IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 August 2023, accepted 29 August 2023, date of publication 1 September 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3311140

RESEARCH ARTICLE

LPPDA: A Light-Weight Privacy-Preserving Data Aggregation Protocol for Smart Grids

NAHEEL FAISAL KAMAL[1,2], ABDULLA KHALID AL-ALI[3], (Member, IEEE), ABDULAZIZ AL-ALI[4], SERTAC BAYHAN[5], (Senior Member, IEEE), AND QUTAIBAH M. MALLUHI[3], (Member, IEEE)

[1]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Qatar Foundation, Doha, Qatar
[2]Electrical and Computer Engineering Department, Texas A&M University, College Station, TX 77843, USA
[3]Department of Computer Science and Engineering, Qatar University, Doha, Qatar
[4]KINDI Center for Computing Research, Qatar University, Doha, Qatar
[5]Qatar Environment and Energy Research Institute (QEERI), HBKU, Qatar Foundation, Doha, Qatar

Corresponding author: Naheel Faisal Kamal (naheel@tamu.edu)

This work was supported in part by the Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP12C-33905-SP-66, and in part by the Open Access funding provided by the Qatar National Library.

ABSTRACT Smart meters are continuously being deployed in several countries as a step in the direction of modernizing the power grid. Smart meters allow for automatic electricity consumption reporting to energy providers to facilitate billing and demand-based power generation. However, research has shown that such high resolution reporting to suppliers can potentially be used to invade consumers' privacy; by identifying and predicting their behavior based on their consumption readings. This work presents a new protocol to preserve users' privacy while maintaining the benefits of smart grids. The proposed method utilizes different techniques like randomization, masking, and differential privacy to build the scheme. The proposed method is shown to be more efficient compared to previous work in terms of performance and communication overhead. The implementation, simulation, and analysis are performed on datasets of real smart meters readings of households and electric vehicle chargers.

INDEX TERMS Privacy, security, smart meters, electric vehicle charging.

I. INTRODUCTION

The smart grid is an ever growing field of engineering and technology [1]. It attempts to transform traditional power grids into smart and reactive grids. Such grids need to operate several components and enable communication between many entities like energy suppliers; companies for power generation, transmission, and distribution; and down to the consumers with metering devices.

Using such a large scale network raises security concerns and urges researchers to investigate suitable solutions to provide acceptable confidentiality, integrity, and availability. In addition to typical security concerns, users' privacy is another major issue [2]. Power consumption data can allow entities with access to this data to infer private information about the users. Examples of such information include human location and activity patterns [3], detection of household devices and occupants [4], and profiling of electric vehicles (EVs) [5].

Privacy preserving smart grid systems have been surveyed in the literature [6]. Several techniques and tools have been used by researchers to address this issue. One notable commonly used concept is known as "differential privacy" that is a definition to formalize data privacy [7]. By adding controlled noise to the data, it ensures that the presence or absence of an individual does not affect the final result. "Local" differential privacy is also commonly utilized where each node adds noise locally to avoid the need for a trusted data curator. These techniques have had a significant role in advancing privacy preservation of power consumption data [8]. Other methods have been used to preserve energy consumption privacy like

The associate editor coordinating the review of this manuscript and approving it for publication was Leandros Maglaras.

homomorphic encryption [9], [10] and blockchain-based systems [10], [11], [12].

This paper attempts to provide a complete protocol to secure and preserve the privacy of users in smart grids without relying on a central or trusted entity. The contribution of this work can be summarized in the following points.

1) Introduce a novel comprehensive and efficient protocol to secure and preserve users' privacy in smart grids.
2) Incorporate hash chain keys to achieve local differential privacy without relying on a trusted party.
3) Tolerate faulty or disconnected smart meters at the supplier side without revealing individual meters' measurements.
4) Simulate the implemented protocol with real datasets and evaluate with respect to existing similar protocols.

This paper is organized as follows: Section II lists and compares some related methods in the literature. Section III introduces the grid model, adversary model, and the requirements of the problem. The proposed solution and preliminary concepts are described in Section IV. The scheme is analyzed and evaluated in Section V. Section VI concludes the paper with final remarks and future research directions.

II. RELATED WORKS

Several researchers have tried to approach smart grid data aggregation in a secure and privacy preserving manner. This is due to the fact that aggregation is a basic required function in smart grids and attacks have been developed to violate users privacy from the data being aggregated [2].

Blockchains have been widely used in recent work in this area [10], [11], [12]. A blockchain based grid aggregation scheme is proposed in [11] to preserve the data privacy. Blockchains are used to enable data immutability and unforgeability. The consumption data is sent to one of the nodes in a residential area to function as an aggregator. This node acts as a mining node in the blockchain and is elected by other nodes using peer-to-peer communication. Thus, trust is indirectly placed on the mining node. A consortium blockchain is used in another work with ring learning with errors (RLWE) as a post-quantum encryption method for metering data [12]. Meter measurements are encrypted and signed before sending to the aggregator. Homomorphic encryption and hash chains are used in combination with blockchains in [10] to secure the grid data. The blockchain in this work functions on the edge to reduce the overhead on the smart meters. Different homomorphic cryptography techniques were applied to ensure privacy of smart meter readings in the smart grid. Paillier cryptosystem is used in [9] to hide individuals' usage data. This method aggregates data, perform batch verification, and tolerate faults. Such homomorphic operations are, however, relatively expensive and might not be the best for low-cost micro controllers.

The systems mentioned previously rely on a trusted authority to initialize the system. Additionally, using blockchain brings a significant overhead to the system that can be problematic in the context of smart grids with weak edge devices.

Aggregation of meter data can be done over multiple parties as presented in [13], after initialization of a trusted authority, each smart meter has multiple "proxies" where each proxy is another smart meter. Smart meters mask their readings before sending and each of their proxies mask again to eventually cancel the initial mask after aggregating. This means that all proxies need to collude to reconstruct the original readings, but this also requires strict coordination between all nodes.

Utilizing differential privacy in the smart grid has been also studied in the literature [14], [15], [16]. It was shown that differential privacy is able to prevent load monitoring in smart grids [15]. A distributed Laplacian or Gaussian noise is masked with smart meter readings in [16]. This work introduces some form of local differential privacy where each meter adds Gamma noise. When the meters are summed together, the total becomes Laplacian. In this work, Keys are shared with the supplier and peer users. "dummy" keys are generated between groups of users to force the supplier to aggregate the results and cancel out those keys. This scheme is light in terms of cryptographic operations, however, the communication between smart meters is increased. A similar work follows the same noise generation mechanism but makes the data more private by shuffling the measurements over a time window [17]. Gamma noise is added to the readings before sending to an aggregator that sums the measurements and send them to the supplier. This method requires less communication overhead, however, shuffling the data to hide the consumption information would effectively be equivalent to reducing the sampling rate.

Another work used a more traditional approach of randomized response in smart grids [18]. Knowing that users report consumption data in known ranges, this work divides this data into intervals. This idea is used to transform the readings to discrete values. Using k-Randomized Response, these discrete values are perturbed. The frequencies of the resulting values are then calculated by the aggregator to effectively get a differentially private aggregate of the data.

Differential privacy has been also used along with the concept of virtual batteries [19]. The virtual batteries act as a method to preserve users' privacy by charging and discharging while providing accurate billing. This work utilizes aggregators functioning in a fog architecture and verifies the authenticity of messages.

EV chargers have been attacked in a similar manner using metering data [5], [20]. Power readings are exploited in [20] whereas current measurements are used in [5]; both from real EV charging stations datasets. Discrete Wavelet Transform (DWT) is used to extract the power load profiles in [20]. To prevent such attacks, authors propose additive charging load patterns. Such technique should preserve EV users' privacy.

TABLE 1. Related works features comparison; F1: Privacy preservation, F2: Differencially private, F3: Locally differencially private, F4: No trusted entity, F5: Integrity verification, F6: Fault tolerance, F7: Private billing.

	LPPDA	[11]	[9]	[13]	[16]	[17]	[18]	[19]
F1	✓	✓	✓	✓	✓	✓	✓	✓
F2	✓	-	-	-	✓	✓	✓	✓
F3	✓	-	-	-	✓	✓	✓	-
F4	✓	✓	-	-	✓	-	✓	✓
F5	✓	✓	✓	✓	-	-	-	✓
F6	✓	-	✓	-	✓	-	-	-
F7	✓	✓	-	✓	-	✓	-	✓

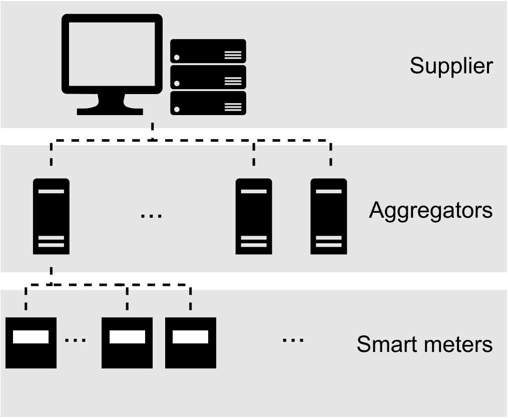


FIGURE 1. Grid network model.

Overall, the proposed LPPDA preserves smart meters' data privacy with locally differential noise, does not rely on a trusted authority, verifies the integrity, tolerate faulty nodes, and allows billing in a privately. Up to the knowledge of the authors, none of the existing literature offer those features in one protocol. Table 1 lists the discussed features in the proposed LPPDA and some related works found in the literature.

III. GRID AND ADVERSARY MODEL

The components of the smart grid can be abstractly modeled using three main categories as shown in Figure 1. A smart meter (SM) is the meter installed on the customers' side. It is the sensing device in the system and is assumed to report real measurements. A supplier (S) is the entity responsible for supplying the main grid electricity and charges consumers. An aggregator (A) collects consumption readings from a cluster of N SMs and submits the aggregates of those readings to S.

A. ADVERSARY

LPPDA considers three types of adversaries as follows:

1) SUPPLIERS AND AGGREGATORS

are considered to be honest-but-curious. They follow the protocols honestly but they are curious in accessing the users' data. They may attempt to read personal data of users, infer behavioral data, and violate their privacy. Such a situation is practically applicable even though it allows a level of trust in the honesty of the suppliers. This is due to the nature of the system where the supplier is the entity in charge of providing the service. To reduce the complexity of the system, collusion between aggregators and suppliers is not considered in this work. Such a threat can be avoided using different methods like collaborative generation of keys between SMs [16] or using different SMs as proxies [13]

2) SMART METERS

are also considered to be honest-but-curious. They follow the protocol and send correct metering values being the physical measuring devices in the system. However, smart meters can be curious to violate the privacy of smart meters other than themselves.

3) EXTERNAL ADVERSARIES

are malicious and untrusted. They may attempt to eavesdrop the channel, inject false data, and modify readings.

B. REQUIREMENTS

Requirements concerning the privacy of customers in the context of smart grids focus on power consumption data. Suppliers typically use this data for billing and for analysis. Analytics, prediction, and on-demand power generation require high resolution readings from smart meters. This is because suppliers need to adjust their power generation based on the data gathered from consumers. Billing, on the other hand, can be performed over longer periods of time. This process has little impact on the user's privacy as it cannot be used to predict consumer behavior on daily basis.

In addition to preserving the privacy of consumers, security should also be preserved between all involved parties. The communication channel must always remain confidential. There should be no entity capable of accessing the individual power consumption readings from a meter, other than the consumer himself. Integrity of the communicated data must be preserved. Otherwise, attackers can modify the data sent and cause harm to the connected devices. For example, an attacker can increase the power readings causing the grid to supply more power and possibly charge the customer higher fees, or reduce the readings for the benefit of the customer to be charged less.

IV. PROPOSED SOLUTION

The solution proposed in LPPDA relies on some preliminary concepts. Below are descriptions of the basic concepts that act as the building blocks of the proposed work.

A. HASH CHAINS

Chains of one-way functions were proposed initially by Lamport in [21] as an authentication method for insecure communication. The technique was later adopted in [22] to build a broadcast authentication protocol. Hash functions were used as the one-way functions in their approach. Similar concepts are now adopted in blockchains to ensure consensus of the distributed ledger data between nodes.

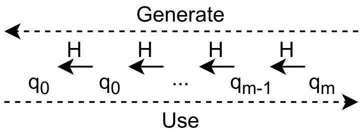


FIGURE 2. Hash chain key generation and use.

The idea is illustrated in Figure 2. The process starts with a secret seed key q_m where m is the number of hashes to be applied. The larger m is, the less initializations needed for the chain. In each step j, the key q_j is constructed as $H(q_{j-1})$ and is used to generate the next key $q_{j+1} = H(q_j)$ down to $j = 0$. After $m - 1$ steps, the keys are used in reverse order starting from q_0 up to q_m.

B. DIFFERENTIAL PRIVACY

Differential privacy is a definition to achieve privacy-preserving analysis over a given computational task [7]. An algorithm is said to be (ϵ, δ)-differentially private if the following condition holds.

$$Pr[\mathcal{M}(DB_1) \in S] \leq e^{\epsilon} Pr[\mathcal{M}(DB_2) \in S] + \delta \tag{1}$$

where $\mathcal{M}$ is randomized algorithm (mechanism) and $S \subseteq Range(\mathcal{M})$. DB_1 and DB_2 are neighboring databases where they differ by one record. That is, $||DB_1 - DB_2||_1 \leq 1$ where $||.||_1$ is the l_1 distance.

This kind of setup usually requires a "curator" who applies the algorithm in a differentially private manner by adding controlled noise to the algorithm's output. Such an entity needs to be fully trusted as it has access to the original data. To avoid this situation, *local differential privacy* aims to add the noise by each user's (e.g. smart meter's) data locally [7]. This hides the original data from all entities other than the data owners themselves.

Differential privacy is needed in such scenarios to preserve the privacy of a group of users. It provides the methods to ensure privacy even when adversaries collude with other internal entities [16]. This concept is even stronger with local differential privacy as it functions as a superior model where only individual users have access to their data [7].

C. NOISE GENERATION MECHANISM

The noise added in the perturbation part is generated using a differentially private mechanism. In particular, a Laplace-based mechanism $\mathcal{M}$ is used in this work.

$$\mathcal{M}(X, f, \epsilon) = f(X) + \mathcal{L}(\lambda) \tag{2}$$

where mechanism $\mathcal{M}$ curates the value X over the function f with privacy budget ϵ. The noise $\mathcal{L}(\lambda)$ is drawn from a Laplace distribution with scale λ

$$\lambda = \Delta f / \epsilon \tag{3}$$

And $\Delta f = max||f(DB_1) - f(DB_2)||_1$ where DB_1 and DB_2 differ by one entry and $||.||_1$ is the l_1 distance.

This mechanism provides $(\epsilon, 0)$-differential privacy to the system [7].

One problem with the basic solution above is that it requires a trusted curator to add the noise to the aggregated results. To reduce the attack surface, a local differentially private protocol can be applied with little additional overhead. This can be achieved by distributing the Laplace mechanism over individual users using other distributions like Gamma and Gauss [23].

An example is to use the Gamma distribution as proposed in [16]:

$$\mathcal{L}(\lambda) = \sum_{i=1}^{N} (\Gamma_1(N, \lambda) - \Gamma_2(N, \lambda)) \tag{4}$$

where the noise is picked individually by each SM_i as two independent variables $\Gamma_1(N, \lambda)$ and $\Gamma_2(N, \lambda)$ from the same gamma distribution. Once the aggregator sums the noisy readings of meters in the cluster, the aggregated result gains a Laplacian noise. However, this technique can leak some information from individual users even when the noise is added locally. This is because the differential privacy is achieved when all those individual readings are summed together and not at the single reading level. For that reason, adding masking keys can be used to force aggregating those readings as shown below in the scheme construction.

D. SCHEME CONSTRUCTION

The solution proposed for LPPDA can be summarized in Figure 3. Note that a secure communication channel is assumed for the data transmitted between different entities of the system. This includes traditional encryption to keep the shared keys and data confidential and any form of public key infrastructure (PKI) to manage public keys shared at the initialization phase.

The terminology used in this paper is listed in table 2 and the steps of the scheme are given below.

1) INITIALIZATION

- The system is initialized at the beginning and is reinitialized after m time slots. m is publicly chosen by S.
- S shares a random large secret, q_0^i, with SM_i. This key is used to hide readings from aggregators and external adversary but not from supplier.
- A shares a random large secret, r_m^i, with SM_i. This key is used to ensure the integrity and authenticity of readings and to hide the readings from the supplier.
- S uses q_0^i to seed a pseudo random number generator RNG to generate a chain of keys for each time slot t up to q_m^i.
- A generates a hash chain starting with r_m^i down to r_0^i for each time slot t. The chain is then used in reverse. Each r_t^i at timestamp t is used to seed a RNG to generate $r_{t(mask)}^i$ and $r_{t(verify)}^i$.

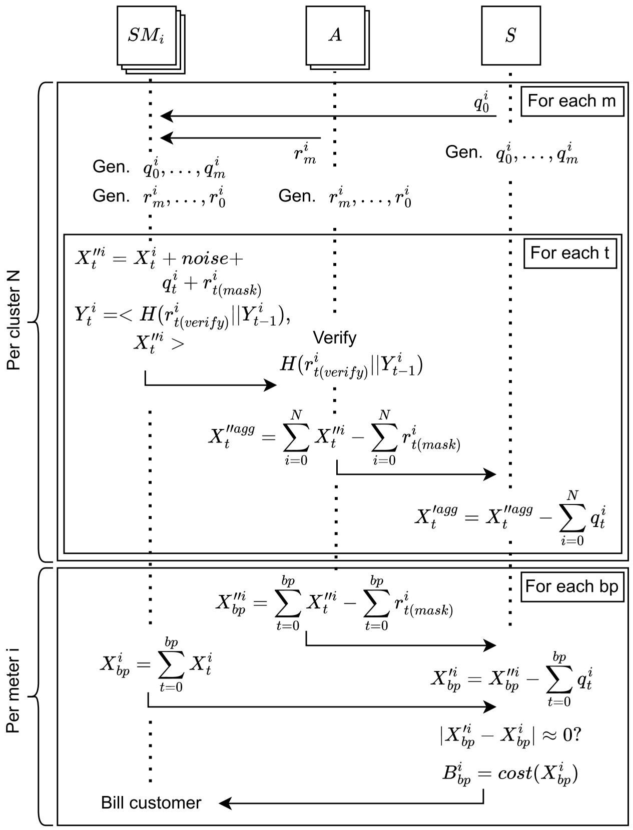

SM_i
A
S
q_0^i
For each m
r_m^i
Gen. $q_0^i,\dots,q_m^i$
Gen. $r_m^i,\dots,r_0^i$
Gen. $r_m^i,\dots,r_0^i$
Gen. $q_0^i,\dots,q_m^i$
For each t
$X''^i_t = X^i_t + noise + q^i_t + r^i_{t(mask)}$,
$Y^i_t =< H(r^i_{t(verify)}||Y^i_{t-1})$, $X''^i_t >$
Verify
$H(r^i_{t(verify)}||Y^i_{t-1})$
$X''^{agg}_t = \sum_{i=0}^{N} X''^i_t - \sum_{i=0}^{N} r^i_{t(mask)}$
$X'^{agg}_t = X''^{agg}_t - \sum_{i=0}^{N} q^i_t$
Per cluster N
For each bp
$X''^i_{bp} = \sum_{t=0}^{bp} X''^i_t - \sum_{t=0}^{bp} r^i_{t(mask)}$
$X^i_{bp} = \sum_{t=0}^{bp} X^i_t$
$X'^i_{bp} = X''^i_{bp} - \sum_{t=0}^{bp} q^i_t$
$|X'^i_{bp} - X^i_{bp}| \approx 0?$
$B^i_{bp} = cost(X^i_{bp})$
Bill customer
Per meter i

FIGURE 3. Overall system timing diagram.

TABLE 2. Terminology.

Abbr.	Description
S	Supplier
A	Aggregator
SM_i	Smart meter of customer i
v^i_t	A variable for customer i at time t
q, r	Large randoms
X	Power consumption reading
X'	Noisy power consumption reading
X''	Noisy and encrypted power consumption reading
N	Cluster size
m	Size of the chains
$H()$	Hash function
RNG	A secure random number generator
$a\|\|b$	Concatenation between a and b
$< a, b >$	A tuple of elements a and b

- SM_i uses q^i_0 and r^i_m to generate both chains identical to what S and A have.

2) ENCRYPTION AND PERTURBING AT SMi

- SM_i perturbs and encrypts each reading X^i_t as

$$X'^i_t = X^i_t + noise \tag{5}$$

$$X''^i_t = X'^i_t + q^i_t + r^i_{t(mask)} \tag{6}$$

where the noise generation is described in subsection IV-D6.
- For each reading X^i_t, a message Y^i_t is formed by concatenating the reading with the signature of the previous reading

$$Y^i_t =< H(r^i_{t(verify)}||Y^i_{t-1}), X''^i_t > \tag{7}$$

3) VERIFICATION AND AGGREGATION AT A

A verifies that the received message Y^i_t is from the authorized meter SM_i using the hash chain and the previous message Y^i_{t-1}.

- A receives and expands Y^i_t from each SM_i for $i \in N$
- Knowing that A has r^i_t and Y^i_{t-1}, it generates $r^i_{t(verify)}$ and verifies the newly sent $H(r^i_{t(verify)}||Y^i_{t-1})$ from each SM_i
- The values are then aggregated, unmasked of $r^i_{t(mask)}$, and sent to S

$$X''^{agg}_t = \sum_{i=0}^{N} X''^i_t - (\sum_{i=0}^{N} r^i_{t(mask)}) \tag{8}$$

4) DECRYPTION AT S

- S decrypts X''^{agg}_t by removing the sum of the q keys to get the noisy aggregate

$$X'^{agg}_t = X''^{agg}_t - (\sum_{i=0}^{N} q^i_t) \tag{9}$$

5) BILLING AT SMi, A, AND S

- Billing occurs after each billing period bp.
- A keeps track of each user's reported X''^i_t and calculates X''^i_{bp}

$$X''^i_{bp} = \sum_{t=0}^{bp} X''^i_t - (\sum_{t=0}^{bp} r^i_{t(mask)}) \tag{10}$$

- X''^i_{bp} is sent to S and S calculates X'^i_{bp}

$$X'^i_{bp} = X''^i_{bp} - (\sum_{t=0}^{bp} q^i_t) \tag{11}$$

- SM_i reports what should be the true value of X^i_{bp} to S

$$X^i_{bp} = \sum_{t=0}^{bp} X^i_t \tag{12}$$

- S validates that $|X'^i_{bp} - X^i_{bp}| \approx \mathbb{E}(|(\sum_{i=1}^{N}(\Gamma_1(N,\lambda) - \Gamma_2(N,\lambda))) - \mathcal{L}(\lambda)|)/bp \approx 0$
- If the condition above does not hold, S rejects X^i_{bp} and flags an issue on SM_i
- Otherwise, S applies a cost function over the given power consumption X^i_{bp}

$$B^i_{bp} = cost(X^i_{bp}) \tag{13}$$

- S sends the bill B^i_{bp} back to SM_i.

The process of calculating the bill at SM_i and at A is useful to avoid situations where users report different values at the time of billing than the values sent periodically at each time slot. This billing verification approach only focuses on matching the submitted measurements, but does not attempt to find anomalies of energy theft situations. For that, machine learning-based methods can be used as found in the literature [24].

To enable time of use billing, SM_i and A can aggregate multiple values; one corresponding to each rate category and send these multiple aggregated values for billing instead of sending one lump sum value.

6) DIFFERENTIALLY PRIVATE NOISE

Noise is drawn from a Gamma distribution as described earlier. Each SM_i adds the noise to its readings in subsection IV-D2 as the difference of two random variables. Over the whole cluster, this adds up to become a Laplacian random.

$$noise = \Gamma_1(N, \lambda) - \Gamma_1(N, \lambda) \tag{14}$$

where $\lambda = \Delta f / \epsilon$ and Δf is calibrated to the expected values of the power readings.

E. COMMUNICATION FAULT TOLERANCE

Verification and aggregation are performed by the aggregator A. The supplier S expects the aggregated result from A to be masked by N meters. This can be problematic if at least one key is missing as S subtracts back all the keys it expects the meters have added. Such case can happen if verification fails or the node is down or not responding. If such an occasion occur, A creates a set of IDs $\beta \subseteq N$ of malfunctioning meters, adds $\sum_{j \in \beta} q^j_{t(mask)}$ to $X^{\prime\prime agg}_t$, and sends β along with the new $X^{\prime\prime agg}_t$ to S. Decryption at S is then performed as $X^{\prime agg}_t = (\sum_{i=0}^{N} X^{\prime\prime i}_t) - (\sum_{i \in N} q^i_t) + (\sum_{j \in \beta} q^j_t)$. Additionally, knowing that X^{agg}_t is differentially private means that the impact of loosing some readings will not cause any significant effect on the aggregated values. This is due to the statistical nature of the aggregation operation where the output describes the community and not any particular individual [7].

V. ANALYSIS AND EVALUATION

The proposed scheme needs to be lightweight and efficient to be deployed on low-cost hardware at a large scale while still protecting the privacy of the consumers. Improving the security and privacy of smart grid systems should have as little impact on its performance as possible. The communication overhead must also be as minimal as possible. This is important to avoid overwhelming the network while handling a large number of nodes.

A. EXPERIMENTAL SETUP

To test the practicality of the proposed solution, the system was implemented and simulated. The implementation was done in Python and run on a PC with an Intel Core i7 and 16GB of RAM. In this implementation, smart meters, aggregators, and suppliers function as separate nodes communicating over a TCP connection. Smart meters have an abstract interface to collect data. It was built this way to allow future expansions to different kinds of meters. LPPDA should work for any electricity consumption data. In our experiments, two real meters datasets are used. One is for home smart meters in London[1] and the other is for EV chargers from Caltech University, California[2] [25]. Each smart meter node is assigned to a randomly selected values of one meter from the used datasets. Nodes are then run in parallel as separate processes communicating with the aggregator process and the supplier process.

The proposed LPPDA protocol is evaluated against two other related works; DPPDA [11] and NHP3 [9]. DPPDA relies on blockchains where each meter submit the readings to one chosen node to be the "mining node" that publish the readings to the blockchain where Paillier cryptosystem to encrypt the data sent to the mining node. NHP3, on the other hand, uses another Paillier-based homomorphic encryption to ensure that no adversary can access private metering values. Performance and communication overheads are discussed in the following subsections.

B. SECURITY ANALYSIS

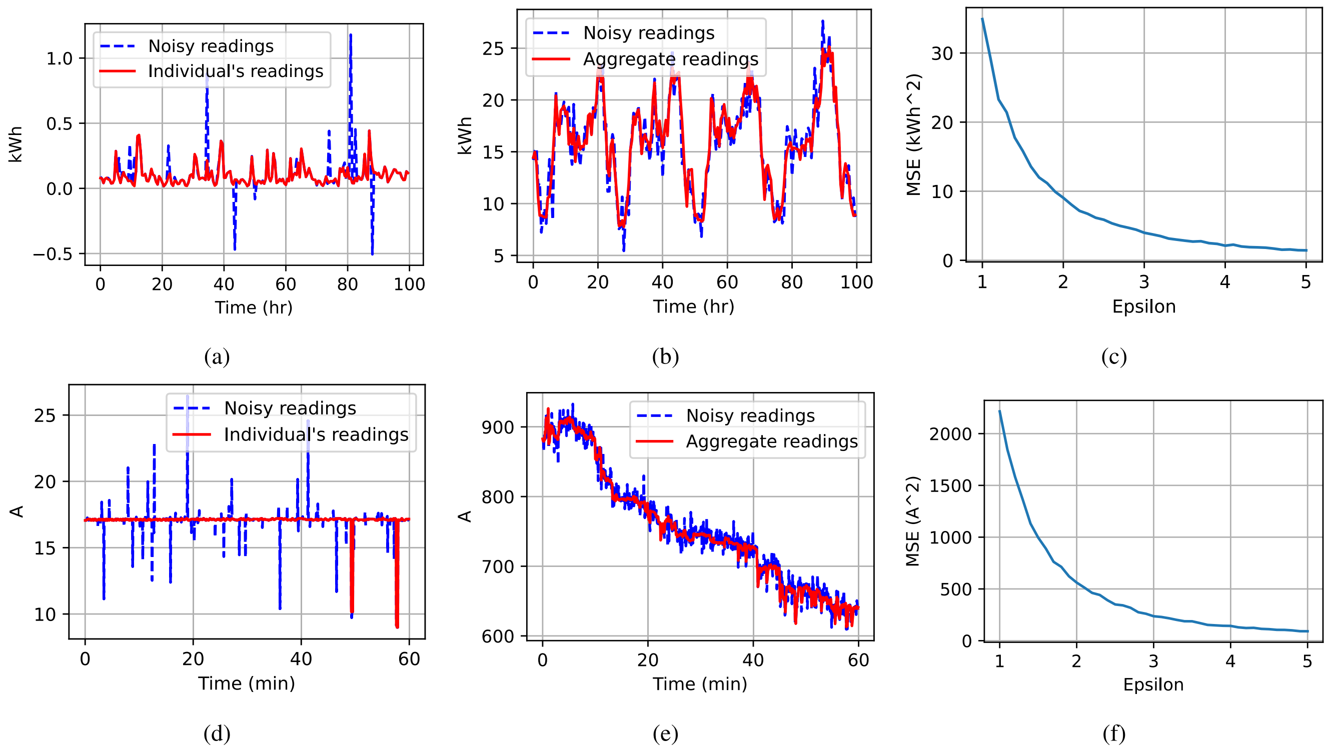
The security of the proposed protocol is analyzed to meet the assumption of the adversary model in Section III. A curious adversary would be interested to find X^i_t for a particular meter SM_i. If an aggregator A attempts to recover X^i_t, it monitors $X^{\prime\prime i}_t$ sent by SM_i and subtracts $r^i_{t(verify)}$ to be left with $X^i_t + noise + q^i_t$ which is indistinguishable from random noise. An aggregator adversary cannot subtract q^i_t as this key is only known to SM_i and supplier S. Recovering X^i_t is also hard for a curious supplier S. First, S needs to eavesdrop the communication channel between SM_i and A as S only receives $X^{\prime\prime agg}_t$ and have no direct access to $X^{\prime\prime i}_t$. Assuming S manages to gain such access, which is hard under the scheme assumptions, it would still only be able to calculate $X^i_t + noise + r^i_{t(verify)}$ after subtracting q^i_t. An external adversaries or a curious smart meter SM_j with access to the communication channel between target meter SM_i and A would also fail to recover any useful data as neither of the keys q^i_t nor r^i_t are known. If such an adversary attempt to modify Y^i_t before it reaches A, send their own Y^i_t, or replay previously sent value of Y^i_t, then A would recognize that the received Y^i_t is corrupted as the calculated $H(r^i_{t(verify)} || Y^i_{t-1})$ would fail to match with the one in Y^i_t. In an extreme case, S would collude with A or with $N - 1$ smart meters to recover X^i_t of a target SM_i. In this case, S may attempt to calculate $X^{\prime agg}_t - (\sum_{j \in N \setminus i} X^j_t - q^j_{t(mask)})$, however, this still would not equal to X^i_t thanks to the differentially private perturbation.

C. DIFFERENTIAL PRIVACY

Applying a differentially private noise to the power consumption readings needs to be calibrated to match the given data. This is done by choosing values for ϵ, the privacy budget,

[1] https://www.kaggle.com/jeanmidev/smart-meters-in-london

[2] https://ev.caltech.edu/dataset



(a) (b) (c)

(d) (e) (f)

FIGURE 4. **Clean and noisy readings of an individual meter 4a and 4d, clean and noisy readings of aggregate of N=70 4b and N=100 4e meters, and MSE for different ϵ values 4c and 4f.**

to change the scale of the noise distribution as described in Equation 3.

The utility can be estimated by calculating the error in the noisy readings. For a sample period of time tp, the mean square error (MSE) is used in this work, where

$$MSE = \frac{1}{tp}\sum_{t=0}^{tp}(X_t'^{agg} - X_t^{agg})^2 \tag{15}$$

Figure 4 shows the MSE for different values of ϵ along with measurements examples from the two datasets. Figures 4c and 4f show the MSE with respect to different values of ϵ. The noisy and clean power readings of an individual user is illustrated in Figure 4a for a period of 4 days and in Figure 4d for 1 hour. Figure 4b shows an aggregate of readings over a cluster of 70 meters and Figure 4e for a cluster of 100 EV chargers' meters. ϵ is chosen to be 3 in both test cases. This values is chosen to provide a good balance between the privacy budget and the added error as shown in Figures 4c and 4f. It can be seen that the noise has much less effect on the aggregated data than on an individual, which can be considered good for individuals. Masking is then used as described in Subsection IV-D to completely hide the readings from the aggregator.

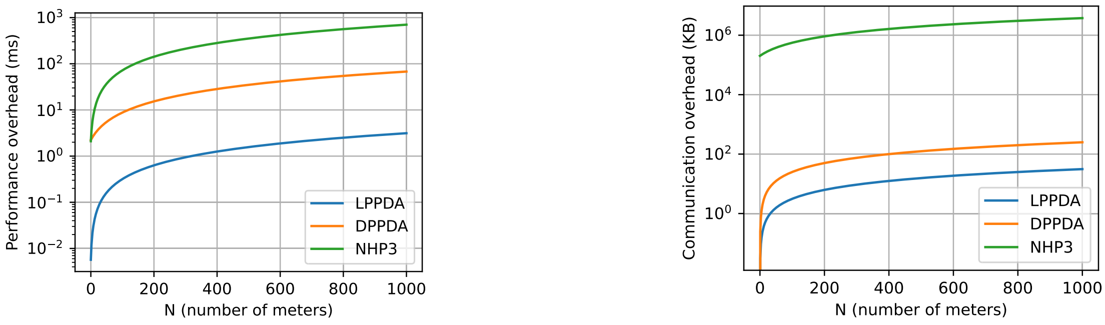
D. PERFORMANCE OVERHEAD

The performance of the system depends on several variables and network conditions. To measure the overhead effect of our protocol, the time measurements were done over the operations described in Subsection IV-D. Formulas for performance overhead of this protocol are given and compared to two techniques in Table 3.

The performance of LPPDA is shown in Figure 5. Figure 5a shows the time overhead with respect to the size of the cluster N. The time is measured starting from getting the readings from N meters and ending with the final decrypted aggregate reading at the supplier S. Processing a single reading from a single meter has an $O(1)$ complexity and takes about 0.08 milliseconds. This is because all operations are done locally and independently regardless of the size of the network. Aggregating all those readings, however, requires summing all readings from meters which results to an $O(N)$ time complexity. However, even with such overhead, the plot in Figure 5a shows that a 1000 users cluster have an overhead of around 3 milliseconds. Another performance concern is the generation of hash chains and random values after each m time slots. Measurements show that generating long chains for 10,000 time slots can be done within around 10 milliseconds. Also, this process does not happen very frequently. This means that it is suitable to run on different entities of the system.

E. COMMUNICATION OVERHEAD

Additional data traffic is typically needed in such communication systems. In LPPDA, all operations to encrypt and perturb the consumption readings do not change



(a)

(b)

FIGURE 5. Time performance overhead in 5a and Communication overhead in 5b.

TABLE 3. Performance overhead compared to previous work.

	LPPDA	DPPDA	NHP3
SM	$3T_{Add} + T_{Mul} + T_{SHA256}$	$2T_{PrExp} + 4T_{PrMul}$	$T_{PrExp} + T_{PrMul}$
A	$N(T_{SHA256} + T_{Add})$	$T_{PrExp} + (N+1)T_{PrMul}$	$(N+1)T_{PrExp} + (N+1)T_{PrMul}$
S	$(N+1)T_{Add}$	–	$T_{PrExp} + T_{PrMul}$

TABLE 4. Communication overhead, in KB, compared to previous work.

	LPPDA	DPPDA	NHP3
SM to A	32 N	256 N	3560 N
A to S	-	-	2048 + 160 + 2e5

TABLE 5. Measured operations times for performance evaluation.

Operation	Time	Operation
T_{SHA256}	1.51944 μs	SHA256 hash
T_{Noise}	14.62650 μs	Gamma noise generation
T_{Add}	0.80466 μs	Number addition
T_{Mul}	0.94748 μs	Number multiplication
T_{PrMul}	65.53411 μs	Paillier multiplication
T_{PrExp}	0.63488 ms	Paillier exponentiation

the size of data itself. The only overhead is the hash $H(r^i_{t(verify)}||Y^i_{t-1})$ sent from smart meters to the aggregator to be used for verification for Y^i_t. The value of this overhead depend on the hashing algorithm used to implement the protocol. Secure Hashing Algorithm (SHA256) is used in the system implemented in this work. This adds a 256 bits long hash with each reading. As shown in Figure 5b, the communication overhead is reasonably small. Sending readings from a cluster of size 1000 adds about 32KB of traffic.

F. OVERHEAD COMPARED TO RELATED WORK

Protocols implemented in DPPDA and NHP3 function in a distinct way to transmit smart meter readings to the supplier. Having a blockchain as in DPPDA would require a more powerful micro controller to be installed on each metering device. This is especially important for a mining node as it is responsible for mining the data and publishing it to the blockchain. As for NHP3, the nature of operations in the proposed cryptosystem can be computationally expensive.

Table 3 describes the performance overhead in terms of consumed time for the different parts of the grid. Table 4 shows the communication overhead from smart meters to aggregators and from the aggregators to the suppliers. Note that DPPDA does not have a normal aggregator. Instead, the protocol uses a "mining node" to publish the data to the blockchain. For this reason, the mining node is treated here as an aggregator.

We use the general term T_{op} to refer to the time required for operation *op*. The terminology used to formalize those equations is further described in Table 5, along with the time required for each operation. Those measurements were averaged over 1000 rounds for consistency. It is advised to refer to the compared protocols in [9] and [11] for more details related to the measured operations.

The performance overhead is presented in Figure 5a. Logarithmic scaling is used here as plot figures and growth rates vary widely. The plot shows that NHP3 has the highest performance overhead. It takes about 700 ms to process 1000 meters. DPPDA follows NHP3 with significant performance gain processing the same number of meters within around 67.8 ms. This is because DPPDA increases complexity based on N only as a multiple of a few Paillier multiplication operation, whereas NHP3 requires a multiple of Paillier exponentiation as well. The figure shows that our work has the least performance overhead. This matches the equations listed in Table 3 where our protocol scales in time with respect to basic hashing (T_{SHA256}) and addition (T_{Add}) operations.

The communication overhead is plotted in a similar fashion in Figure 5b. In terms of communication, DPPDA has less overhead than NHP3 as NHP3 needs to transmit more parameters and larger keys. Our work has the least overhead in communications as well because the only added overhead is the hashes used for verification.

VI. CONCLUSION

LPPDA presents a secure and privacy preserving protocol for smart meters. The protocol achieves differential privacy by adding distributed Laplacian noise locally on each smart meter. Random keys are used to hide readings from the aggregator and external adversaries while the aggregator hides individuals' data from the supplier. This combination allows efficient local differential privacy while keeping the protocol computationally minimal. Additionally, hash chains are used to verify the integrity of readings and authenticate each reading. This eliminates the need for a trusted third party to manage public keys while the protocol is running. Such third party is only needed once at the initialization phase, when random key seeds are distributed. All of these operations can be performed with minimal computational and communication overheads. Experimental results on two real-world datasets show that the performance of LPPDA compares favorable with existing state-of-the-art privacy preserving protocols. The proposed method can be applied to other similar applications such as microgrid monitoring, energy trading, and others for the sake of maintaining privacy.

ACKNOWLEDGMENT

The statements made herein are solely the responsibility of the authors.

REFERENCES

[1] S. S. Refaat, O. Ellabban, S. Bayhan, H. Abu-Rub, F. Blaabjerg, and M. M. Begovic, *Smart Grid Enabling Technologies*. Hoboken, NJ, USA: Wiley, 2021.

[2] E. Proedrou, "A comprehensive review of residential electricity load profile models," *IEEE Access*, vol. 9, pp. 12114–12133, 2021.

[3] C. Chalmers, P. Fergus, C. A. C. Montanez, S. Sikdar, F. Ball, and B. Kendall, "Detecting activities of daily living and routine behaviours in dementia patients living alone using smart meter load disaggregation," *IEEE Trans. Emerg. Topics Comput.*, vol. 10, no. 1, pp. 157–169, Jan. 2022.

[4] G. Tanoni, E. Principi, and S. Squartini, "Multilabel appliance classification with weakly labeled data for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 440–452, Jan. 2023.

[5] A. Brighente, M. Conti, D. Donadel, and F. Turrin, "EVScout2.0: Electric vehicle profiling through charging profile," 2021, *arXiv:2106.16016*.

[6] P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin, "Smart grid metering networks: A survey on security, privacy and open research issues," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2886–2927, 3rd Quart., 2019.

[7] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends® Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014.

[8] M. A. Husnoo, A. Anwar, R. K. Chakrabortty, R. Doss, and M. J. Ryan, "Differential privacy for IoT-enabled critical infrastructure: A comprehensive survey," *IEEE Access*, vol. 9, pp. 153276–153304, 2021.

[9] A. Mohammadali and M. S. Haghighi, "A privacy-preserving homomorphic scheme with multiple dimensions and fault tolerance for metering data aggregation in smart grid," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5212–5220, Nov. 2021.

[10] W. Lu, Z. Ren, J. Xu, and S. Chen, "Edge blockchain assisted lightweight privacy-preserving data aggregation for smart grid," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 2, pp. 1246–1259, Jun. 2021.

[11] H. Fan, Y. Liu, and Z. Zeng, "Decentralized privacy-preserving data aggregation scheme for smart grid based on blockchain," *Sensors*, vol. 20, no. 18, p. 5282, Sep. 2020.

[12] S. Zhang, Y. Zhang, and B. Wang, "Antiquantum privacy protection scheme in advanced metering infrastructure of smart grid based on consortium blockchain and RLWE," *IEEE Syst. J.*, vol. 17, no. 2, pp. 3036–3046, Jun. 2023.

[13] A. Alsharif, M. Nabil, S. Tonyali, H. Mohammed, M. Mahmoud, and K. Akkaya, "EPIC: Efficient privacy-preserving scheme with EtoE data integrity and authenticity for AMI networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3309–3321, Apr. 2019.

[14] N. Ravi, A. Scaglione, S. Kadam, R. Gentz, S. Peisert, B. Lunghino, E. Levijarvi, and A. Shumavon, "Differentially private K-means clustering applied to meter data analysis and synthesis," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4801–4814, Nov. 2022.

[15] H. Wang, J. Zhang, C. Lu, and C. Wu, "Privacy preserving in non-intrusive load monitoring: A differential privacy perspective," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2529–2543, May 2021.

[16] G. Ács and C. Castelluccia, "I have a DREAM!(DIffeRentially privatE smArt Metering)," in *Proc. Int. Workshop Inf. Hiding*. Cham, Switzerland: Springer, 2011, pp. 118–132.

[17] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz, and X. Wang, "A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid," *IEEE Trans. Comput.*, vol. 71, no. 11, pp. 2915–2926, Nov. 2022.

[18] N. Gai, K. Xue, B. Zhu, J. Yang, J. Liu, and D. He, "An efficient data aggregation scheme with local differential privacy in smart grid," *Digit. Commun. Netw.*, vol. 8, no. 3, pp. 333–342, Jun. 2022.

[19] F. Kserawi, S. Al-Marri, and Q. Malluhi, "Privacy-preserving fog aggregation of smart grid data using dynamic differentially-private data perturbation," *IEEE Access*, vol. 10, pp. 43159–43174, 2022.

[20] R. Zhou, Y. Xiang, Y. Wang, and X. Yan, "Non-intrusive identification and privacy-preserving of residential electric vehicle," *Energy Rep.*, vol. 8, pp. 1322–1329, Apr. 2022.

[21] L. Lamport, "Password authentication with insecure communication," *Commun. ACM*, vol. 24, no. 11, pp. 770–772, Nov. 1981.

[22] A. Perrig and J. Tygar, "TESLA broadcast authentication," in *Secure Broadcast Communication*. Cham, Switzerland: Springer, 2003, pp. 29–53.

[23] S. Goryczka and L. Xiong, "A comprehensive comparison of multiparty secure additions with differential privacy," *IEEE Trans. Dependable Secure Comput.*, vol. 14, no. 5, pp. 463–477, Sep. 2017.

[24] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120043–120065, 2021.

[25] Z. J. Lee, T. Li, and S. H. Low, "ACN-data: Analysis and applications of an open EV charging dataset," in *Proc. 10th ACM Int. Conf. Future Energy Syst.*, 2019, pp. 139–149.

NAHEEL FAISAL KAMAL received the B.S. and M.S. degrees in computer engineering from Qatar University, in 2019 and 2021, respectively. He is currently pursuing the Ph.D. degree with Texas A&M University. He worked on multiple research projects as a Research Assistant and as a Teaching Assistant with the College of Computer Engineering, Qatar University. He is a Research Graduate Assistant with Texas A&M University at Qatar. His research interests include the privacy and security of the smart grids and electric vehicle charging communication systems.

ABDULLA KHALID AL-ALI (Member, IEEE) received the master's degree in software design engineering and the Ph.D. degree in computer engineering from Northeastern University, Boston, MA, USA, in 2008 and 2014, respectively. He is currently the Head of the Computer Science and Engineering Department, College of Engineering, Qatar University, Doha, Qatar. He is an active Researcher in the field of communication, cyber–physical systems security, mHealth, and sensor networks. He has published a number of peer-reviewed papers in journals and conferences.

ABDULAZIZ AL-ALI received the Ph.D. degree from the University of Miami, Coral Gables, FL, USA, in 2016, with a concentration on machine learning. His previous research involved textual, image, and video-based data. He is currently an Assistant Professor with the Computer Science and Engineering Department, Qatar University. He is also taking the role of the Director of the KINDI Center of Computing Research, College of Engineering, Qatar University. He is also involved in diverse multi-disciplinary projects with industrial collaborators in the medical, security, and information retrieval fields. His research interests include machine learning and artificial intelligence. He is an Awardee of several research grants, of which are the National Priorities Research Program from the Qatar National Research Fund.

SERTAC BAYHAN (Senior Member, IEEE) received the bachelor's degree and the M.S. and Ph.D. degrees in electrical engineering from Gazi University, Ankara, Turkey, in 2008 and 2012, respectively. He is currently a Senior Scientist with the Qatar Environment and Energy Research Institute (QEERI). He is also a Faculty Member with the rank of an Associate Professor with the Sustainable Division, College of Science and Engineering, Hamad Bin Khalifa University. He has acquired U.S. $13M in research funding and published more than 170 papers in mostly prestigious IEEE journals and conferences. He is the coauthor of two books and five book chapters. He was a recipient of many prestigious international awards.

QUTAIBAH M. MALLUHI (Member, IEEE) received the B.S. and M.S. degrees in computer engineering from the King Fahd University of Petroleum and Minerals, Saudi Arabia, and the M.S. and Ph.D. degrees in computer science from the University of Louisiana at Lafayette. He was the Head of the Department, from 2006 to 2012, and the Director of the KINDI Center for Computing Research, Qatar University (QU), from 2012 to 2016. He was a Professor with Jackson State University. He was the Co-Founder and the CTO of Data Reliability Inc. He was a consultant for several telecommunication companies, where he built networks, designed distributed applications, and developed telecommunication management software. He is currently a Professor with the Department of Computer Science and Engineering, QU. He has received the QU Research Award, the JSU Technology Transfer Award, the JSU Faculty Excellence Award, and several best paper awards.