), M_{sm}, \Delta_{sm}$.
- Each AP has a unique ID issued by TA, $\{AP_1, \ldots, AP_m\}$, where m is the total number of APs in HAN. APs are arranged in a fixed order according to their IDs. The aggregator, AP_s, for each aggregation round is known to SM and APs. This order is fixed and securely sent to all APs in the HAN so that each one automatically knows its turn to be the aggregator.

For instance, if there are five APs in the house, AP_1, AP_2, AP_3, AP_4, and AP_5, then SM arranges them so that AP_1 is the aggregator for first round, AP_2 is the aggregator for second round,. . ., AP_5 is the aggregator for fifth round, and then AP_1 is the aggregator for sixth round, and so on.

- Each AP stores its encrypted ID ID_{j-enc} in a secure place.

$$ID_{j-enc} = ID_j * M_{sm0} + \sum_i r_i * M_{smi} \tag{8}$$

The ID is encrypted by SM's public key, because AP needs to prove its identity to SM when it becomes the aggregator during the aggregation phase.

Moreover, TA assigns certain extra number of IDs for each HAN in the area, e.g., TA sets 20 IDs per HAN so that HAN can have 20 APs at maximum. Then, if the customer needs to add or remove APs, the following procedure is applied:

- If a new extra AP is added to the network, the customer selects one ID for the new AP from the assigned range of IDs for that house (i.e., the IDs are assigned by TA).
- If an old AP is replaced by a new same AP, such as an old AC is replaced by a new AC, then the new AP uses the same ID as the old one.
- If an AP is removed from the network, no procedure is required. As other APs in HAN still can aggregate their readings utilizing the homoromphic feature of the applied cryptosystem.

B. Reading Aggregation Phase

1) Inside Home Area Networks (HANs): At the beginning of each readings' aggregation round, each AP_j in HAN encrypts its reading vector $m_j = (m_1, \ldots, m_w)$ using CC's public key.

$$c_j = m_j * M_{cc0} + \sum_i r_i * M_{cci} \tag{9}$$

Then, it sends c_j to AP_s, i.e., the aggregator for the current round.

$$AP_j \xrightarrow{c_j} AP_s$$

- AP_s computes the total reading c by aggregating the received readings employing the homomorphic addition feature.

$$c = \sum_j c_j \tag{10}$$

- AP_s attaches its encrypted ID (ID_{s-enc}) to the aggregated message and then forwards the message to SM.

$$AP_s \xrightarrow{c,\, ID_{s-enc}} SM$$

- After checking the validity of AP_s's ID, SM attaches timestamp T_v and nonce f vectors, and then signs the received message using $\mathscr{P}_{sm}(.), M_{sm}, \Delta_{sm}$:

$$x = c \| T_v \| f \tag{11}$$
$$\dot{x} = \mathscr{P}_{sm}^{-1}(x) \tag{12}$$
$$e = \dot{x_D} - \dot{x_U} A_{sm}^{-1} B_{sm} \tag{13}$$
$$\dot{e} = e \Delta_{sm}^{-1} = [\dot{e_1}, \ldots, \dot{e_N}] \tag{14}$$

For each $\dot{e_j}$, $i \in \{1, 2, \ldots, N\}$, SM computes

$$\ddot{e_j} = \dot{e_j} - \mu, \tag{15}$$

where $\mu = \begin{cases} \dot{e_j} \bmod q & \dot{e_j} \bmod q < \frac{q}{2} \\ (\dot{e_j} \bmod q) - q & \text{else} \end{cases}$

$$y_j = \ddot{e_j}\, q^{-1}, i \in \{1, 2, \ldots, N\} \tag{16}$$
$$Y = (y_1, \ldots, y_N) \tag{17}$$

SM then forwards Y to the local BS.

$$SM \xrightarrow{Y} BS.$$

2) At Local Base Station (BS): BS first verifies each SM's signature and obtains its message ($x = c \mid T_v \mid f$):

$$x = Y * M_{sm0} + \sum_i r_i * M_{smi} \tag{18}$$

Also, it checks the validity of timestamp T_v and nonce f.

- Then, BS aggregates the received aggregated consumptions from different SMs in the area

$$C = \sum_k c_k, \tag{19}$$

where k is the number of connected HANs to local BS.

- Next, BS signs the total aggregated consumption for the area by its private key $\mathscr{P}_{bs}(.), M_{bs}, \Delta_{bs}$:

$$g = C \| T_u \| q \tag{20}$$
$$\dot{g} = \mathscr{P}_{bs}^{-1}(g) \tag{21}$$
$$w = \dot{g_D} - \dot{g_U} A_{bs}^{-1} B_{bs} \tag{22}$$
$$\dot{w} = w \Delta_{bs}^{-1} = [\dot{w_1}, \ldots, \dot{w_N}] \tag{23}$$

For each $\dot{w_j}$, $i \in \{1, 2, \ldots, N\}$, BS computes

$$\ddot{w_j} = \dot{w_j} - \mu, \tag{24}$$

where $\mu = \begin{cases} \dot{w_j} \bmod q & \dot{w_j} \bmod q < \frac{q}{2} \\ (\dot{w_j} \bmod q) - q & \text{else} \end{cases}$

$$d_j = \ddot{w_j}\, q^{-1}, i \in \{1, 2, \ldots, N\} \tag{25}$$
$$D = (d_1, \ldots, d_N) \tag{26}$$

BS then forwards the aggregated D to CC as an electricity reading message for one unit, i.e., CC deals with BS and its connected cluster of HANs as one HAN.

- CC verifies BS's signature on D and then decrypts C using its private key parameters:

$$\dot{c} = \mathscr{P}_{cc}^{-1}(C) \tag{27}$$
$$s = \dot{c_D} - \dot{c_U} A_{cc}^{-1} B_{cc} \tag{28}$$
$$\dot{s} = s \Delta_{cc}^{-1} = [\dot{s_1} \ldots \dot{s_N}] \tag{29}$$

For each $\dot{s_k}$, $k \in \{1, 2, \ldots, N\}$, CC computes

$$\ddot{s_k} = \dot{sk} - \mu_0, \tag{30}$$

where $\mu_0 = \begin{cases} \dot{s_k} \bmod q & \dot{s_k} \bmod q < \frac{q}{2} \\ (\dot{s_k} \bmod q) - q & \text{else} \end{cases}$

$$m_k = \ddot{s_k}\, q^{-1}, k \in \{1, 2, \ldots, N\} \tag{31}$$
$$m = (m_1, \ldots, m_N) \tag{32}$$

CC now obtains the total aggregated consumption for BS's area in plaintext m.

3) Control Messages: If any AP from group 3 or 4 needs to send a request to CC, e.g., to change its load, or duration of usage, it can directly send its request to CC via SM and BS and does not wait for the new aggregation round. These messages are called control messages.

For instance, if AP_j wants to send a request R, it first adds a timestamp T_d and random nonce L to R, concatenates its ID (ID_{j-enc}), $n_j = R \| ID_{j-enc} \| T_d \| L$, and then encrypts n_j by CC's public key:

$$z_j = n_j * M_{cc0} + \sum_i r_i * M_{cci} \tag{33}$$

- AP_j then sends control message z_j to SM, which signs the message and forwards it to CC via BS, i.e., BS verifies SM's signature and signs z_j again before forwards it to CC.

VI. Security Analysis

The main objective of the proposed scheme is to preserve the privacy of HAN's customers and does not expose daily habits and lifestyle of houses' owners from their electricity consumptions. In addition, our scheme aims to satisfy basic security requirements, such as confidentiality and messages' integrity.

Privacy: The electricity consumption for HAN can reveal the daily behaviours of householders so that preserving their privacy is a major concern. The proposed scheme guarantees that no party even SM or BS knows the individual reading for each AP. AP_s cannot analyze the daily life pattern

for householders too, as the received individual readings are encrypted; in addition, the aggregator is different for each reading round. Although we assume that APs' secret IDs are protected, $\mathscr{A}$ will not gain a lot of information if he/she manages to compromise one of the APs, i.e., he/she can only know the reading for that AP and cannot analyze the private life for householders by that data only. Furthermore, if the compromised AP by chance is the aggregator $\boldsymbol{AP_s}$, $\mathscr{A}$ cannot extract any data, as the messages sent to $\boldsymbol{AP_s}$ are encrypted and only CC has the decryption key. However, we assume that $\mathscr{A}$ cannot compromise APs and cannot obtain their secure IDs, because if $\mathscr{A}$ can physically compromise APs, e.g., $\mathscr{A}$ can enter the house, he/she can easily obtain the readings for APs by him/herself and does not need to snoop/modify and analyze the messages. The same happens when $\mathscr{A}$ attempts to compromise SM. Since SM is just a relay node, it only forwards the received encrypted messages. According to BS, it receives the total house's consumption, but this received message is encrypted so that BS cannot extract any knowledge about electricity consumption for householders. While, CC receives the total aggregated consumption for the whole area so that it cannot extract any private data about real-time consumption pattern of a specific house/customer.

Authenticity and Messages' Confidentiality and Integrity: Only authorized parties have access to messages' contents. APs' readings cannot be revealed to anybody even $\boldsymbol{AP_s}$, which receives only the encrypted version of these readings. According to the total aggregated consumption for HAN, neither $\boldsymbol{AP_s}$ nor related SM can decrypt it, only CC can. As well, BS cannot extract any knowledge from the total consumption of the whole area, as it processes the messages in their encrypted versions. Moreover, messages integrity and confidentiality are also guaranteed. The proposed scheme guarantees the data confidentiality, as none of the participated parties can know the reading of each AP; only the authorized party, CC, can decrypt the total aggregated consumption for the whole area. Furthermore, the message is protected against different attacks. If $\mathscr{A}$ successes to compromise SM, he/she cannot interpret the message's contents, as SM does not own the decryption key. The same happens if $\mathscr{A}$ is powerful enough to compromise the BS (BS is more protected and can resist attacks more than SM). Therefore, any eavesdropping attack does not succeed. In addition, the proposed scheme assures that no illegal party can modify/access the exchanged messages. $\mathscr{A}$ cannot forge the transmitted messages from HANs to the connected BS, as $\mathscr{A}$ does not know SM's private key to falsify its signature. In addition, $\mathscr{A}$ does not have access to BS's private key so that he/she cannot mimic its signature. Replay attacks do not succeed too, because of the attached timestamps and nonce values.

Security of utilized cryptosystem is guaranteed by the hardness of hidden lattice problem (HLP) [30] that is based on disorganizing the lattice in a way so that no party can extract the original lattice from its disturbed version, i.e., HLP disturbs lattice $\boldsymbol{l}$ by a specified technique to be $\bar{\boldsymbol{l}}$ so that $\mathscr{A}$ cannot extract $\boldsymbol{l}$ from $\bar{\boldsymbol{l}}$. Lattice-based homomorphic encryption scheme exploits HLP to select the private key parameters:

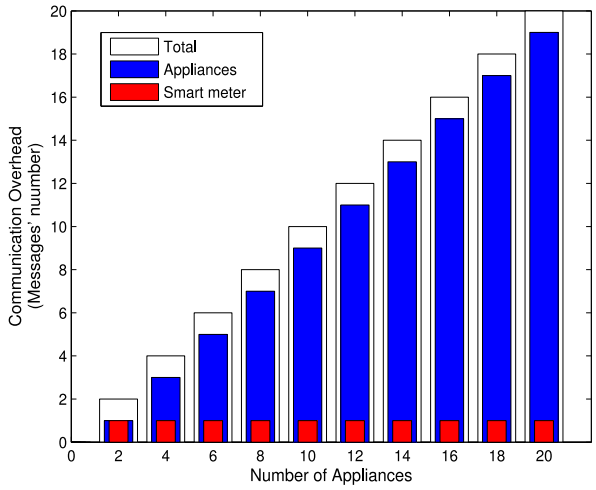

Fig. 2. Communication Overhead per Reading Round.

the hidden lattice $\boldsymbol{M}$, the scrambling matrix Δ and the permutation $\mathscr{P}(.)$. Accordingly, to break the system, $\mathscr{A}$ should obtain the secret permutation and then retrieve the disturbed columns' indexes to solve the corresponding HLP. To guarantee the security and robustness of the system against attacks specially chosen plaintext and lattice-based attacks, the main four parameters of the cryptosystem, i.e., $\boldsymbol{l}$, $\boldsymbol{r}$, $\boldsymbol{N}$, and $\boldsymbol{p}$, must be chosen carefully so that the cost of non-disturbed columns searching operation is unbearable for powerful attackers. If $\boldsymbol{N \geq 50}$, this operation requires $\binom{2N}{N} \geq \binom{100}{50} \approx \boldsymbol{2^{100}}$. In addition, choosing a high-dimension lattice, e.g., $\boldsymbol{600}$, can further increase the hardness of HLP problem. Following these constraints during parameters' selection can enhance the proposed scheme's resistance to attacks [30].

VII. Performance Evaluation

This section analyzes the performance of the proposed scheme in terms of communication and computation overhead.

A. Communication Overhead

The number of exchanged messages between different parties every readings' aggregation round is too small. The number of messages that should be sent by the limited-capability devices SM and APs is trivial. During each readings' aggregation round, each AP sends only one reading message. As well, $\boldsymbol{AP_s}$ just sends the aggregated message. According to SM, it only forwards the total aggregated readings message for the house to BS, which in turn forwards the total aggregated readings message for the whole area to CC. According to control messages, APs from group 3 and 4 only need to send their requests to CC. These messages are sent directly to SM, which forwards them to CC via BS. Generally, the house could contain two or three of these APs, e.g., a house includes one EV, one AC, and one clothes washer. Assume that each HAN has three of these APs for maximum; that means three control messages. These messages are sent occasionally; assume that each AP needs to send one or two control messages per day so that the maximum number of control messages per day for

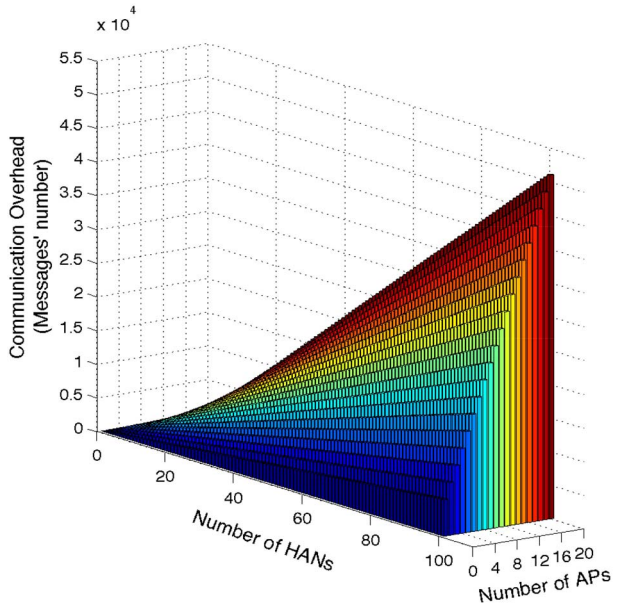

Fig. 3. Communication Overhead per Day.

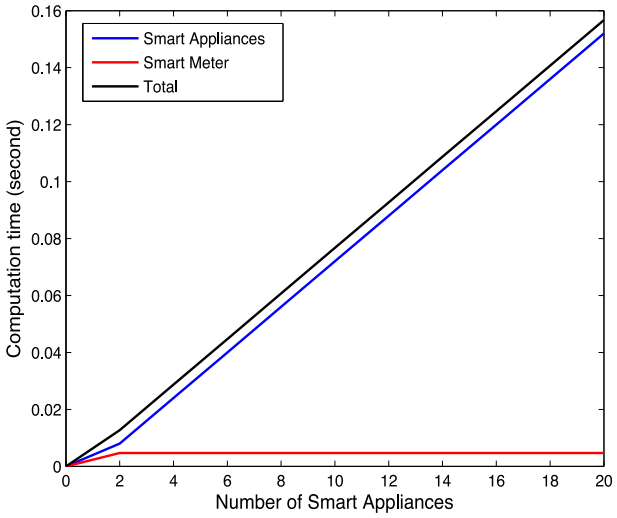

Fig. 4. Computation Delay per Reading Round.

one HAN is six messages. In summary, the total communication burden for BS, SM, and for each AP is one message per reading round, which considers insignificant load.

Fig 2 shows the communication overhead inside HAN for every reading round. It can be seen that the communication delay is increased from **2** messages in two-appliances case to **20** messages in twenty-appliances case. Although the communication overhead is expected to increase as the number of APs increases, its growth is limited and affordable by the restricted-resources devices in the house. Fig 3 shows the total communication delay for the area per day, after adding the control messages overhead, as the number of APs in the house and HANs in the area increases. Although the APs' number increases, each AP, SM, and BS still have to send one message only per round, which means a fixed number of messages are sent from each HAN in the area per day.

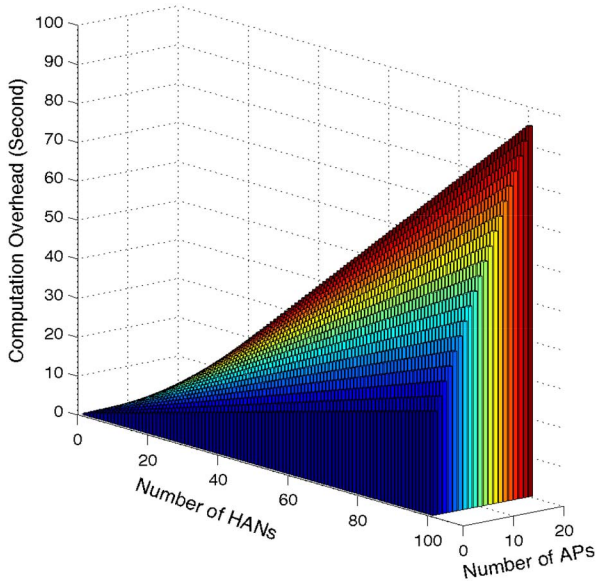

Fig. 5. Computation Delay per Day.

TABLE I
THE NUMBER OF OPERATIONS FOR SMART DEVICES

Number of Operations	Per Round	Per Day
$APs(Group-1\&2)$	$1*T_e$	$h*T_e$
$APs(Group-3\&4)$	$1*T_e$	$(h+2)*T_e$
SM	$1*T_s$	$(h+6)*T_s$

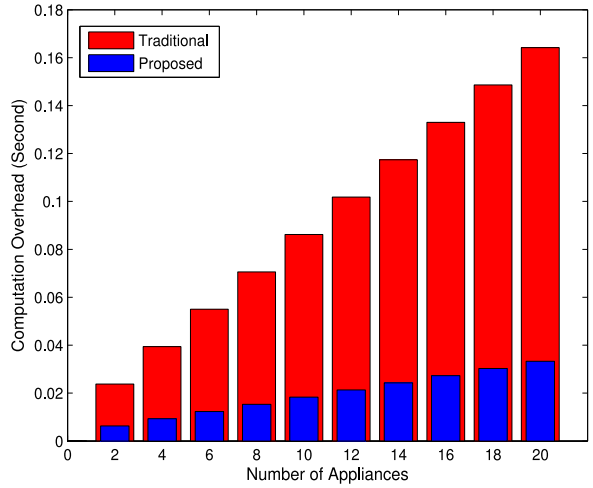

Fig. 6. Computation Delay per Reading Round.

According to number of HANs, the total communication overhead for the area is increased as the number of connected HANs increases. However, the communication load is affordable by different parties in the network, which means light communication overhead.

B. Computation Overhead

Each AP in the HAN has to perform one encryption process per round for its reading except APs, which only performs simple neglected summation operation. The total number of

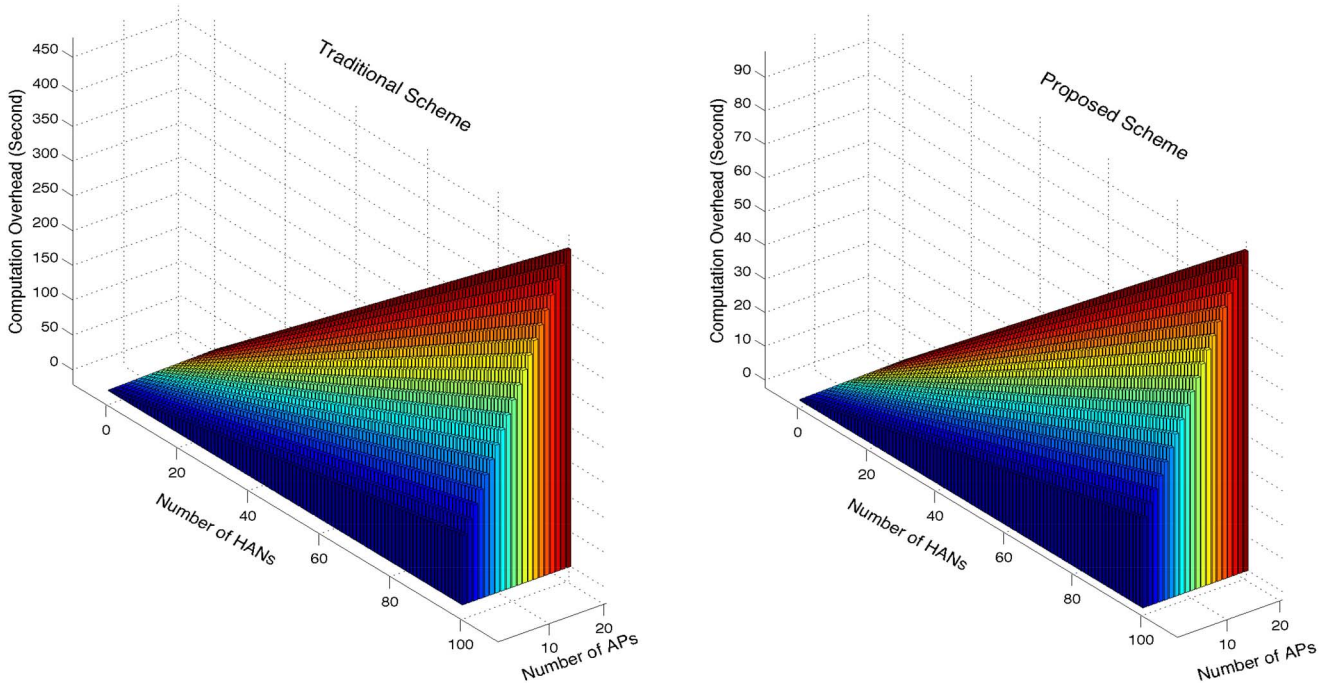

Fig. 7. Computation Delay per Day.

encryption operation per HAN is $n-1$, where n is the number of APs in the HAN. These encryption processes do not require high computation capabilities, because the deployed cryptographic scheme is lightweight scheme especially for encryption; it only consists of simple addition and multiplication operations. Next, SM signs the received aggregated message from AP_s before forwarding it to CC via BS. That means one signing process for each HAN per round. Then, BS aggregates the received readings from the connected cluster, signs the result, and then sends it to CC. So, if the total number of HANs in BS's area is m, then the total computation overhead for the area equals $[m*([(n-1)*T_e]+T_s+T_v)]+T_s+T_v+T_d$, where T_e is the computation time for one encryption process, T_d is the time for one decryption process, T_s is the time for one signing process, and T_v is the time for one verification process.

In addition, If HANs have group 3 or 4 APs, then control messages have to be encrypted by these APs and sent to CC via SM and BS. Assume that each house has up to three group 3 or 4 APs, then six control messages are sent from HAN per day that need six encryption operation maximally. SM, i.e., also BS, needs to sign these control messages. However, these messages will not impact on the total computation duty for different parties per day, because they are few and sent occasionally. Table I presents the number of operations per reading round and per day for each AP and SM, where h is the number of rounds per day.

Assume that the hidden lattice dimension is **600** to resist the lattice-based attack, $n = 9, r = 2, p \approx 260, \varepsilon_{max} = 1024, l_0 \approx 219,$ and $q = 2^{21}.2^{38}$. The size of the public parameters is $2N^2(n+1)Log_2(p)$. The speed of one encryption/verification operation is the cost of $(n+1).N$ addition processes of vectors of lengths $2N$ and $Log_2(p)$ bits plus the cost of multiplying two random vectors with length $Log_2(\varepsilon_{max})$ bits. While the decryption/signing speed is the cost of two $N \times N$ matrix multiplication in $GF(p)$, which equals $2N^2.Log_2(p)^2$. Using a MATLAB simulator on a 3.20 GHz-processor with 6.00GB RAM, we study the computation delay for our proposed scheme.

Fig 4 presents the computation load for each AP and for SM every reading round. It can be noticed that the computation overhead for each AP is the same and does not affected by the number of APs in the house. Moreover, SM load does not change, which is expected, as it requires to perform one signing process regardless the number of messages included. Fig 5 points out the total computation load for the whole cluster per day as numbers of APs and HANs increase. As indicated, the computation overhead increases by the increase of APs and HANs' numbers, but still within a low maximum limit; the total computation delay for a cluster of **100** HANs that each one of them has **20** APs is around **90** second per day.

Practically speaking, our proposed scheme is feasible for the restricted-computation capabilities APs. Considering the Smart Grid Smart City data set provided by the Australian Department of Industry, Innovation and Science [31], for instance, which includes a data set of different household APs' readings at different times per one HAN in Sydney, e.g., at November 16 - 2013, almost at 3:12 pm, the current APs' readings for customer **11178213** are measured as: television **388.68**, computer **124.799**, stove **44.474**, AC **1.711**, and light **11.474** kWh, it can be seen that the APs readings' range is limited by the maximum value **731625** kWh. So, the readings' values do not require a significant storage memory. In addition, encrypting the APs' readings does not provide high computation burden on the AP, as the encryption operation consists of generating a fixed series of random

numbers and then computing trivial summation and multiplication operations. Consequently, simple cheap processing device embedded on the APs, such as Raspberry PI [32], is enough to implement that light encryption process.

The existing privacy schemes cannot be applied on APs, since they have to perform complex crypto-operations, such as exponentiation and pairing, which require high computation capabilities not owned by APs. Although the current cryptosystems are not applicable in these restricted-resources APs, we compare the performance of the proposed scheme with a traditional homomorphic Paillier-based scheme, as several privacy-preserving schemes in the literature utilize Paillier cryptosystem because of its additive homomorphic feature, also it is robust against privacy attacks.

In Fig 6, we compare the total computation time per HAN for the proposed scheme versus the traditional one per round, as the number of APs at house increases. Clearly, the computation delay for the proposed scheme is much less than the traditional one, especially at the high number of APs in the house. The computation delay goes from 6.3 to 33.3 ms for the proposed scheme while the traditional scheme's delay increases from 23.8 to 164.2 ms, when the number of APs increases from **2** to **20** APs.

While Fig 7 shows the total computation delay for the whole BS's connected area per day and presents the effect of the increase in APs' number also the increase in the number of HANs of the BS's cluster. It is shown that our scheme takes less computation time compared to the Paillier-based schemes, especially as numbers of APs and HANs increase. Although the computation overhead increases in our scheme, the resulted computation delay is limited and affordable by APs. The computation delay per day for a cluster of **100** HANs with **20** APs each is around **90** second for proposed scheme versus **450** second for traditional schemes. In conclusion, the proposed scheme guarantees privacy and security requirements for the residential consumers with low computation and communication overhead even for limited-computation capabilities APs.

VIII. Conclusion

In this paper, we have proposed a lightweight lattice-based homomorphic security and privacy-preserving scheme that secures the electricity consumption aggregation operation for HANs in residential areas. The proposed scheme depends on house's APs to aggregate their consumption among themselves without involving the connected SM utilizing the lightweight lattice-based homomorphic cryptosystem to secure their readings. However, SMs and the intermediate BS can validate the messages' authenticity without decrypting them. The security analysis and simulation results show that the proposed scheme guarantees consumers' privacy and messages' confidentiality and integrity, at the same time, ascertains lightweight communication and computation overhead. So, our proposed scheme is suitable for limited-computation resources APs. In the future work, we will study the impact of connected EVs as storage units on the performance of our proposed scheme.

References

[1] A. R. Abdallah and X. S. Shen, "Lightweight lattice-based homomorphic privacy-preserving aggregation scheme for home area networks," in *Proc. WCSP*, Hefei, China, Oct. 2014, pp. 1–6.

[2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 944–980, Dec. 2012.

[3] J. Liu, Y. Xiao, S. Li, W. Liang, and C. L. P. Chen, "Cyber security and privacy issues in smart grids," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 981–997, Dec. 2012.

[4] M. Kuzlu, M. Pipattanasomporn, and S. Rahman, "Communication network requirements for major smart grid applications in HAN, NAN and WAN," *Comput. Netw.*, vol. 67, pp. 74–88, Jul. 2014.

[5] Z. Erkin, J. R. Troncoso-Pastoriza, R. L. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 75–86, Mar. 2013.

[6] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proc. IEEE*, vol. 100, no. 1, pp. 210–224, Jan. 2012.

[7] O. Tan, D. Gunduz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1331–1341, Jul. 2013.

[8] Z. Chen and L. Wu, "Residential appliance DR energy management with electric privacy protection by online stochastic optimization," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1861–1869, Dec. 2013.

[9] X. He and C. Kuo, "A distortion-based approach to privacy-preserving metering in smart grids," *IEEE Access*, vol. 1, pp. 67–78, May 2013.

[10] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor, "Smart meter privacy: A theoretical framework," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 837–846, Jun. 2013.

[11] W. Jia, H. Zhu, Z. Cao, X. Dong, and C. Xiao, "Human-factor-aware privacy-preserving aggregation in smart grid," *IEEE Syst. J.*, vol. 8, no. 2, pp. 598–607, Jun. 2014.

[12] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 99–107, Jun. 2010.

[13] Z. M. Fadlullah, N. Kato, R. Lu, X. Shen, and Y. Nozaki, "Toward secure targeted broadcast in smart grid," *IEEE Commun. Mag.*, vol. 50, no. 5, pp. 150–156, May 2012.

[14] D. He *et al.*, "Secure service provision in smart grid communications," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 53–61, Aug. 2012.

[15] J. Liu, Y. Xiao, and J. Gao, "Achieving accountability in smart grid," *IEEE Syst. J.*, vol. 8, no. 2, pp. 493–508, Jun. 2014.

[16] X. Li *et al.*, "Securing smart grid: Cyber attacks, countermeasures, and challenges," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 38–45, Aug. 2012.

[17] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. S. Shen, "A lightweight message authentication scheme for smart grid communications," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 675–685, Dec. 2011.

[18] H. Li, R. Lu, and X. Shen, "An efficient Merkle-tree-based authentication scheme for smart grid," *IEEE Syst. J.*, vol. 8, no. 2, pp. 655–663, Jun. 2014.

[19] H. Nicanfar, P. Jokar, and V. Leung, "Efficient authentication and key management mechanisms for smart grid communications," *IEEE Syst. J.*, vol. 8, no. 2, pp. 629–640, Jun. 2014.

[20] C. I. Fan and Y. L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 666–675, Feb. 2014.

[21] C. Rottondi, G. Verticale, and C. Krauß, "Distributed privacy-preserving aggregation of metering data in smart grids," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1342–1354, Jul. 2013.

[22] T. W. Chim, S. M. Yiu, L. C. K. Hui, and V. O. K. Li, "Privacy-preserving advance power reservation," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 18–23, Aug. 2012.

[23] S. Ruj and A. Nayak, "A decentralized security framework for data aggregation and access control in smart grids," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 196–205, Mar. 2013.

[24] Y. S. Kim and J. Heo, "Device authentication protocol for smart grid systems using homomorphic hash," *J. Commun. Netw.*, vol. 14, no. 6, pp. 606–613, Dec. 2012.

[25] H. Li *et al.*, "EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 8, pp. 2053–2064, Aug. 2014.

[26] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1621–1631, Sep. 2012.
[27] A. R. Abdallah and X. S. Shen, "A lightweight lattice-based security and privacy-preserving scheme for smart grid," in *Proc. IEEE GLOBECOM*, Austin, TX, USA, Dec. 2014, pp. 668–674.
[28] Y. Yan, Y. Qian, and H. Sharif, "A secure data aggregation and dispatch scheme for home area networks in smart grid," in *Proc. IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011, pp. 1–6.
[29] C. A. Melchor, G. Castagnos, and P. Gaborit, "Lattice-based homomorphic encryption of vector spaces," in *Proc. IEEE ISIT*, Toronto, ON, Canada, Jul. 2008, pp. 1858–1862.
[30] C. Melchor and P. Gaborit, "A lattice-based computationally-efficient private information retrieval protocol," in *Proc. WEWoRC*, Bochum, Germany, Jul. 2007, pp. 50–54.
[31] *Smart Grid Smart City*. Accessed on Jan. 20, 2016. [Online]. Available: https://data.gov.au/dataset/smart-grid-smart-city-customer-trial-data/resource/63d2b1cd-f453-4440-8bb7-ed083326f5ae?inner_span=True
[32] *Rasberry PI*. Accessed on Jan. 20, 2016. [Online]. Available: https://www.raspberrypi.org

Asmaa Abdallah received the B.Sc degree in computer and control engineering, and the M.Sc. degree in mobile networks from Suez Canal University, Egypt, in 2003 and 2007, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo. Her research interests include security and privacy in smart grid, wireless network security, and mobile computing.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the B.Sc. degree from Dalian Maritime University, China, in 1982, and the M.Sc. and Ph.D. degrees from Rutgers University, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering. He is a Professor and the University Research Chair of the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He is an elected member of the IEEE ComSoc Board of Governors and the Chair of Distinguished Lecturers Selection Committee. He served as the Technical Program Committee Chair/Co-Chair for IEEE Infocom'14 and IEEE VTC'10 Fall; the Symposia Chair for IEEE ICC'10; the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08; the Technical Program Committee Chair for IEEE Globecom'07; the General Co-Chair for ACM Mobihoc'15, Chinacom'07 and QShine'06; and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for IEEE NETWORK, *Peer-to-Peer Networking and Application*, and *IET Communications*; the Founding Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Networks*; and *ACM/Wireless Networks*, and the Guest Editor for IEEE JSAC, IEEE WIRELESS COMMUNICATIONS, the IEEE COMMUNICATIONS MAGAZINE, and *ACM Mobile Networks and Applications*. He was a recipient of the Excellent Graduate Supervision Award in 2006; the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo; the Premier's Research Excellence Award in 2003 from the province of Ontario, Canada; and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is a registered Professional Engineer in Ontario, Canada; an Engineering Institute of Canada Fellow; a Canadian Academy of Engineering Fellow; and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.