Cluster Computing (2022) 25:263–278
https://doi.org/10.1007/s10586-021-03387-0

LPM2DA: a lattice-based privacy-preserving multi-functional and multi-dimensional data aggregation scheme for smart grid

Saleh Darzi[1] · Bahareh Akhbari[1] · Hassan Khodaiemehr[2,3]

Received: 16 September 2020 / Revised: 3 August 2021 / Accepted: 7 August 2021 / Published online: 27 August 2021

Abstract

The smart grid provides efficient transmission of energy and data. However, the frequent gathering of users' consumption data discloses users' privacy. Plenty of data aggregation schemes have been introduced to preserve the privacy of users' private information. Unfortunately, with the advent of quantum machines, most of these schemes will be rendered vulnerable and insecure. Hence, to preserve privacy and provide other security services like integrity and authentication in smart grid, we attempt to introduce a secure scheme based on lattice-based cryptography named LPM2DA: a lattice-based privacy-preserving multi-functional and multi-dimensional data aggregation scheme. The proposed scheme enables the control center to acquire temporal and spatial aggregation of multi-dimensional data in a privacy-preserving manner. Also, it empowers the control center to calculate different statistical functions such as mean, variance, and skewness on users' multi-dimensional data. Eventually, through analytical evaluation, we illustrate the efficiency of the proposed scheme.

Keywords Privacy-preserving · Lattice-based homomorphic encryption · Chinese remainder theorem · Integrity and authentication · Multi-functional · Smart grid

1 Introduction

With the improvement of communication structures and information technologies, the smart grid as the new generation of the power grid was officially defined by Energy Independence, and Security Act of 2007 (EISA-2007) [1]. As opposed to the traditional grid that was comprised of presentational components with no data connections between them, the smart grid allows for two-way communication of data and energy between its components [2]. The smart grid is a distributed network that utilizes additional renewable sources of energy; therefore it is reliable and makes it easy to monitor and control the energy distribution. Specifically, the smart grid provides fault detection and self-healing. As a result, it is dependable and efficient for the transmission of energy. Thus, the term "smart" is expressed in defining all these new grid's abilities and facilities. Concretely, the traditional grid's failures and blackouts like Northeastern Blackout of 2003, which affected more than fifty-five million people, stand in total contrast to the new energy grid [3].

Among seven domains of smart grid defined by the National Institute of Standards and Technology (NIST), consider the customer-side network in which all the appliances in the residential area (or in a house) are reporting their data to the smart meter (SM) installed in each home area network (HAN). Subsequently, SM reports the collection of consumption data of all these appliances to the control center (CC), whose obligation is to scrutinize these data and monitor the energy distribution and the grid's stability [4].

✉ Bahareh Akhbari
akhbari@kntu.ac.ir

Saleh Darzi
salehdarzi@email.kntu.ac.ir

Hassan Khodaiemehr
ha.khodaiemehr@kntu.ac.ir

[1] Faculty of Electrical Engineering, K. N. Toosi University of Technology, P. O. Box: 16315-1355, Tehran, Iran

[2] Faculty of Mathematics, K. N. Toosi University of Technology, P. O. Box: 16765-3381, Tehran, Iran

[3] School of Mathematics, Institute for Research in Fundamental Sciences (IPM),
P. O. Box: 19395-5746, Tehran, Iran

Although two-way communication gave advantages to the smart grid over the traditional grid, it also could bring some vulnerabilities, privacy disclosure, and security menaces. Energy theft, fraud, impersonation, and reducing network reliability, and learning personal patterns are the most common privacy and security threats of the smart grid. Specifically, the privacy of consumers' information is of paramount significance, because some private information can be deduced from users' reports, such as their habits, lifestyle, the number of people residing in a household, and in some cases, the type of appliances can be recognized [3, 4]. It should be noted that privacy is a sophisticated and multidimensional concept, and it should be defined accurately so that we could utilize the proper combination of the privacy model and the privacy-preserving technique to accomplish a practical scheme. Since there exist multiple privacy models and privacy-preserving techniques like perturbation, randomization, generalization, and etc., the primary step is to determine the smart grid's data type [5].

Considering the research directions for smart grid's security concerns, to tackle these threats and to maintain privacy, diverse approaches have been adopted so far: (1) Hardware equipment to mask the transmitting data. This approach is costly and not suitable for smart grid network considering the considerable number of smart meters. (2) Concealing each smart meter's data with noise. Specifically, in this approach, the addition of private data with a particular noise will be transmitted to the control center. The main drawback to this approach is the low accuracy of data recovered by CC. (3) Utilization of cryptographic techniques to ensure the privacy of users. This approach comprises three different types: anonymization, authentication, and data aggregation (DA) schemes [6]. It is worth noting that the utilization of cryptographic techniques is one of the most practical approaches regarding security concerns such as privacy, integrity, and authentication, and it is employed in many applications like smart grid, internet-of-things (IoT), and e-health systems [7].

By means of anonymization techniques, each entity disguises its true identification by a pseudo-ID. However, for traceability of errant entities, these lightweight techniques need an online trusted authority (TA) who knows the connection between the real and pseudo ID. Authentication techniques could be used with various architectures, and their critical problem is "key management". Hence, a lightweight and low-cost scheme is required to thoroughly address the authentication, which is the primary shield to challenge security concerns [8]. Authentication techniques are deployed for authenticity confirmation of each entity in each phase. Therefore, by the growth of users in the network, these authentication techniques add a certain amount of delay to each phase [6]. For this reason, some schemes (like [8]) try to eliminate the key management and focus on mutual authentication of network components before the communication phase via constrained operations like hash functions. To preserve privacy and to hinder the managing unit CC, or even a fraudulent employee in the operation center from acquiring personal information of each individual, data aggregation (DA) schemes have been suggested. The structure of a DA scheme requires a powerful entity called gateway (GW) between CC and the residential area's smart meters to combine the consumption information of that residential area. In order to preserve the private information of individual HANs, data aggregation schemes deploy homomorphic encryption systems like Paillier, BGN, El-Gamal, RSA, etc. [6].

Lu et al.'s scheme, EPPA [9], preserves the privacy of individual reports by employing the Paillier encryption scheme. By relying on the super-increasing sequence, each user could transmit its multi-dimensional data in a way that the control center could calculate the aggregation of each dimension via a recursive algorithm. It also utilizes batch verification to decrease the authentication overhead. Zhang et al. [10] has also proposed a DA scheme based on the Paillier encryption scheme that could aggregate users' information in both temporal and spatial form. In this scheme's system model, to achieve temporal and spatial aggregation, a network is formed in which users could communicate and transmit shares of their data to each other without using encryption. Chen et al. [11] adopted an additive homomorphic encryption scheme, namely BGN cryptosystem, to present its DA scheme MuDA. By means of computing and transmitting one, two, and three kinds of aggregation by GW, the control center could obtain the average, variance, and one-way analysis of variance (ANOVA) of users' data, respectively. Since the GW adds a particular noise to the aggregated data to make the scheme resistant to differential attack, it would render the aggregated data obtained by CC less accurate.

Ge et al. [12] proposed another DA scheme called FGDA, which is capable of calculating the average, variance, and skewness via transmitting a single aggregated data by GW. Furthermore, it supports users' fault-tolerance by trusted authority's participation in the aggregation phase. Based on a session key between users and GW, the authentication of the message would be assured. Ni et al. [13] proposed a DA scheme relying on the lifted El-Gamal encryption and BBS signature scheme. Based on GW's role in noise addition and aggregation of honest smart meters' data, the features like fault-tolerance and differential privacy are guaranteed. Moreover, it utilizes zero-knowledge-proof techniques to screen unusual reports. In the DA scheme proposed by Ming et al. in [14], to efficiently preserve privacy, heavy operations such as bilinear pairings are eliminated, and the elliptic curve and El-Gamal encryption schemes are employed. Besides, their scheme is also resistant to various attacks and provides the transmission of multi-dimensional data. The LFDA scheme [15] tries to eliminate the homomorphic encryption from the DA scheme and achieves a lightweight DA scheme that supports the fault-tolerance property. Specifically, this scheme is a masking-based data aggregation and has utilized the notion of a flag bit and identity authentication to ascertain the correctness of the aggregated result. Although the general consumption of electricity is monitored in each geographical area, the privacy of each user, the integrity of their data, and the anonymity of their identity would be ensured efficiently.

Security of these schemes relies on the hardness of integer factorization and/or discrete logarithm problems, which have been broken by Shor's quantum algorithm [16]. This obviously indicates the need for using secure and effective post-quantum cryptography like lattice-based cryptography, which has not been broken by quantum attacks yet and is supposed to provide post-quantum security and entails straightforward procedures like polynomial addition and multiplication.

However, only a few researches exploit lattice-based cryptography to preserve privacy [17–20]. LRSPPP [17] utilizes a ring learning with error (R-LWE) based cryptosystem and an R-LWE based signature scheme. Abdallah et al.'s scheme [18] uses the revised version of the NTRU cryptosystem and the new NTRU signature scheme (NSS). Both schemes [17, 18] estimate the energy demand for a constellation of households by means of a forecasting function. Technically, load forecasting provides prudent decision-making in the network and could be achieved through different algorithms [21, 22]. Similar to the DA schemes, the electrical load forecasting strategy also deals with the problem of outsourcing its necessary operations (data pre-processing phase and load prediction phase) to a powerful entity like cloud servers to have a fast and accurate decision [23]. On the one hand, forecasting lessens the communication overhead. On the other hand, it adds a database to each building area network (BAN) for storing the demands that BAN has access to it. Abdallah et al. [19] proposed another scheme, in which a lattice-based homomorphic scheme is used to encrypt and sign the individual and aggregated data. It categorizes the smart household appliances into four groups and lets them aggregate their consumption without SM's participation. Concentrating on privacy assurance, Li et al. [20] introduced a DA scheme, namely PDA relying on an R-LWE based somewhat homomorphic encryption scheme. It accomplishes multi-dimensional data and empowers CC to calculate functions such as the mean and variance on these data.

Therefore, compared to the other privacy-preserving schemes in the smart grid, we take a novel approach to work towards a different goal. More precisely, given that it is hard to update the construction of the power grid and due to the imminent quantum attacks, our goal is to achieve a higher level of security, i.e., post-quantum security. However, since the lattice-based schemes in smart grid network are truly heavy or these schemes tend to change the typical system model to propose a solution for the privacy of users' data, we show that the proposed approach does not only have the appropriate structure, but also is more efficient. Thus, to preserve privacy, ensure integrity and provide authentication, we propose a secure Lattice-based Privacy-preserving Multi-functional and Multi-dimensional Data Aggregation scheme (LPM2DA) in smart grid, which enables CC to acquire temporal and spatial aggregation of multi-dimensional data. Specifically, by means of acquiring the spatial aggregation of users' data, CC could realize electricity theft or power leakage and make better conscious pricing choices; and by computing temporal aggregation, CC could attain the bill. Also, to find out about the balance and uniformity of usage data in the smart grid network, our scheme empowers CC to calculate different statistical functions such as the mean, variance, and skewness on users' multi-dimensional data. Concretely, we show that the proposed scheme is resistant against various attacks like modification, impersonation, replay, and man-in-the-middle attack. Due to the system model and employed encryption in our scheme, the decryption of aggregated data does not depend on the presence of all the SMs in the residential area, which makes our scheme fault-tolerant. Especially, not only our DA scheme sustains intact by decreasing the number of users, but also can support an acceptable increase in the number of users. Therefore, the proposed scheme allows dynamic users.

The outline of this article is structured as follows. In Sect. 2, we demonstrate the network model, attack scenarios and establish security requirements and the scheme's objectives. Section 3 introduces the fundamental notations and assumptions of our homomorphic encryption scheme. The explication of the LPM2DA scheme is in Sect. 4. We assess the security and performance of the proposed scheme separately in Sects. 5 and 6 and then draw the article's conclusion and present the future work within Sect. 7.

2 System model

In this section, we describe the network model and its components thoroughly, then establish our aim and the security requirements.

2.1 Network model

Typically, the smart grid communication network consists of one managing unit that administers the communication and information flow, a large number of users who mainly consume energy, and an authority that monitors all entities. In this paper, we name the managing unit, control center (CC). In addition, inside each building area network (BAN), we categorize users in the collection of home area networks (HAN) which could be units or apartments equipped with smart meters to gather energy consumption data. And lastly, TA signifies the trusted authority. As shown in Fig. 1, our system model is comprised of one control center, one trusted authority, and a building area network composed of ω home area networks.

2.1.1 CC

The CC's role could be played by an organization or government. Its responsibility is mainly administering the load balance, collecting consumption information, and evaluating different functions on these data. It also settles down any wrongdoings and even produces bills.

2.1.2 BAN

The BAN is a powerful entity whose gateway (GW) acts as a communication relay between CC and smart meters. Precisely, the GW primarily executes statistical functions on the legitimate data so that CC obtains various aggregations. The communication between GW and smart meters is by means of somewhat inexpensive WiFi technology, whereas the GW communicates with the CC via high bandwidth or wired links.

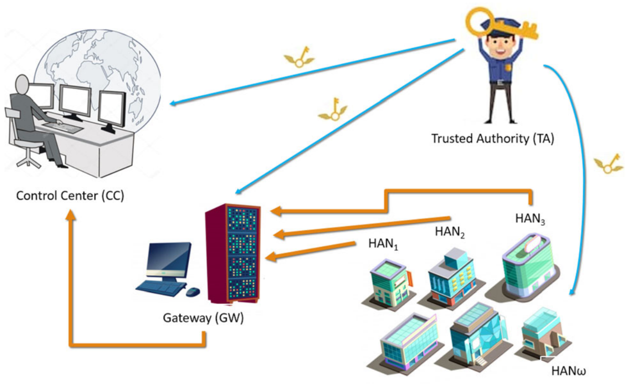

Fig. 1 System model

2.1.3 HAN

Each HAN could be units or apartments equipped with smart meters to gather real-time information of the energy consumed by the appliances in the apartment. At some intervals, like every 15 or 30 min, each smart meter has a unique ID and records the activity and energy usage of all the appliances. Then, it sends these consumption data to CC through the BAN gateway.

2.1.4 TA

If we take CC as the brain of the smart grid network, TA is representative of this network's eyes. The TA is a trustworthy entity that is responsible for the whole system setup and key distributions.

2.2 Assumptions and security requirements

To consider the real-world status, we identify CC and GW as honest-but-curious i.e., they play their role in the system honestly, but they are inquisitive entities and will take any chances they have got to find out about comprehensive consumption data of each customer. As opposed to traditional meters, it is really more complicated for users to tamper with smart meters, so we consider the SM to be an honest entity.

Concretely, there exists a strong adversary $\mathcal{A}$ residing in the smart grid network, which can damage the smart meters and interfere with the databases of CC and GW. The adversary mostly eavesdrops and alters the communication flow or even sets up some undetectable malware in GW to get a hold of consumption data and exploits it to its advantage. More seriously, to disclose privacy and violate integrity and authentication of the network, not only adversary could be passive and just eavesdrop on the communication, but also could carry out some active attacks such as impersonation, modification, man-in-the-middle (MITM), and replay attack.

Therefore, to transmit data in a privacy-preserving manner, ensuring the integrity of the information, authenticating the source, and impeding malicious deeds, we should gratify the security requirements established beneath:

Privacy preservation: In the smart grid communication, neither smart grid components (GW, CC) nor the adversary needs to know the comprehensive consumption data. Therefore, the eavesdropping adversary who deployed malicious malware or even an unhappy employee in the control center should not be capable of unveiling the consumption information. Accordingly, knowing only the residential area's data will suffice CC to balance loads and monitor the grid.

Integrity and authentication: Since the transmission of data is through public links, each entity in the smart grid network should be uniquely identified in order to participate in the communication. Therefore, to hinder adversaries from forging or altering the transmitted data, every entity's authenticity and validity of its data should be checked after each transaction.

Resistance against attacks: Typically, like any other open network, the smart grid is susceptible to numerous attacks, for instance: replay attack, impersonation, modification, differential, MITM, and other internal attacks.

2.3 Design goal

In this paper, we aim to preserve privacy and guarantee data integrity and source authentication meanwhile empowering CC to scrutinize residential area's data thoroughly through computing different kinds of aggregation. Furthermore, based on the utilization of lattice-based cryptography, our data aggregation scheme is supposed to be resistant to imminent quantum attacks. Concretely, the addressed security requirements and calculation of the statistical functions should be accomplished efficiently. Albeit, considering the computation overhead is indispensable to ensure the availability of the smart grid network (providing CC with real-time data every 15 min), the communication costs should also be minimized.

3 Preliminaries

3.1 Notation

For a sufficiently large prime modulus q, we define the ring $\mathbb{Z}/q\mathbb{Z}$ (denoted as $\mathbb{Z}_q$) in $(-q/2, q/2) \cap \mathbb{Z}$. We consider the quotient ring $R_q = R/qR$ based on the prime modulus q and a cyclotomic ring R which is defined next. The ring R is a polynomial ring of the form $R = \mathbb{Z}[x]/ < f_m(x) >$ where $f_m(x) = x^n + 1$ is the irreducible mth cyclotomic polynomial; n is a power of 2; and $m = 2n$. We identify the error distribution χ as the discrete Gaussian distribution $D_{\mathbb{Z}^n,r}$, in which every vector drawn from $D_{\mathbb{Z}^n,r}$ with standard deviation $r > 0$ is of length $r\sqrt{n}$ [24].

The symbol "$\cdot$" specifies all types of multiplications such as matrix and polynomial multiplications. The notation like $A_{M\times N}$ are used to represent matrices of rings with dimensions M and N, where all elements are from R_q; and the identity matrix of size $N \times N$ is denoted by $I_{N\times N}$. For simplicity, we denote the row vectors of length n as $[a_1, \ldots, a_n]$, and the column vectors of length n as $[a_1; \ldots; a_n]$.

In the encryption system [25], deployed in our scheme, there exists one function named "bit decomposition" denoted as BD(.), which can take integers, polynomials, or matrix of polynomials as input and yields the extended version of them. Specifically, BD(.) function outputs a size-ℓ vector consisting of the bit decomposition of the input elements. Therefore, it outputs an $\ell \times n$ matrix for the degree-n polynomial with ℓ-bit integer coefficients as input. Subsequently, by substituting the bit representation of each integer coefficient, BD(.) function produces a matrix of size-$x \times y\ell$ with depth n for the input matrix of polynomials with dimension $x \times y$ (wherein each polynomial is of degree n with ℓ-bit integer coefficients). The inverse of the BD(.) function, which is a function denoted by BDI(.), is also used in the encryption part of our scheme. Technically, BDI(.) function collects ℓ successive coefficients and yields the ℓ-bit integer of the BD(.)'s input. Precisely, BDI(.) function can be identified as the multiplication of its input matrix and a matrix named $\Upsilon_{y\ell\times y}$ as described below [25]:

$$B_{x\times y} = \text{ BDI } (\tilde{B}_{x\times y\ell}) = \tilde{B}_{x\times y\ell} \cdot \Upsilon_{y\ell\times y}.$$

$$\Upsilon_{y\ell\times y} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 2 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 2^{\ell} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 2^{\ell} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2^{\ell} \end{bmatrix}. \tag{1}$$

3.2 Assumption

Ajtai's breakthrough paper [26] which showed a connection between the worst-case and the average-case problems for lattices, affords confidence in adopting lattice-based schemes in cryptography which holds a grand promise for post-quantum cryptography. In 2005, with the seminal work of Regev [27], the average-case problem of learning with errors was announced. Since then, the LWE and its variants seemed to be a versatile problem in most encryption schemes. In this paper, we utilize a homomorphic scheme named SHIELD [25] which is based on the ring variant of the LWE problem introduced by Lyubashevsky et al. [28]. The hardness of the R-LWE problem is linked to the worst-case problems on ideal lattices. For a uniformly random $s \in R_q$ (called the secret key), there are two distributions on $R_q \times R_q$: the first distribution is $(a, b = a \cdot s + e) \in R_q \times R_q$ where a is uniform, randomly chosen from R_q and e is an independent error chosen from the error distribution $\chi \subset R_q$; and the second distribution is formed by choosing $(a, b) \in R_q \times R_q$ uniformly. The decision R-LWE problem is distinguishing between these two distributions with non-negligible advantage.

4 The proposed scheme: LPM2DA

The LPM2DA scheme is formed by five phases, named: system setup, user report generation, report aggregation, secure report reading, and temporal aggregation. First, we present an overview of the proposed DA scheme. The LPM2DA scheme mainly focuses on computational competence, such as computing multiple statistical functions, different aggregations, and calculation proficiency.

In the beginning, TA initializes the system via settling both system and Chinese remainder theorem's (CRT) parameters [29]. Then, it generates and issues different key pairs of all entities. One of the efficient points of this scheme is its ability to handle multi-dimensional data by exploiting polynomial CRT. Technically, based on polynomial CRT, each smart meter gathers all of its multi-dimensional data into a single appropriate data and then encrypts that data. Subsequently, each smart meter signs its encrypted data to achieve integrity, authentication, and securing the scheme against various attacks.

Despite the fact that the GW is merely a relay between the residential area's smart meters and the CC, it verifies the signatures, computes different kinds of aggregations on the valid data, and also computes the bill at the end of each billing period. Finally, after verifying the aggregated data, which is signed by GW, CC decrypts the aggregated data. Due to the deployed homomorphic encryption scheme and GW's collaboration, CC could obtain any kind of aggregation it prefers. Specifically, relying on the CRT's features, CC could acquire various aggregations of each dimension's data without disclosing privacy. Now, the comprehensive description of each phase of our scheme is presented below.

4.1 System setup

In this phase, TA plays the leading role in the configuration of the system parameters and generates the keys for signing and encryption. First, TA yields system parameters including the maximum degree of polynomials n; the prime modulus $q = 1(\text{mod } 2\text{n})$; the standard deviation of discrete Gaussian distributions of keyspace and ciphertext space from which the key and the ciphertext error are taken from and denoted by σ_k and σ_c, respectively; $\ell = \lceil \log q \rceil$, and $N = 2\ell$ is a parameter used in the ciphertext matrix dimensions and manages the ciphertext size. The entire system parameters and the complexity of the operations evaluated on ciphertexts are managed by λ, which is the security parameter and typically is set to be 80 or 120 bits.

Then, to utilize the CRT technique [29] for transmitting k-dimensional data, TA chooses k pairwise coprime polynomials in the ring R, denoted by $p_j(x)$ for $j = 1, 2, \ldots, k$. We denote the degree of each $p_j(x)$ with $\deg(p_j)$ and set $D = \sum_{j=1}^{k} \deg(p_j)$. These parameters are chosen such that for each smart meters' data polynomial like $p_{data}(x)$, $\deg(p_{data}) < \deg(p_j)$. TA denotes the multiplication of all of these k polynomials with $Q(x) = \prod_{j=1}^{k} p_j(x)$ and defines $Q_j(x) = Q(x)/p_j(x)$ for $j = 1, 2, \ldots, k$. Then, it finds k polynomials $T_j(x)$ with $\deg(T_j(x)) < \deg(p_j(x))$ by the partial fraction decomposition of $1/Q(x) = \sum_{j=1}^{k} T_j(x)/p_j(x)$.

Based on the notations and the system parameters, to build CC's public and private keys, TA chooses the polynomial t_{cc} from the discrete Gaussian error distribution D_{R_q,σ_k}. Subsequently, the CC's secret key is set to be the column vector of polynomials $SK_{cc} = [1; -t_{cc}]$. Then, it samples $a_{cc} \leftarrow R_q$, $e_{cc} \leftarrow D_{R_q,\sigma_k}$ uniformly and yields $b_{cc} = a_{cc} \cdot t_{cc} + e_{cc}$. Hence, CC's public key is the row vector of polynomials $PK_{cc} = [b_{cc}\, a_{cc}]$. From the definition of b_{cc}, we can see that the inner product of public and private keys over the ring R_q is:

$$PK_{cc} \cdot SK_{cc} = [b_{cc}\, a_{cc}] \cdot [1; -t_{cc}] = b_{cc} - a_{cc} \cdot t_{cc} = e_{cc}. \tag{2}$$

Next, TA samples GW's secret key s_{GW} from R_q. Given a prime integer $t_{GW} \in \mathbb{Z}_q{}^*$ and drawing $a_{GW} \leftarrow R_q$ and $e_{GW} \leftarrow \chi$, where $\chi = D_{\mathbb{Z}^n,r}$ is the error distribution, TA sets $b_{GW} = a_{GW} \cdot s_{GW} + t_{GW} \cdot e_{GW}$ and lets the GW's public key to be $PK_{GW} = (a_{GW}, b_{GW}) \in R_q \times R_q$.

Similarly, to generate keys for each smart meter, TA draws a ring element $s_i \leftarrow R_q$ for $i = 1, 2, \ldots, \omega$ and $SK_i = s_i$. By computing $b_i = a_i \cdot s_i + t_i \cdot e_i \in R_q$ where $t_i \in \mathbb{Z}_q{}^*$, a_i is drawn from R_q and e_i is chosen from the error distribution $\chi = D_{\mathbb{Z}^n,r}$, TA assigns the SM_i's public key as $PK_i = (a_i, b_i) \in R_q \times R_q$ for $i = 1, 2, \ldots, \omega$.

Finally, TA chooses a secure collision-resistant hash function $H : \{0,1\}^* \rightarrow R_q$, which will be utilized in the signing procedure. In the end, TA issues the key pairs to each entity through a secure channel and publishes the public parameters.

4.2 User report generation

At every time instants T_γ, for example every 15 or 30 minutes, each smart meter SM_i for $i = 1, 2, \ldots, \omega$ uses the polynomial CRT to pack its k-dimensional data ($d_{i1}, d_{i2}, \ldots, d_{ik}$ for $i = 1, 2, \ldots, \omega$) into one message and encrypts it with CC's public key. Specifically, each SM_i in residential area performs the following procedure:

1. SM_i multiplies each one of its data by the polynomials $Q_j(x)$ and $T_j(x)$, and computes the summation of them to attain $m_i(x) \in R_q$:

$$\begin{aligned} m_i(x) &= d_{i1}(x) \cdot T_1(x)Q_1(x) + d_{i2}(x) \cdot T_2(x)Q_2(x) \\ &\quad + \cdots + d_{ik}(x) \cdot T_k(x)Q_k(x) \\ &= \sum_{j=1}^{k} d_{ij} \cdot T_j(x)Q_j(x) \in R_q. \end{aligned} \tag{3}$$

2. At the interval T_γ, to encrypt the polynomial message $m_i \in R_q$ with CC's public key, SM_i utilizes an error matrix $E_{N\times 2} \leftarrow D_{R_q^{N\times 2},\sigma_c}$ and an $N \times 1$ matrix of polynomials in which every random coefficient is in $\{0,1\}$. Then, it can obtain the $N \times 2$ matrix of ciphertext, as shown below:

$$C_{i\gamma} = m_i \cdot \text{ BDI } (I_{N\times N}) + r_{N\times 1} \cdot PK_{cc} + E_{N\times 2}. \tag{4}$$

To achieve data integrity and source authentication, every smart meter SM_i needs to sign the ciphertext with its own secret key. Specially, SM_i for $i = 1, 2, \ldots, \omega$ hashes the ciphertext C_i (for simplicity we use the notation C_i instead of $C_{i\gamma}$) and the timestamp T_γ, and then creates the signature as presented below:

$$u_i = \big(v_i + H(C_i, T_\gamma)\big) \cdot s_i + t_i \cdot e'_i, \sigma_i = (u_i, v_i), \tag{5}$$

where v_i is drawn from a uniform distribution over R_q and the error term $e'_i \leftarrow \chi$ is different from e_i which was sampled in the system setup phase. Finally, SM_i sends the ciphertext C_i, timestamp T_γ, and the signature $\sigma_i = (u_i, v_i)$ to the GW.

4.3 Report aggregation

The GW checks the integrity and authenticity of the data received by verifying the signature and the timestamp. Precisely, GW verifies the signature by assessing the following conditions for every $SM_i, i = 1, 2, \ldots, \omega$:

1) $\sigma_i \in R_q \times R_q$,
2) $[-a_i \cdot u_i + b_i \cdot v_i] \bmod t_i = -b_i \cdot H(C_i, T_\gamma) \bmod t_i$.

If both of the mentioned conditions hold, the GW aggregates ciphertexts in a privacy-preserving manner. The verification correctness is demonstrated at the end of this section. In the proposed scheme, the utilization of the homomorphic encryption scheme lets CC compute multiple functions with various circuit depth[1], such as mean, variance, skewness, and one-way analysis of variance (ANOVA), etc. Nevertheless, here we just demonstrate the aggregation needed for the computation of mean, variance, and skewness. Surely, GW accomplishes different aggregations as steps shown below:

1. GW computes the homomorphic addition of all the verified ciphertexts:

$$C_{Add} = \sum_{i=1}^{\omega} C_i. \tag{6}$$

2. Then, GW calculates the homomorphic multiplication of each ciphertext with itself. Moreover, it aggregates the result:

$$\begin{aligned} A_i &= \text{ BD } (C_i) \cdot C_i \text{ for } i = 1, 2, \ldots, \omega, \\ C_{Mult} &= \sum_{i=1}^{\omega} A_i. \end{aligned} \tag{7}$$

3. For the last aggregation, GW calculates the homomorphic multiplication of each C_i with A_i obtained from the Eq. (7), and then computes the summation of them:

$$\begin{aligned} B_i &= \text{ BD } (A_i) \cdot C_i \text{ for } i = 1, 2, \ldots, \omega, \\ C_{Skew} &= \sum_{i=1}^{\omega} B_i. \end{aligned} \tag{8}$$

Now, to provide integrity and authentication, GW signs the timestamp and the aggregations. Specifically, after calculating the hash value of $T_\gamma, C_{Add}, C_{Mult}$, and C_{Skew}, GW achieves $\sigma_{GW} = (u_{GW}, v_{GW})$ by evaluating $u_{GW} = (v_{GW} + H(C_{Add}, C_{Mult}, C_{Skew}, T_\gamma)) \cdot s_{GW} + t_{GW} \cdot e'_{GW}$ where $v_{GW} \in R_q$ and the error term e'_{GW} is different from e_{GW} in the system setup phase, and drawn from the error distribution χ. Finally, GW transmits different aggregations $C_{Add}, C_{Mult}, C_{Skew}$, timestamp T_γ, and $\sigma_{GW} = (u_{GW}, v_{GW})$ to the CC.

[1] The circuit depth of a function is the number of multiplication levels required for the implementation of that function.

4.3.1 Verification correctness

The first condition's correctness is apparent, and the second condition's correctness is presented below. We first analyze the left-hand side of the second condition's equation:

$$\begin{aligned}
&[-a_i \cdot u_i + b_i \cdot v_i] \bmod t_i \\
&= \left[-a_i\left(\left(v_i + H(C_i, T_\gamma)\right)s_i + t_i e'_i\right) + (a_i s_i + t_i e_i)v_i\right] \bmod t_i \\
&= [-a_i v_i s_i - a_i H(C_i, T_\gamma)s_i - a_i t_i e'_i + a_i s_i v_i + t_i e_i v_i] \bmod t_i \\
&= -a_i \cdot H(C_i, T_\gamma) \cdot s_i.
\end{aligned} \tag{9}$$

Since the right-hand side of the second condition's equation is simplified as below, it is obvious that both sides of the equation are equal, and the verification is correct.

$$\begin{aligned}
&[-b_i \cdot H(C_i, T_\gamma) \cdot s_i] \bmod t_i \\
&= [-(a_i \cdot s_i + t_i \cdot e_i) \cdot H(C_i, T_\gamma)] \bmod t_i \\
&= -a_i \cdot s_i \cdot H(C_i, T_\gamma).
\end{aligned} \tag{10}$$

4.4 Secure report reading

When CC receives data from GW, it first checks the validity of the timestamp and signatures. Indeed, CC examines the equations $\sigma_{GW} \in R_q \times R_q$ and $[-a_{GW} \cdot u_{GW} + b_{GW} \cdot v_{GW}] \bmod t_{GW} = -b_{GW} \cdot H(C_{Add}, C_{Mult}, C_{Skew}, T_\gamma) \bmod t_{GW}$, and if they hold, CC could decrypt the legitimate data to obtain different aggregations of user's data (in each dimension). Surely, by decrypting each of the ciphertexts C_{Add}, C_{Mult}, and C_{Skew}, CC will acquire $\sum_{i=1}^{\omega} m_i$, $\sum_{i=1}^{\omega} m_i^2$, and $\sum_{i=1}^{\omega} m_i^3$, respectively. For decryption, CC just needs to multiply the ciphertext by its private key SK_{cc}. The decryption and homomorphic correctness are demonstrated at the end of this section.

Next, CC acquires the k-dimensional data, by applying CRT techniques. These aggregations are exhibited in the form:

$$\sum_{i=1}^{\omega} m_i = \sum_{i=1}^{\omega} \left(\sum_{j=1}^{k} d_{ij}(x) T_j(x) Q_j(x) \right), \tag{11}$$

$$\sum_{i=1}^{\omega} m_i^2 = \sum_{i=1}^{\omega} \left(\sum_{j=1}^{k} d_{ij}(x) T_j(x) Q_j(x) \right)^2, \tag{12}$$

$$\sum_{i=1}^{\omega} m_i^3 = \sum_{i=1}^{\omega} \left(\sum_{j=1}^{k} d_{ij}(x) T_j(x) Q_j(x) \right)^3. \tag{13}$$

Based on CRT's features, the remainder of the Euclidean division of $\sum_{i=1}^{\omega} m_i, \sum_{i=1}^{\omega} m_i^2$, and $\sum_{i=1}^{\omega} m_i^3$, by $p_j(x)$ is $\sum_{i=1}^{\omega} d_{ij}(x), \sum_{i=1}^{\omega} d_{ij}^2(x)$, and $\sum_{i=1}^{\omega} d_{ij}^3(x)$, respectively. The correctness of CRT features used in above statement is demonstrated at the end of this section. Eventually, CC could compute various functions of the aggregated k-dimensional data for $j = 1, 2, \ldots, k$, as follows:

$$\text{Mean } = \frac{1}{\omega} \sum_{i=1}^{\omega} d_{ij}, \tag{14}$$

$$\text{Variance } = \frac{1}{\omega} \left(\sum_{i=1}^{\omega} d_{ij}^2\right) - \text{ Mean}^2, \tag{15}$$

$$\text{Skewness } = \frac{\frac{1}{\omega}\left(\sum_{i=1}^{\omega} d_{ij}^3\right) - 3\,\text{Mean} \cdot \text{Variance} - \text{Mean}^3}{\text{Variance}^{3/2}}. \tag{16}$$

Based on the computation of the mean and the variance of the users' multi-dimensional data, the CC could realize electricity theft, power leakage, and the uniformity of the usage data. Since the skewness measures the asymmetry of the probability distribution, the computation of the skewness on the users' data helps the CC to know the differences between sample distribution and normal distribution with the users' electricity usage data distribution. Hence, it helps him to balance the grid electricity more thoroughly.

4.4.1 Decryption correctness

Consider the ciphertext C_i as a representative of m_i's encryption, then we have:

$$\begin{aligned}
C_i \cdot SK_{cc} &= \big(m_i \cdot \text{ BDI } (I_{N\times N}) + r_{N\times 1} \cdot PK_{cc} \\
&\quad + E_{N\times 2}\big) \cdot SK_{cc} \\
&= m_i \cdot \text{ BDI } (I_{N\times N}) \cdot SK_{cc} + r_{N\times 1} \cdot PK_{cc} \cdot SK_{cc} \\
&\quad + E_{N\times 2} \cdot SK_{cc} \\
&= m_i \cdot \text{ BDI } (I_{N\times N}) \cdot SK_{cc} + r_{N\times 1} \cdot e_{cc} + E_{N\times 2} \cdot SK_{cc} \\
&= m_i \cdot \text{ BDI } (I_{N\times N}) \cdot SK_{cc} + error.
\end{aligned} \tag{17}$$

Due to errors e_{cc} and $E_{N\times 2}$, we have $error = r_{N\times 1} \cdot e_{cc} + E_{N\times 2} \cdot SK_{cc}$. Now, the first ℓ coefficients in (17) are of the form $m_i, 2m_i, \ldots, 2^{\ell-1}m_i$ in addition to $error$, as shown below:

$$\begin{aligned}
&m_i \cdot \text{ BDI } (I_{N\times N}) \cdot [1\,; -t_{cc}] \\
&= m_i \cdot \begin{bmatrix} 1 & 2 & \ldots & 2^{\ell-1} & -t_{cc} & -2t_{cc} & \ldots & -2^{\ell-1}t_{cc} \end{bmatrix}^T.
\end{aligned} \tag{18}$$

So, considering n-degree polynomial, for each coefficient of this polynomial and with the assumption that $error < q/2$, the most significant bit of each part of the above matrix has one bit of m_i.

4.4.2 Homomorphic correctness

Considering the above correctness, the legitimacy of homomorphic addition is evident. Nevertheless, for the homomorphic multiplication, the correctness is slightly tricky. Assuming C_i is the encryption of m_i, the homomorphic multiplication is asymmetric based on the input; and as a result, the noise growth is lower. Matrix dimensions are omitted for simplicity.

$$\begin{aligned}
&\text{BD }(C_i)\cdot C_i\cdot SK_{cc}=\text{ BD }(C_i)\cdot\big(m_i\cdot\text{ BDI }(I)+\\
&r\cdot PK_{cc}+E\big)\cdot SK_{cc}\\
&=\text{ BD }(C_i)\cdot\big(m_i\cdot\text{ BDI }(I)\cdot SK_{cc}+r\cdot e_{cc}+E\cdot SK_{cc}\big)\\
&=m_i\cdot\text{ BD }(C_i)\cdot\text{ BDI }(I)\cdot SK_{cc}+\text{ BD }(C_i)\big(r\cdot e_{cc}+E\cdot SK_{cc}\big)\\
&=m_i\cdot\text{ BD }(C_i)\cdot\text{ BDI }(I)\cdot SK_{cc}+error_1\\
&\overset{(a)}{=}m_i\cdot C_i\cdot SK_{cc}+error_1\\
&=m_i\cdot\big(m_i\cdot\text{ BDI }(I)+r\cdot PK_{cc}+E\big)\cdot SK_{cc}+error_1\\
&=m_i^2\cdot\text{ BDI }(I)\cdot SK_{cc}+m_i\cdot r\cdot PK_{cc}\cdot SK_{cc}\\
&+m_i\cdot E\cdot SK_{cc}+error_1\\
&=m_i^2\cdot\text{ BDI }(I)\cdot SK_{cc}+m_i\cdot\big(r\cdot e_{cc}+E\cdot SK_{cc}\big)+error_1\\
&=m_i^2\cdot\text{ BDI }(I)\cdot SK_{cc}+error_2
\end{aligned}\tag{19}$$

where (a) is due to the fact that $\text{BD}(C_i)\cdot\text{ BDI }(I_{N\times N})=I_{N\times N}\cdot C_{N\times 2}=C_i$. Based on the errors e_{cc} and E, we consider $error_1=\text{ BD }(C_i)\big(r\cdot e_{cc}+E\cdot SK_{cc}\big)$ and $error_2=m_i\cdot\big(r\cdot e_{cc}+E\cdot SK_{cc}\big)+error_1$. It should be observed that the decryption of $\text{BD}(C_i)\cdot C_i$ is of the form m_i^2. It can be noted that through an accumulator-like procedure, we can multiply ciphertexts efficiently [25]. The homomorphic correctness of depth-3 multiplication can be done similarly.

4.4.3 CRT correctness

The Chinese remainder theorem for polynomials states that for the modulo $p_j(x)$, for $j=1,2,\ldots,k$, with degree $deg(p_j)$, $D=\sum_{j=1}^{k}deg(p_j)$ and polynomials $d_{ij}(x)$ where $d_{ij}(x)=0$ or $deg(d_{ij}(x))<deg(p_j(x))$, there is a unique polynomials $P(x)$ with $deg(P(x))<D$ for which the following equation holds [29]:

$$P(x)=d_{ij}(x)\ \big(\bmod p_j(x)\big)\ \text{ for } j=1,2,\ldots,k.\tag{20}$$

Using the parameters determined in the system setup phase like $Q(x)=\prod_{j=1}^{k}p_j(x), Q_j(x)=Q(x)/p_j(x)$, and by substituting $p_j(x)=Q(x)/Q_j(x)$ in $1/Q(x)=\sum_{j=1}^{k}T_j(x)/p_j(x)$, we have:

$$\sum_{j=1}^{k}T_j(x)Q_j(x)=1.\tag{21}$$

Now, by considering $d_1,d_2,...,d_k$ as the k-dimensional data, and based on the above analysis, we have:

$$\begin{aligned}
&\sum_{j=1}^{k}d_j(x)T_j(x)Q_j(x)=\\
&d_j(x)+\sum_{J=1}^{k}\big(d_J(x)-d_j(x)\big)T_J(x)Q_J(x)\\
&\equiv d_j(x)\ \big(\bmod p_j(x)\big),\quad \text{ for } j=1,2,\ldots,k.
\end{aligned}\tag{22}$$

Therefore, to compute each dimension of SM_i's aggregated data, CC could perform as follows:

$$\begin{aligned}
\sum_{i=1}^{\omega}m_i&=\sum_{i=1}^{\omega}\Big(\sum_{j=1}^{k}d_{ij}(x)\,T_j(x)\,Q_j(x)\Big)\\
&=\sum_{j=1}^{k}\Big(\big(\sum_{i=1}^{\omega}d_{ij}(x)\big)\,T_j(x)\,Q_j(x)\Big)\\
&=\Big(\sum_{i=1}^{\omega}d_{i1}(x)\Big)T_1(x)Q_1(x)+\Big(\sum_{i=1}^{\omega}d_{i2}(x)\Big)T_2(x)\\
&\quad Q_2(x)+\cdots+\Big(\sum_{i=1}^{\omega}d_{ik}(x)\Big)T_k(x)Q_k(x).
\end{aligned}\tag{23}$$

Then, CC could acquire the aggregation of each dimension with the following congruences:

$$\sum_{i=1}^{\omega}m_i\equiv\sum_{i=1}^{\omega}d_{ij}(x)\ \big(\bmod p_j(x)\big),\ \text{ for } j=1,2,\ldots,k.\tag{24}$$

Similarly, the CC could compute the congruences for the $\sum_{i=1}^{\omega}m_i^2$, $\sum_{i=1}^{\omega}m_i^3$.

4.5 Temporal aggregation

The purpose of this phase is to compute an imperative aggregation named temporal aggregation. As opposed to spatial aggregation computed in the prior phases to accomplish privacy of users' real-time data, the temporal aggregation will be calculated regularly (every billing period) to obtain the bill. Given the timestamp T_γ for $\gamma=1,2,\ldots,b,b+1,\ldots,2b,\ldots,3b,\ldots$; at the end of each billing period, GW computes the temporal aggregation of each smart meter's data. Let's assume that we are at the end of the first bill period T_b. The GW aggregates the SM_i's data relying on the data stored in its database connected to the SM_i's identity (ID_i), and then signs the aggregated data as shown below:

$$C_{Bi} = \sum_{\gamma=1}^{b} C_{i\gamma}, \tag{25}$$

$$v_{GW} \leftarrow R_q\ ,\ u_{Gw} = \big(v_{GW} + H(C_{Bi}, ID_i, T_b)\big) \cdot s_{GW} + t_{GW} \cdot e'_{GW}. \tag{26}$$

Lastly, GW sends C_{Bi}, ID_i, T_b, and $\sigma_{GW} = (u_{GW}, v_{GW})$ to CC. After verifying the timestamp and signature by checking the validity of two conditions:

$$1)\ \sigma_{GW} \in R_q \times R_q \tag{27}$$

$$2)\ [-a_{GW} \cdot u_{GW} + b_{GW} \cdot v_{GW}] \bmod t_{GW} = -b_{GW} \cdot H(C_{Bi}, ID_i, T_b) \bmod t_{GW} \tag{28}$$

then, CC decrypts the temporal aggregation in a similar way and obtains the bill.

4.6 Structure of the proposed CRT-based technique

In this subsection, we aim to analyze various aspects of the proposed CRT-based technique more thoroughly. Since the lattice-based encryption scheme has a large plaintext space and the smart grid communications are usually very small, it would not be efficient to send a single data via each transmission. One way to make a scheme more efficient is to transmit multi-dimensional data in each transmission instead of sending single data. Moreover, transmitting multi-dimensional data is considerably practical in the smart grid network because the consumption data is composed of various information about multiple appliances of each house and the heating and lighting systems' data. It should be noted that based on the same reasons, other applications and constructions besides the smart grid may also require transmitting multi-dimensional data. As mentioned before, in our scheme, we have used the Chinese remainder theorem to handle multi-dimensional data transmission. Technically, based on polynomial CRT, each smart meter gathers all of its multi-dimensional data into a single appropriate data and then encrypts that data. Practically, our technique requires ℓ multiplications and $\ell - 1$ additions to transmit ℓ-dimensional data.

In contrast to the other techniques, which are specialized for one specific encryption scheme, the proposed technique is independent of the encryption scheme and could be utilized with various encryptions and different structures. More specifically, our technique only requires to be adapted for the plaintext space of the employed encryption scheme; and that is achieved through choosing the proper construction of CRT. For example, in order to use number-theoretic encryptions with integer plaintext or lattice-based encryptions with polynomial plaintext, providing an appropriate selection of CRT's parameters, we could utilize an integer CRT or polynomial CRT, respectively. Besides, it also supports homomorphic encryptions and different depths of computations on the encrypted data.

Typically, to transmit multi-dimensional data, most schemes use super-increasing sequence, Horner's rule, or design an algorithm to fill the multi-dimensional data into a blank plaintext. First, in comparison to our CRT-based technique, these techniques are adopted for a specific type of encryption scheme, and thereby they are practically not applicable in our case. Furthermore, these techniques impose heavy computations on the input multi-dimensional data; and to obtain each component of data after decryption, they require a recursive algorithm. Hence, they are not computationally efficient. Finally, given the security requirements and space restrictions, the number of dimensions that these techniques could support is quite limited. Thus, we can conclude that the proposed CRT-based technique handles the multi-dimensional data transmission more efficient.

5 Security analysis

This section comprises the analysis of three parts: privacy preservation, integrity and authentication, and resistance against other attacks. Concretely, as mentioned in the adversary model, assume there exists a strong adversary $\mathcal{A}$ who aims to obtain the consumption data through eavesdropping, using malware, and attacking entities. Therefore, we focus on the privacy-preserving assessment because it is of paramount importance for the smart grid network. Besides, we demonstrate that the proposed scheme can guarantee the integrity of data, the authenticity of sources, and resist attacks carried out by the adversary.

5.1 Privacy preservation

There exist three components in the smart grid network that communicate the consumption reports through two links: SM-to-GW and GW-to-CC. Since every report is transmitted through public channels, the strong adversary $\mathcal{A}$ inhabits the customer-side of the network to eavesdrop on the energy consumption data. However, the energy consumption data is a really small value, and the adversary might try a brute-force attack. Since our utilized homomorphic encryption scheme [25] is IND-CPA secure (i.e., indistinguishability under chosen-plaintext attack), and every message is encrypted, no adversary is capable of unveiling smart meters or aggregated data without the knowledge of CC's private key.

Let's imagine that the honest but curious entity, GW, desires to acquire the energy consumption data. Since GW is simply a relay and its principal obligation is to gather encrypted data and aggregate them homomorphically, GW cannot get any individual's data. Thus, the role of the GW as a relay will not disclose any private information about the transmitted data. Hence, any computationally strong entity can play GW's role.

Envision the brain of the smart grid, CC, wants to obtain the individual reports. The CC cannot attain any private information about the households because we have designed the system model to be a data aggregation scheme and only aggregated data is received by the CC. Therefore, our system structure permits CC to achieve the bill and the aggregated data as it needs.

Now, consider that the strong adversary could utilize malicious malware in the CC or GW, or even they are potent enough to interfere with GW or CC's database. Evidently, we can infer from the above assessments, this adversary could not get any individual information because the GW and CC themselves could not acquire the individual's energy consumption information. As a result of deploying a homomorphic encryption scheme which is IND-CPA secure and the fact that the CC's private key is distributed by TA over a secure channel, the adversary's attempts to achieve private information about individual users will fail, or at best, gives them aggregated reports' statistics. Hence, the LPM2DA scheme preserves the confidentiality of users' information.

5.2 Integrity and authentication

In our proposed scheme, to achieve integrity of communicated data, each entity must sign its reporting data. Especially, by relying on a collision resistance hash function and the unforgeability of the R-LWE based signature scheme [24] under chosen message attack, the integrity of each transmitted data is assured. Recall that only TA allocates private keys via a secure channel, and there is a straightforward equivalence between discovering private keys from transmitted data and solving the R-LWE problem. Thus, since the users report the signatures alongside their data (C_i, σ_i), send the signature of aggregated data by GW (C_{GW}, σ_{GW}) and report the IDs in the billing period, no adversary can modify the reports without the private keys. Since the first step in every phase is to authorize each entity and its data, and the invalid data would be discarded, the adversary cannot tamper with the aggregated data and interrupt the system. Consequently, there is no way in which the adversary could forge the signature or modify the data without detection.

Since each entity has its public and private keys assigned by TA over a secure channel, it would be reasonably hard for the adversary to impersonate an entity. As a result of sending every report with its signature, each entity will be authenticated before any further steps. Accordingly, data integrity and source authentication of users' data is provided.

5.3 Resistance against other attacks

In addition to the criteria addressed above, the proposed scheme is resistant to various attacks and supports some practical properties. As mentioned earlier, no adversary could yield a valid signature of his/her reports to engage in fraud or corrupt the residential area's data. Consequently, modification and impersonation attacks can be detected. Similarly, if the adversary resides in the links between SM-to-GW or GW-to-CC, the execution of a MITM attack will be detected by checking the legitimacy of each entity. The timestamp is used in every message and signature; hence the adversary could not carry out a replay attack. Furthermore, whether the adversary comprises some smart meters or even occasionally smart meters do not send any data, the GW still can aggregate the received ciphertexts due to the employed homomorphic encryption scheme. Therefore, the LPM2DA scheme achieves fault tolerance. In the proposed scheme, each GW is generally connected with ω number of users on average. We determined ω in a way that with a few increases or decreases in the number of users, the CC could decrypt the aggregated data appropriately. As a consequence, if a new SM needs to connect to a near GW, it would suffice to register himself to the TA to get his ID and key pairs. Thus, our LPM2DA scheme supports dynamic users. Although the employed homomorphic encryption scheme is secure [25], like any other scheme that uses homomorphic encryptions, our scheme is also susceptible to some collusion attacks. Technically, in all the schemes with homomorphic encryptions, the authority that can decrypt the aggregated data, can also decrypt any individual data. Thereby, the authorities with the decryption key can collude with other network entities to acquire individual data and disclose its privacy. The detection of collusion between CC and GW is our goal in future work.

6 Performance evaluation

In this section, we aim to assess the computational performance and communicational overhead of our LPM2DA scheme. To show the efficiency of the proposed scheme, we compare its performance with schemes like EPPA [9], MuDA [11], FGDA [12], Shen's scheme [30] and lattice-based schemes like PDA [20] and Abdallah's scheme [19]. It should be noted that traditional schemes are not secure against quantum attacks. The lattice-based schemes in [17, 18] are relying on forecasting the electricity demand, and in their system model, a secure database is required,

Table 1 Computational costs 1

Operation	Time (ms)
C_{ez}	12.4
C_m	6.4
C_{et}	8.4
C_{bp}	20
C_{pl}	18.3 (ω =50), 25.8 (ω =100) 31.5 (ω=150) 36.48 (ω=200)

which makes their system model different from DA schemes' model. Therefore, the performance of these schemes is not fairly comparable. Abdallah's scheme utilizes homomorphic encryption based on lattices. However, its system model considers the appliances before the smart meters, which is different from our considered model that is a typical system model in DA schemes. Hence, we assess the efficiency of LPM2DA and Abdallah's scheme in a separate subsection.

6.1 Computation costs

Due to the fact that the cost of the system setup phase is truly a one-time expense, we only focus on the cost of user report generation, report aggregation, and secure report reading phase. To analyze the computational efficiency of the proposed scheme, we compare LPM2DA with various schemes described subsequently. The lattice-based DA scheme, namely PDA [20], uses a somewhat homomorphic encryption and could transmit multi-dimensional data via an algorithm, and it is capable of computing the mean and variance of users' data. The available traditional schemes are as follows: EPPA [9], which uses Paillier cryptosystem and bilinear pairing and could transmit multi-dimensional data via super-increasing sequence; DA schemes like MuDA [11] which uses BGN cryptosystem and bilinear pairing, and could calculate functions like average, variance, and one-way ANOVA on users' data; and FGDA [12] which is also capable of calculating the average, variance, and skewness of users' data; and finally, Shen's scheme [30] which also utilizes Paillier cryptosystem and bilinear pairing, and could transmit multi-dimensional data via Horner's rule. It should be noted that the computational costs are from [7, 14, 21, 23].

Based on PBC [31] and MIRACL [32] libraries running on a 3.0-GHz Pentium IV processor with 512MB memory, and considering $|N|^2 = 2048$ and a 160-bit cyclic group $\mathbb{G}$, the computation costs of exponentiation operation in $\mathbb{Z}_{n^2}$ which is denoted by C_e, group-based multiplication C_m, group-based exponentiation C_{et}, bilinear pairing C_{bp} and Pollard's lambda method C_{pl} are demonstrated in Table 1.

Based on an 1126 MHz GPU with 4 GB memory and a 3.5 GHz core i7 5930K with 15 MB cache size CPU; and using the abbreviations Enc, Dec, Add, Mult for encryption, decryption, addition, and multiplication, respectively, and the notation SH for somewhat homomorphic, the computational costs are denoted in Table 2.

To lessen the computation overhead, the multiplication can be done using the number-theoretic transform (NTT), which is similar to fast Fourier transform (FFT) in terms of computational cost in $O(n \log n)$. Our homomorphic encryption scheme can exploit the parallelism when implemented on a GPU platform, which achieves huge speed up compared to the implementations based on CPU, IBM HLib, and PDA's somewhat homomorphic encryption and other traditional encryption schemes. In the case of using NTT [33, 34], since each dimensional data in smart grid communication is typically small, transmitting k-dimensional data via the Chinese remainder theorem could cost nearly $k \times 30\mu$s which is practically insignificant in contrast to the encryption part of our scheme. According to the structure of each scheme and the costs presented above, each smart meter's cost is presented in Table 3.

It should be noted that in Shen's scheme [30], there exists two gateways: district gateway (DGW) and Residential area gateway (RAGW); and n_i depicts the number of smart meters in the ith residential area, where $n_i < \omega$. The GW's computational cost is demonstrated in the Table 4, and the clear comparison of average and variance aggregation's costs are depicted in Figs. 2 and 3.

For the homomorphic multiplication part, our encryption scheme uses the components of ciphertext C_i and multiplies it by C_i's bit-wise decompositions, which means it is asymmetric based on the input ciphertexts. Thus, due

Table 2 Computational costs 2

LPM2DA Enc (CPU)	< 100 ms
LPM2DA Enc (GPU)	< 10 ms
LPM2DA Dec (CPU)	< 100 ms
LPM2DA Dec (GPU)	< 3 ms
LPM2DA add (CPU)	< 1 ms
LPM2DA add (GPU)	< 0.1 ms
LPM2DA mult (CPU)	< 104 ms
LPM2DA mult (GPU)	< 1.4 ms
LPM2DA sign	0.36 ms
LPM2DA verify	0.57 ms
PDA SHEnc	348 ms
PDA SHDec	26 ms
PDA SHAdd	1 ms
PDA SHMult	41 ms

Table 3 SM's computational cost

	SM
LPM2DA (CPU)	Enc + Sign <100 ms
LPM2DA (GPU)	Enc + Sign <10 ms
PDA [20]	SHEnc = 348 ms
EPPA [9]	$(k+1)C_{ez}+C_m+4C_p = 12.4k + 98.8$ ms
MuDA [11]	$3C_{et}+C_m = 31.4$ ms
FGDA [12]	$C_{et}+2C_m = 21.2$ ms
Shen's scheme [30]	$2C_{ez}+C_m = 46.4$ ms

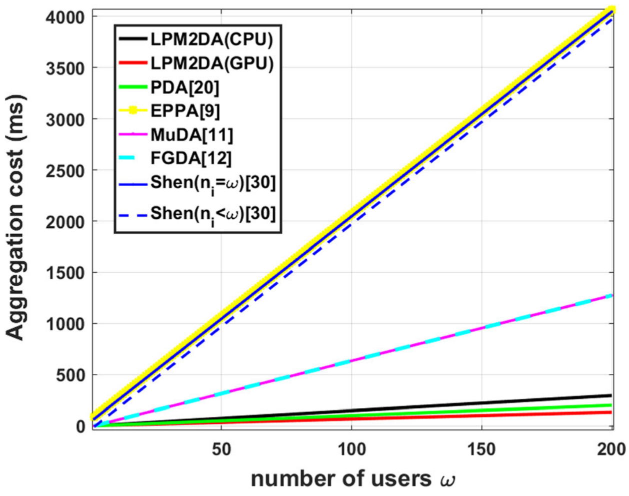

Fig. 2 Cost of aggregation computation

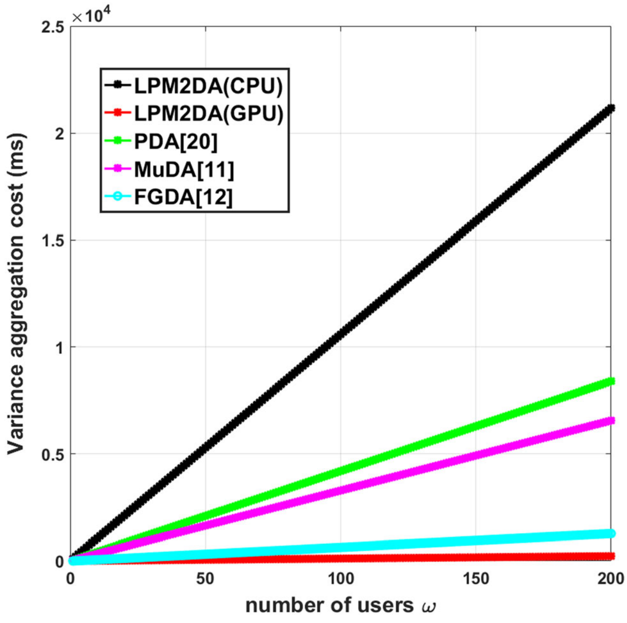

Fig. 3 Cost of variance aggregation computation

Table 4 GW's computational cost

	GW (aggr.)	GW (variance aggr.)
LPM2DA	$(\omega$-1)Add+1Sign +ω Verification	$(\omega$-1)Add+ωMult
PDA [20]	$(\omega$-1)SHAdd	$(\omega$-1)SHAdd+ωSHMult
EPPA [9]	$(\omega+3)C_{bp}+C_m$	–
MuDA [11]	$(\omega\text{-}1)C_m$	$2(\omega\text{-}1)C_m+(\omega+1)C_{bp}$
FGDA [12]	$(\omega\text{-}1)C_m$	$(\omega\text{-}1)C_m$
Shen's scheme [30]	RAGW:$(n_i+2)C_{bp}+(\omega-n_i)C_{ez}+C_m$ DGW:$(\omega_2+2)C_{bp}+C_m$	–

Table 5 CC's computational cost

	CC (aggregation) (ms)	CC (variance) (ms)
LPM2DA (CPU)	Dec <100	Dec <100
LPM2DA (GPU)	Dec <3	Dec <3
PDA [20]	SHDec $=26$	SHDec $=26$
EPPA [9]	$2C_{bp}+C_{ez}+4C_m+C_{et}=86.4$	–
MuDA [11]	$C_{et}+C_{pl}=34.3$ms	$2C_{et}+2C_{pl}=68.5$ms
FGDA [12]	$C_{et}+C_m=14.8$ms	$C_{et}+C_m=14.8$ms
Shen's [30]	$2C_{bp}=40$ms	–

to the asymmetric error growth depending on two subsequent ciphertexts, this multiplication method gives a slow noise growth, which is clearly better than PDA's symmetric noise growth. Hence, it is clear that our scheme not only has lower computation costs, but also can support more users in a residential area. Although the CC is a potent entity, we analyze CC's computational costs in obtaining the aggregation and computing the average and variance of the users' data in Table 5.

Due to the evaluation performed earlier, we can safely conclude that in terms of computation cost and calculating average and variance aggregation, the LPM2DA scheme is efficient in CPU and achieves the lowest computation cost with GPU. Especially, considering PDA's algorithm and EPPA's super-increasing sequence and Horner's rule utilized in Shen's scheme, the Chinese remainder theorem provides the handling of k-dimensional data more efficiently.

6.2 Comparison with Abdallah's scheme [19]

Since the scheme [19] is a DA scheme with a distinct system model; in this subsection, we aim to compare LPM2DA with this lattice-based DA scheme based on computational costs. If we consider each household appliance's data in [19] equivalent to each component of data in our scheme, we can practically compare the schemes. Considering 20 appliances in each HAN, the smart meter and appliances in Abdallah's scheme need nearly 160 ms for computations [19], while the smart meter in our scheme needs less than 100ms (with CPU) and 10ms (with GPU) to compute the 20-dimensional data. It should be noted that based on the data presented in [19], we can see that the key size of [19] is 2.95 Mb while our scheme's key is 63.4 kb, which is significantly better. Therefore, our LPM2DA is much more efficient compared to Abdallah's scheme [19] in terms of computational and communication costs.

6.3 Communication efficiency

To analyze the communication performance of each scheme in smart grid networks, we should consider two parts: communication between SM to GW and between GW to CC. Since GW and CC are powerful entities, the communicational efficiency of a DA scheme relies on the efficiency of smart meters' performance. Specifically, to compare the efficiency of each scheme, we utilize the *expansion metric* as formulated below:

$$\text{Exp} = \frac{\text{bits transmitted}}{\text{data (bit)}} = \frac{1}{\text{Eff}}. \tag{29}$$

To be accurate and fair, we compare the efficiency of LPM2DA with schemes that could transmit multi-dimensional data, such as PDA [20], EPPA [9], and Shen's scheme [30]. Since EPPA and Shen's scheme are using Paillier cryptosystem and expand the k-dimensional data to $2\log(\mathcal{N})$, their expansions are better than lattice-based schemes like LPM2DA and PDA, but they are significantly slower and do not have quantum security. Moreover, the EPPA utilizes a super-increasing sequence and Shen's scheme utilizes Horner's rule to transmit multi-dimensional data. Therefore, those schemes lose computational efficiency to gain k-dimensional data. For communication comparison with PDA, the expansions are presented in the Table 6.

In Table 6, k denotes the number of dimensions of data. The PDA utilizes an algorithm to achieve k-dimensional data in which every component of data can be shown with $\mathcal{K}_1 = 10$ bits as mentioned in [20] and for 128-bit security, $n = 1024$ and $\log(q) = 58$. It should be noted that based on PDA's algorithm with $\mathcal{K}_1 = 10$, the k-dimensional data is practically limited to $k = 4$. However, LPM2DA achieves k-dimensional data by deploying polynomial CRT. Technically, in our scheme, $\log(q) = 31$, and each component of data is a polynomial with average degree $p_j = 100$ and the coefficients are of size $\log(q)$, and there exists no limit on the dimension of data k. Therefore, considering the proposed scheme's quantum security and computational competence, LPM2DA is acceptable in communication overhead comparing to traditional schemes like EPPA [9] and Shen's scheme [30], and it achieves better communication efficiency in contrast to lattice-based schemes like PDA and Abdallah's scheme.

Table 6 Communication overhead

	Bits transmitted/data
LPM2DA	$\frac{(N\times 2)\times n\times \log(q)}{k\times p_j\times \log(q)}$
PDA [20]	$\frac{2\times n\times \log(q)}{k\times \mathcal{K}_1}$

7 Conclusion and future work

Due to the fact that one of the most important design criteria in smart grid networks is preserving the privacy of users' data, in this paper, we have introduced a secure lattice-based multi-functional and multi-dimensional data aggregation scheme, namely LPM2DA. In the situations where networks are in imminent danger of quantum attacks, the LPM2DA scheme not only preserves the privacy of users' consumption data, but also can resist all the potential mentioned threats to the integrity of the transmitted data or the authentication of all the transmissions. Moreover, the LPM2DA scheme maintains fault-tolerant, allows dynamic users, and is resistant against various attacks like impersonation, modification, MITM, and replay attack. Based on the homomorphic encryption scheme and Chinese remainder theorem, the control center could acquire temporal and spatial aggregations of users' multi-dimensional data efficiently; and it is capable of calculating various statistical functions like mean, variance, and skewness. Finally, we have demonstrated the computational and communication efficiency of the proposed scheme by comparing its performance with other data aggregation and lattice-based schemes. In future works, we intend to accomplish other features like preserving differential privacy and detecting collusion between different entities.

Acknowledgements The authors are grateful to the editor and the referees for their very meticulous reading of this manuscript. Their suggestions were very helpful in creating the improved final version. This work was supported in part by the Niroo Research Institute (NRI) under Grant No. 95/10615/51. The research of the third author was supported by a Grant from IPM (No. 1400050115).

References

1. Li, H., Lai, L., Caiming Qiu, R.: Scheduling of wireless metering for power market pricing in smart grid. IEEE Trans. Smart Grid **3**(4), 1611–1620 (2012)
2. Desai, S., Alhadad, R., Chilamkurti, N., et al.: A survey of privacy preserving schemes in IoE enabled Smart Grid Advanced Metering Infrastructure. Clust. Comput. **22**, 43–69 (2019)
3. Jokar, P., Arianpoo, N., Leung, V.C.M.: A survey on security issues in smart grid. Secur. Commun. Netw. **9**, 262–273 (2012)
4. Vahedi, E., Bayat, M., Pakravan, M.R., Aref, M.R.: A secure ECC-based privacy preserving data aggregation scheme for smart grids. Comput. Netw. **129**(P1), 28–36 (2017)
5. Kanwal, T., Anjum, A., Khan, A.: Privacy preservation in e-health cloud: taxonomy, privacy requirements, feasibility analysis, and opportunities. Clust. Comput. **24**, 293–317 (2021)
6. Abdallah, A., Shen, X.: Security and privacy of customer-side networks. In: Abdallah, S., Shen, X. (eds.) Security and Privacy in Smart Grid, pp. 27–64. Springer, Berlin (2018)
7. Aghili, S.F., Mala, H., Shojafar, M., Peris-Lopez, P.: LACO: lightweight three-factor authentication, access control and ownership transfer scheme for E-health systems in IoT. Future Gener. Comput. Syst. **96**, 410–424 (2019)
8. Singh, S., Chaurasiya, V.K.: Mutual authentication scheme of IoT devices in fog computing environment. Clust. Comput. (2020). https://doi.org/10.1007/s10586-020-03211-1
9. Lu, R., Liang, X., Li, X., Lin, X., Shen, X.: EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications. IEEE Trans. Parallel Distrib. Syst. **23**(9), 1621–1631 (2012)
10. Zhang, L., Zhang, J., Hen Hu, Y.: A privacy-preserving distributed smart metering temporal and spatial aggregation scheme. IEEE Access **7**, 28372–28382 (2019)
11. Chen, L., Lu, R., Cao, Zh., Alharbi, Kh., Lin, X.: MuDA: maultifunctional data aggregation in privacy-preserving smart grid communications. Peer-to-Peer Netw. Appl. **8**, 777–792 (2015)
12. Ge, Sh., Zeng, P., Lu, R., Choo, K.-K.R.: FGDA: fine-grained data analysis in privacy-preserving smart grid communications. Peer-to-Peer Netw. Appl. **11**(5), 966–978 (2018)
13. Ni, J., Zhang, K., Alharbi, Kh., Lin, X., Zhang, N., Shen, XSh.: Differentially private smart metering with fault tolerance and ranged-based filtering. IEEE Trans. Smart Grid **8**(5), 2483–2493 (2017)
14. Ming, Y., Zhang, X., Shen, X.: Efficient privacy-preserving multi-dimensional data aggregation scheme in smart grid. IEEE Access **7**, 32907–32921 (2019)
15. Huang, C., Wang, X., Gan, Q., et al.: A lightweight and fault-tolerable data aggregation scheme for privacy-friendly smart grids environment. Clust. Comput. (2021). https://doi.org/10.1007/s10586-021-03345-w
16. Shor, P.W.: Algorithm for quantum computations: discrete logarithm and factoring. In: Proceedings of 35th annual symposium on foundations of computer science, pp. 124–134 (1994)
17. Agarkar, A.A., Agrawal, H.: LRSPPP: lightweight R-LWE-based secure and privacy-preserving scheme for prosumer side network in smart grid. Heliyon **5**(3), e01321 (2019)
18. Abdallah, A., Shen, X.: Lightweight security and privacy preserving scheme for smart grid customer-side networks. IEEE Trans. Smart Grid **8**(3), 1064–1074 (2017)
19. Abdallah, A., Shen, X.S.H.: A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid. IEEE Trans. Smart Grid **9**(1), 396–405 (2018)
20. Li, Ch., Lu, R., Li, H., Chen, L., Chen, J.: PDA: a privacy-preserving dual-functional aggregation scheme for smart grid. Secur. Commun. Netw. **8**(15), 2494–2506 (2015)
21. Arun Jees, S., Gomathi, V.: Load forecasting for smart grid using non-linear model in Hadoop distributed file system. Clust. Comput. **22**, 13533–13545 (2019)
22. Rabie, A.H., Ali, S.H., Ali, H.A., et al.: A fog based load forecasting strategy for smart grids using big electrical data. Clust. Comput. **22**, 241–270 (2019)
23. Rabie, A.H., Ali, S.H., Saleh, A.I., Ali, H.A.: A new outlier rejection methodology for supporting load forecasting in smart grids based on big data. Clust. Comput. **23**, 509–535 (2020)
24. Wu, Y., Huang, Z., Zhang, J., Wen, Q.: A lattice-based digital signature from the ring-LWE. In: Proceedings of 3rd IEEE International conference on network infrastructure and digital content, Beijing, China (2012)
25. Khedr, A., Gulak, G., Vaikuntanathan, V.: SHIELD: scalable homomorphic implementation of encrypted data-classifiers. IEEE Trans. Comput. **65**, 2848–2858 (2016)
26. Ajtai, M.: Generating hard instances of lattice problems. Electron. Colloq. Comput. Complex. (ECCC) **3**, 1 (1996)
27. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. J. ACM (JACM) **56**, 84–93 (2005)
28. Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. In: Proceedings of 29th International conference on the theory and applications of cryptographic techniques (EUROCRYPT), Berlin, Heidelberg (2010)
29. Ding, C., Pei, D., Salomaa, A.: Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography, pp. 1–213. World Scientific Publishing, Singapore (1996)
30. Shen, H., Zhang, M., Shen, J.: Efficient privacy-preserving cube-data aggregation scheme for smart grids. IEEE Trans. Inf. Forensics Secur. **12**(6), 1369–1381 (2017)
31. Lynn, B.: PBC Library. http://crypto.stanford.edu/pbc/ (2012)
32. http://www.shamus.ie/, Multiprecision Integer and Rational Arithmetic c/c++ Library (2012)
33. Longa, P., Naehrig, M.: Speeding up the number theoretic transform for faster ideal lattice-based cryptography. In: Proceedings of International conference on cryptology and network security (CANS) (2016)
34. Can Mert, A., Öztürk, E., Savas, E.: Design and implementation of a fast and scalable NTT-based polynomial multiplier architecture. In: Proceedings of 22nd Euromicro conference on digital system design (DSD) (2019)

Saleh Darzi received his B.Sc. in Electrical Engineering (Electronic) from Islamic Azad University of Central Tehran Branch, Tehran, Iran, in 2016. He is currently an M.Sc. student of Electrical Engineering (Communication-System) at K. N. Toosi University of Technology, Tehran, Iran. His main research interests include applied and post-quantum cryptography, privacy and security of IoT, Blockchain, and network security.

Bahareh Akhbari received the B.Sc. degree in 2003, the M.Sc. degree in 2005, and the Ph.D. degree in 2011, all in Electrical Engineering from Sharif University of Technology (SUT), Tehran, Iran. She was also a visiting Ph.D. student at the University of Minnesota for one year, starting in 2010. Since 2012, she is an assistant professor of the Faculty of Electrical Engineering, K. N. Toosi University of Technology (KNTU), Tehran, Iran. Her research interests include information theory, cryptography and network security, communication theory and information-theoretic security.

Hassan Khodaiemehr received the B.Sc. degree in mathematics, the B.A.Sc. degree in electrical engineering, and the M.Sc. and Ph.D. degrees in mathematics from the Amirkabir University of Technology, Tehran, Iran, in 2010, 2012, and 2017, respectively. From October 2015 to August 2016, he was a Visitor with the School of Mathematics and Statistics, Carleton University, Ottawa, Canada. From October 2017 to February 2018, he was a Post-Doctoral Fellow with the Institute for Research in Fundamental Sciences (IPM), Tehran. He is currently an Assistant Professor with the Computer Science and Statistics Department, K. N. Toosi University of Technology, Tehran. His research interests include data science, lattice codes and their applications in wireless communications, coding and information theory, cryptography, and physical layer security.