Privacy Preserving Fine-Grained Data Distribution Aggregation for Smart Grid AMI Networks

Enahoro Oriero* and Mohammad Ashiqur Rahman†
*Department of Electrical and Computer Engineering, Tennessee Technological University, Cookeville, USA
†Department of Computer Science, Tennessee Technological University, Cookeville, USA
Email: eoriero42@students.tntech.edu, marahman@tntech.edu

***Abstract*—An advanced metering infrastructure (AMI) allows real-time fine-grained monitoring of the energy consumption data of individual consumers. Collected metering data can be used for a multitude of applications. For example, energy demand forecasting, based on the reported fine-grained consumption, can help manage the near future energy production. However, fine-grained metering data reporting can lead to privacy concerns. It is, therefore, imperative that the utility company receives the fine-grained data needed to perform the intended demand response service, without learning any sensitive information about individual consumers. In this paper, we propose an anonymous privacy preserving fine-grained data aggregation scheme for AMI networks. In this scheme, the utility company receives only the distribution of the energy consumption by the consumers at different time slots. We leverage a network tree topology structure in which each smart meter randomly reports its energy consumption data to its parent smart meter (according to the tree). The parent node updates the consumption distribution and forwards the data to the utility company. Our analysis results show that the proposed scheme can preserve the privacy and security of individual consumers while guaranteeing the demand response service.**

***Index Terms*—Smart Grid; Advanced Metering Infrastructure; Privacy; Smart Metering; Fine-Grained; Distribution**

I. Introduction

The smart grid (SG) is the modernized version of the traditional electric grid. The SG upgrades electricity distribution and management by incorporating advanced two-way communication and pervasive computing capabilities for improved control, efficiency, reliability and safety [1]. The two-way communication allows the SG to collect users electricity consumption data in order to analyze the power generation, transmission, and consumption status in real-time.

Collecting energy consumption data is one of the most important processes of the smart grid. The utility company (UC) utilizes collected metering data to obtain useful real time information about the SG and to prepare the energy consumption bill for the users. To obtain real-time electricity consumption data, smart meters (SMs) on the consumers side usually record and report these data periodically, e.g., every 15 minutes. A full measurement and data collection system called the advanced metering infrastructure (AMI) replaces traditional electromechanical meters. The AMI allows for data aggregation where intermediate nodes relay packets on behalf of other smart meters to the UC, through a communication network. This enables the UC to integrate demand-side resources into the wholesale market, and hence, the demand response (DR) [2]. DR allows generators and loads to interact in an automated fashion in real-time, coordinating demand to flatten consumption spikes. Eliminating the demand that occurs during these spikes eliminates the cost of adding reserve generators, which prolongs the lifespan of equipment and reduces consumers' bills.

The granularity of the reported metering data (every 15 minutes as compared to every 24 hours) provides room for more detailed statistical analysis. This high-frequency data necessitates the need for privacy against malicious and curious adversaries. Careful analysis of the metering data can allow an adversary to categorize the data with alarmingly high accuracy. From the consumers' perspective, privacy preservation is important because reported energy consumption data can expose private information about the consumers habits and real-time daily activities as each electrical appliance has its unique load signature (i.e., energy consumption pattern).

In this paper, we leverage a tree topology structure to model the communication network, in which each node of the tree represents an SM, and each SM is connected to a parent node, which can also be a child to another parent node. This structure ensures each child node records and reports metering data periodically to the parent node. The parent node links between child nodes and the UC by forwarding the metering data to the UC. The control centers of the UC use the aggregated energy consumption data to support the pricing and decision-making. Utilizing this tree communication structure, we present an anonymous fine-grained data distribution aggregation scheme in which the UC learns only the distribution of the aggregated total energy consumption for a specific time slot.

Fig. 1 shows a bar chart representing the distribution of energy consumption for a time slot or period (i.e., 1 p.m. to 2 p.m.). Each bar represents the energy consumption in KW by a specific number of users. As can be observed from the chart, the distribution shows that 5 users consumed 50 KW of energy, 8 users consumed 70 KW and 20 users consumed 120 KW. Similarly, 10 users consumed 100 KW and 25 users consumed 150 KW for the considered time period. In this research, we assume a DR application in which the UC utilizes the distributed aggregated data for short-term load forecasting (STLF), which can be used to advise customers on the best time to use energy based on predicted price. In summary, our main contributions are as follows:

- We propose a privacy preserving fine-grained data dis-

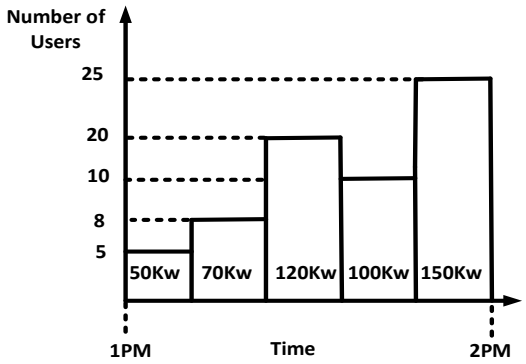


Fig. 1: Distribution of users according to their energy usage during a specific time slot.

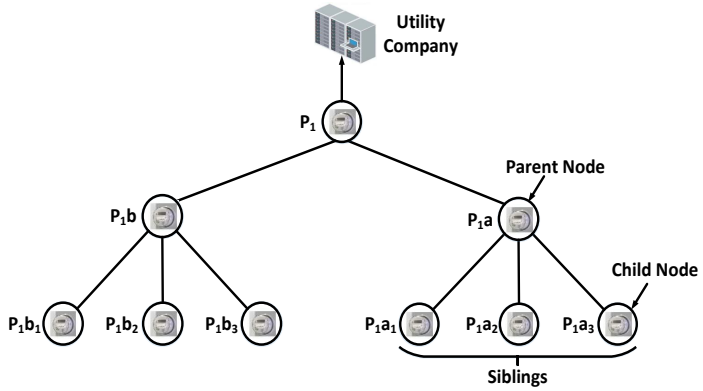


Fig. 2: The network architecture for reporting energy consumption data.

tribution aggregation scheme in which the UC obtains only the distribution of the total energy consumption of its users for a specific time period.

- We design an anonymous privacy preserving random collaborative forwarding protocol that leverages a tree topology structure for secure meter data reporting.
- UC utilizes fine-grained data for short-term load forecasting to predict price for future energy demand in order to advise customers how to adjust their consumption patterns according to time-varying electricity prices.

The rest of the paper is organized as follows. The network and threat model is discussed in Section II. Some preliminaries related to the proposed scheme are presented in section III. In Section IV, we elaborate on the proposed privacy preserving fine-grained data distribution aggregation scheme. In section V, we analyze some characteristics of the proposed scheme and we present the energy forecasting and price prediction scheme. The performance evaluation is discussed in Section VI. The related work is presented in section VII. Finally, we present the conclusion in Section VIII.

II. Network and Threat Model

The network model is made up of SMs, UC and an offline-trusted authority (TA). Each consumer in a residential neighborhood has an SM in his/her apartment. Each parent SM and its children node form a group. Therefore, as the tree topology grows, the groups become bigger with grand-parent nodes belonging to multiple sub-groups consisting of various children (parent) nodes. Fig. 2 illustrates the network model – a hierarchal tree topology. As shown in the figure, parent node P_1a has three children nodes (i.e., P_1a_1, P_1a_2 and P_1a_3). Similarly, node P_1 is the parent of nodes P_1a and P_1b. The SMs are connected via a wireless mesh network using Wi-Fi or ZigBee. Siblings can communicate directly amongst themselves and also with the parent SM. The SMs cannot communicate directly with the UC but can report their energy consumption data via the parent SM to the UC according to the network tree topology.

SMs randomly report metering data to the UC periodically and the parent SM links between the child node and the UC. The data received by the UC is the distribution of the energy usage that is aggregated from the metering data reported by all users for a specific time period. This ensures privacy of the individual SM (consumer) because the UC can not isolate a specific reading for a particular SM. However, an exceptional case in which only one parent SM and one child node exist in a group is also possible. This can compromise privacy because the reported energy data always come from the same child node. Thus, the parent node can learn sensitive information about the child node. In this paper, we assume that each parent node has at least two child nodes. The TA is responsible for computing and distributing the group secret and public keys to users. A certificate that binds the groups identity to its public key is also distributed.

For the threat model, we assume an honest-but-curious one, in which all parties follow the protocol properly. That is, nodes are curious about the energy usage data, from which they may try to infer sensitive information about the households. However, each SM forwards received packets from other nodes according to the protocol and does not intentionally drops or distorts any value or intermediate result to disrupt data aggregation for malicious purposes.

III. Preliminaries

In this section, we first provide some background information related to the techniques used in the proposed scheme.

A. Bloom Filter

A Bloom Filter is a space efficient probabilistic data structure that is used to determine time and memory efficiently whether an element is a member of a set. The result of the Bloom Filter search query can either be "possibly in set" or "definitely not in set," which means it can have false positives but never false negatives. Thus, the Bloom Filter has a 100% recall rate [3]. The Bloom Filter has an array of m bits representing a set $S = \{x_1, x_2, \ldots, x_n\}$ of n elements. Initially, all the m bits are zeroes. A k number of independent hash functions $h_1, h_2, \ldots, h_k$ are used to map data to be stored in a particular location of the Bloom Filter. However, it can also be possible that an element that was not added to the Bloom Filter, by coincidence, each of its bit locations in the

vector is set (i.e., true). This comes from the fact that a bit can be set multiple times when the hashes of different elements point to the same location. This situation is called a false positive. The probability of this false-positive in a Bloom Filter is given by the following equation:

$$Pf = (1 - (1 - (\frac{1}{m})^{kn}))^k \quad (1)$$

Here, n represents the number of elements in the vector, m represents the size of the Bloom Filter, and k is the number of unique hash functions used to add elements.

In our scheme, we use the Bloom Filter to store the serial number of the group certificate. Since each parent node forms a group with its children node, the Bloom Filter is used to cope with the increasing storage overhead of the tree topology.

B. Bilinear Groups

The bilinear map notation is defined as follows:

1) G_1 and G_2 are two (multiplicative) cyclic groups of prime order p.
2) g_1 is a generator of G_1 and g_2 is a generator of G_2.
3) ψ is a computable isomorphism from G_2 to G_1 by the isomorphism function $\psi(g_2) = g_1$.
4) e is an efficiently computable bilinear map $e : G_1 \times G_2 \rightarrow G_T$ with the following properties:
   - Bilinear: for all $u \in G_1$, $v \in G_2$ and a, $b \in Z_p*$, $e(u^a, v^b) = e(u, v)^{ab}$.
   - Non-degenerate: e(g_1, g_2) $\neq 1$. In essence, the map does not send all pairs in $G_1 \times G_2$ to the identity in G_T.

C. VLR Group Signature

We utilize the concept of verifier-local revocation (VLR) group signatures introduced by Boneh et al [4]. Three algorithms named *KeyGen*, *Sign*, and *Verify* are used to generate the signature for each group.

KeyGen (n): The algorithm uses the given parameter n (the number of members in the group) to generate a group public key gpk as well as a private key gsk_i and a revocation token grt_i for each user i.

Sign (gpk, gsk_i, M): The algorithm takes as input the group public key gpk, a private key gsk_i, and a message $M \in \{0,1\}^*$ and returns a signature σ.

Verify (gpk, RL, σ, M): The verification algorithm takes several parameters: the group public key gpk, a set of revocation tokens RL (whose elements form a subset of the elements of grt_is), and an alleged signature σ on a message M. It returns either "valid" or "invalid". The "invalid" response specifies that the signature is either invalid or the user is revoked.

IV. Proposed Scheme

In this section, we describe the proposed privacy-preserving metering data distribution reporting scheme in detail. We first present the initialization phase in which an offline TA generates the group signature for each group and distributes cryptographic credentials (public and private keys) to each node. This phase should be run only once when the application is first launched. If a node is revoked, TA updates the secret keys and constructs another signature for that group. Then, we proceed to explain our privacy-preserving metering data reporting scheme in which each parent SM aggregates the distribution of energy consumption for its group. Finally, the UC obtains the total distribution of the reported metering data for the particular time period, which is used to predict future energy prices and inform (or advise) customers.

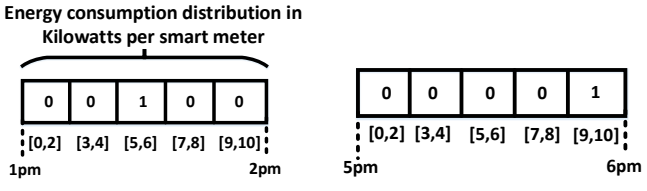
A. Preliminaries

Each SM registers with the TA, which assigns the SM to a group. The TA also distributes the corresponding key materials and group certificate to the SM. Each parent node creates a Bloom Filter containing the serial numbers of all the group certificates the parent node belongs to after receiving the group certificate. The TA runs the key generation algorithm (*KeyGen (n)*) to generate the cryptographic key materials. The algorithm uses an input n that represents the number of user keys to generate. It selects a generator $g_2 \in G_2$ uniformly at random, and a set $g_1 \leftarrow \psi(g_2)$. An element γ is randomly selected from Z_p*, where Z_p* is a finite field of order p. γ is known only to the TA. It generates $w = g_2{}^{\gamma}$. Using γ, (A_i, x_i) is generated for each SM by selecting x_i randomly from Z_p* such that $\gamma + x_i \neq 0$ and setting $A_i \leftarrow g_1{}^{\frac{1}{x_i}}$. Each SM's private key is gsk_i = (A_i, x_i). The group public key is gpk = (g_1, g_2, w). The revocation token corresponding to an SM's key (A_i, x_i) is grt_i = A_i. The total output of the key generation algorithm is (gpk, gsk, grt).

B. Fine-Grained Data Distribution Aggregation Protocol

Each SM reports its energy consumption data via a bit vector to its parent node according to the tree topology. The parent node correspondingly forwards this data to the next parent node or, if it is at the peak of the tree, the data is forwarded directly to the UC. Each entry of the bit vector represents a range of energy consumption per time period. Fig. 5 shows an example of the energy consumption vector for two different time slots, 1–2 p.m. and 5–6 p.m. As shown in the figure, the row represents the consumed energy distribution in kilowatts, where each column per row corresponds to the specific consumption range for that time slot. In the first vector of Fig. 3(a), the distribution of the energy consumption for time slot 1–2 p.m. is in the 5–6 KW range for that household. Similarly, Fig. 3(b) represents the range of the energy consumption distribution between 5 and 6 p.m. (9–10 KW). To report energy consumption, each SM signs the data with its private key according to a digital signature algorithm by using the *sign* algorithm.

The signature algorithm uses a group public key $gpk = (g_1, g_2, w)$, an SM's private key $gsk_i = (A_i, x_i)$, and a message $M \in \{0,1\}$ to generate a signature σ. A random nonce r is selected from $Z_p{}^*$. Generators (û, v̂) $\leftarrow H_0$ (gpk, M, r) $\in {G_2}^2$, where H_0 is an one-way hash function that maps $\{0, 1\}^* \rightarrow {G_2}^2$. The images $u, v \in G_1$ are computed from $u \leftarrow \psi(\hat{u})$ and $v \leftarrow \psi(\hat{v})$. An exponent $\alpha \in Z_p*$ is selected to compute $T_1 \leftarrow u^{\alpha}$ and $T_2 \leftarrow A_i v^{\alpha}$. δ is computed



(a) (b)

Fig. 3: Energy consumption vector

as $\delta \leftarrow x_i\alpha$. Blinding values r_α, r_x, and r_δ are randomly selected from Z_p* to compute helper values R_1, R_2 and R_3 as shown below:

$$R_1 \leftarrow u^{r_\alpha}$$
$$R_2 \leftarrow e(T_2, g_2)^{r_x} \cdot e(v, w)^{-r_\alpha} \cdot e(v, g_2)^{-r_\delta}$$
$$R_3 \leftarrow T_1^{r_x} \cdot u^{-r_\delta}$$

A challenge value c is computed using H where H is a one way hash function that maps $\{0, 1\}^* \rightarrow Z_p*$. $c \leftarrow H(gpk, M, r, T_1, T_2, R_1, R_2, R_3)$. s_α, s_x and s_δ are computed as $s_\alpha = r_\alpha + c\alpha$, $s_x = r_x + cx_i$, and $s_\delta = r_\delta + c\delta$. Finally, the signature σ is generated where $\sigma \leftarrow (r, T_1, T_2, c, s_\alpha, s_x, s_\delta)$. The SMs using the group signature (σ) and group certificate report their energy consumption to the parent node. To forward the packet to the parent node (U_m), an SM (U_i) randomly selects a sibling SM (U_j), which acts as a link to help forward the energy consumption to the parent node (U_m). Here, U_m is the parent of both U_i and U_j. U_i encrypts the metering data packet M_i using a symmetric key shared with U_j (Ek_{ij}).

$$M_i \leftarrow Ek_{ij}(t \| EDT_i)$$
$$EDT_i \leftarrow Epk_m(t \| V_i \| H(t) \| \sigma)$$

Here, V_i represents the energy consumption vector and t is the time stamp for the reported interval. Epk_m is the public key of parent node U_m and H is a secure one-way hash function. U_j decrypts the packet with the symmetric key (Ek_{ij}) and checks the time-stamp to determine whether it is the current time interval or if it is an old or replayed packet. If the packet is old (or replayed), the packet is dropped. If it is the current time interval, it proceeds to forwarding the packet either to the next available node (U_k) which is another sibling or to the parent node (U_m). If it is forwarded to U_k, it is similarly encrypted with a symmetric key (Ek_{jk}) that is shared between U_j and U_k. If U_k, is closest to the parent node, it first checks the time-stamp of the packet to verify the time interval, as described before. If it is the current time interval, it decrypts it with the shared symmetric key and forwards it to U_m.

On receiving the packet , U_m first decrypts it with its private key and proceeds to computing the hash of the time interval. The computed hash output is compared with the sent hash to verify that it tallies with the time interval of the reported energy consumption vector. It proceeds to authenticate the group certificate by checking the Bloom Filter to see if the serial number is added. However, Bloom Filters suffer from false positives, where a search may mistakenly indicate that an

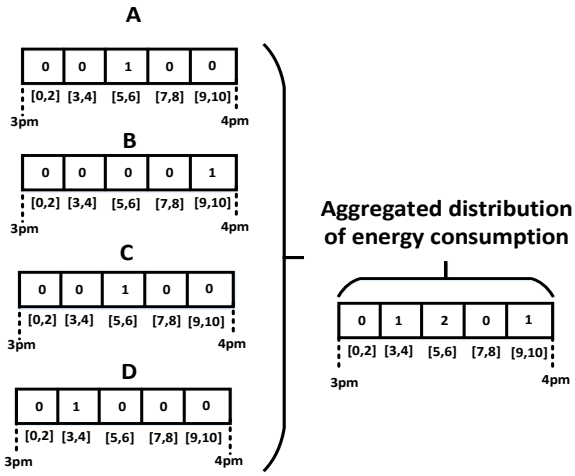


Fig. 4: Aggregated energy consumption vector.

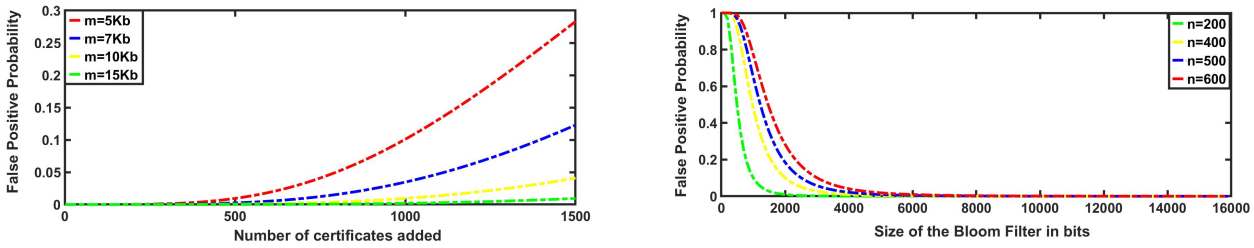
invalid certificate is in the list when it was not originally added. This is not acceptable in AMI networks because meters may authenticate malicious users who may proceed to carry out more harmful actions. In subsequent sections, we show how we managed this challenge by carefully designing the Bloom Filter to reduce the false positive to a very low acceptable percentage. The group certificate provides a first level of authentication before proceeding to verify the group signature (second-level authentication). The two-level authentication is required to cope with the overhead of the complex tree topology, because at higher nodes in the tree structure, more authentication is required. Thus, if an attacker sends a fake group certificate, the first level of authentication via the Bloom Filter can help identify such malicious packets.

The k hash functions used in creating the Bloom Filter are similarly used to hash the serial number of the certificate. If the same bits are set, the parent node proceeds to verify the group signature by using the *Verify* algorithm. The verification algorithm verifies the signature σ using a group public key gpk = (g_1, g_2,w), a set RL of revocation tokens, signature σ = (r, T_1, T_2, c, s_α, s_x, s_δ), and a message $M \in \{0, 1\}^*$. To verify the signature, $\hat{u}$ and $\hat{v}$ and their images u and v are first computed. $u \leftarrow \psi(\hat{u})$, $v \leftarrow \psi(\hat{v})$. R'_1, R'_2, and R'_3 are computed as:

$$R'_1 \leftarrow u^{s_\alpha}/T_1^c$$
$$R'_2 \leftarrow (T_2, g_2)^{s_x} e(v, w)^{-s_\alpha} e(v, g_2)^{-s_\delta} \cdot (e(T_2, w)/e(g_1, g_2))^C$$
$$R'_3 \leftarrow T_1^{s_x} u^{-s_\delta}$$

The verification is done by checking $c \stackrel{?}{=} H(gpk, M, r, T_1, T_2, R'_1, R'_2, R'_3)$. If it is true, the signature is valid; otherwise, the signature is discarded. To ensure that the signature is not from a revoked user, a revocation check is done by determining whether $e(T_2/A, \hat{u}) \stackrel{?}{=} e(T_1, \hat{v})$. If successful, the user is not revoked.

The group signature provides the added advantage of hiding the identity of the signer as it can be any member (SM) in the group. If the authentication steps described above fail, the packet is immediately dropped; otherwise, the parent node



(a) (b)

Fig. 5: False positive probability by varying (a) the Bloom Filter size and (b) the number of items.

proceeds to aggregate all the received energy consumption vectors for that time slot. The parent node aggregates the energy consumption for a particular time slot by adding all the values corresponding to the individual bit vectors. This is shown in Fig. 4 where the distribution of the total energy consumed during 3–4 p.m. by four users is shown in the total aggregated bit vector. The aggregated result shows that between 3 and 4 p.m., one user consumed energy in the range of 3–4 KW. Similarly, another user consumed energy in the range of 9–10 KW, while two users consumed energy in the range of 5–6 KW. Thus, this ensures anonymity of individual users because the utility only receives the distribution of the total aggregated energy consumption.

V. Characteristic Analysis

We discuss two major characteristics of the proposed scheme in this section.

A. Bloom Filter Characteristic

In our scheme, we used the Bloom Filter to store the serial number of each group certificate. However, Bloom Filters suffer from false positives, which occurs when a search on a Bloom Filter indicates that a serial number exists when it actually does not. This negatively affects the performance of our system. Therefore, it is important to carefully design the Bloom Filter to cope with this challenge by reducing the false positive probability to a very low (tolerable) value. We managed this challenge by adjusting two key parameters that affect the false positive probability. As can be observed from the false positive probability equation in Equation (1), the size of the Bloom Filter and the number of added items determine the false positive probability of the Bloom Filter.

We analyzed the effect of the size and number of items on the Bloom Filter, and Fig. 5(a) shows a plot of the Bloom Filter with different sizes. As shown in the figure, increasing the size of the Bloom Filter correspondingly reduces the false positive probability of the Bloom Filter for a fixed number of items. Fig. 5(b) illustrates the effect of increasing the number of added items on the Bloom Filter. As seen in the plot, increasing the number of items correspondingly increases the false positive probability for a fixed size. To maintain a low false positive probability (0.02%) rate, we kept the size of the Bloom Filter to 15 KB and the number of added items to 500. If more than 500 certificates are required, we propose using multiple Bloom Filters to ensure the false positive probability is always within the acceptable range.

B. Energy Forecasting and Price Prediction

One of the most common types of energy forecasting is Short-Term Load Forecasting (STLF) [5], which aims to predict consumption for short time frames, even hourly. Several short-term load forecasting techniques have been proposed, and an overview of the different STLF techniques available can be found in [6]. Moreover, the various STLF techniques can be employed by the grid operator to support resource management on a short-term basis to benefit other entities, such as network operators, transmission operators, and market operators can benefit from. In our scheme, the UC utilizes the obtained distribution of the fine-grained energy consumption data for STLF. By regularly performing the STLF, the UC can predict future energy price based on the frequently collected fine-grained distribution data. The UC broadcasts prediction results to all SMs providing advice to customers on how to save money (best time to use energy) and how to determine optimum usage. Thus, this DR scheme encourages consumers to use less energy during peak hours, or to move the time of energy consumption to off-peak periods based on received energy price forecast information.

VI. Performance Evaluation

We evaluate the proposed scheme with respect to the security and privacy and the incurred communication and computation overhead.

A. Security and Privacy Analysis

Privacy: The UC receives only the distribution of energy consumption data which preserves the privacy of each individual consumption. Also, since energy consumption is randomly reported via the tree topology, it is difficult to identify the SM that generated the data.

Authenticity and Integrity: Signing with the group signature preserves the integrity of the reported energy data because only legitimate nodes that have been issued the correct private keys can generate the signature.

Data Confidentiality: External eavesdroppers and attackers who are interested in learning sensitive information from reported metering data cannot obtain any useful information even if the packets are intercepted because the packets are encrypted with both symmetric and asymmetric keys.

B. Communication Overhead

The communication overhead is measured by the amount of data (in bytes) transmitted in each packet. G_1, G_2, and Z_p* have order p, which is 32 bytes. The elements of Z_p* have 32 bytes but the elements of G_1 and G_2 have 64 bytes, using elliptic curve cryptography. The group signature in our scheme described previously comprises two elements of G_1 and five elements of Z_p*. Thus, the total group signature length is 128 bytes + 160 bytes = 288 bytes. The size of the symmetric key (E_{kij}) is 64 bytes/block. Therefore, packet M has a size of 352 bytes $\times$ the number of blocks. The size of the Bloom Filter is fixed at 15 KB.

C. Computation Overhead

The computation overhead is measured by the time required to perform an operation in milliseconds (ms). We measured the computation times of the multiplication, pairing, and exponentiation operations using MIRACL cryptographic library [7] running on a 2 GHz Intel processor and 4 GB RAM. Our measurements indicate that the multiplication, exponentiation and pairing operations take 0.65, 2.4, and 7 ms, respectively. Using these measurements, the computation of T_1 requires 2.4 ms, T_2 requires $3.05ms$, R_1 requires 2.4 ms, R_2 requires 28.2 ms, and R_3 requires 5.45 ms. Thus, the overall time required for signature generation is 41.5 ms. Similarly, signature verification requires 46.55 ms. AES symmetric key encryption/decryption requires 0.5 ms/block.

VII. Related Work

Efthymiou et al. [8] proposed a trusted third party escrow mechanism for anonymizing frequent metering data sent by an SM without the possibility of linking the metering data to the consumers. In [9], Dimitriou et al. proposed a semi-trusted third-party scheme composed of multiple aggregators for sending the aggregated energy consumption to the UC. A secret sharing mechanism is leveraged to achieve anonymity. Ford et al. [10] introduced an anonymization approach that consists of a three-tier model (smart meters, the utility company, and a trusted third party) to manage a cloud-based storage system for secure smart meter communication and privacy preservation. Ambrosin et al. [11] proposed an anonymization technique for metering data transmission via a random multi-hop path. A collaborative protocol among SMs is used to ensure unlinkability between the raw metering data and the source SM. Unlike our proposed work, the authors did not consider the distribution of the energy consumption per time period, rather they focused on aggregating the energy consumption data only. A different approach to aggregate metering data via the network topology is considered by Lyu et al. [12]. The intermediate fog nodes periodically collect and aggregate data from the connected SMs and report to the utility company. The cloud/utility supplier computes the overall data aggregation by aggregating the intermediate Fog level aggregation of the energy consumption data.

A majority of the existing solutions either require a trusted third-party entity that actively participates in the protocol execution, which introduces a single point of failure into the system, or incur heavy computational and communication overhead, which hinders implementation. Also, they mainly focus on aggregating the energy consumption data and not the distribution of the energy consumption data. Hence, the aforementioned shortcomings complicate existing solutions. Based on the above observations, we propose a privacy preserving fine-grained data distribution aggregation protocol relying on a collaborative mechanism among SMs according to a predefined network tree topology.

VIII. Conclusion

Consumer privacy in fine-grained SM data aggregation is a major problem facing smart grid deployment. To address this challenge, we proposed an anonymous privacy preserving fine-grained data distribution aggregation scheme for smart grid AMI networks in which the utility company receives only the distribution of the aggregated energy consumption data of different time periods. We utilized a random collaborative forwarding protocol to anonymously transmit energy consumption data for each time period. In order to efficiently store and reduce computation overhead, we used Bloom Filters to store the group certificates. Our evaluations and analysis have demonstrated that our scheme can provide security/privacy, and the use of Bloom Filters can make the scheme scalable, with minimum computation and communication overhead. For future work, we plan to consider a dynamic network model.

References

[1] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. "A survey on smart grid communication infrastructures: Motivations, requirements and challenges." IEEE communications surveys & tutorials 15, no. 1 (2013): 5-20.

[2] Y. Gong, Y. Cai, Y. Guo, and Y. Fang. "A privacy-preserving scheme for incentive-based demand response in the smart grid." IEEE Transactions on Smart Grid 7, no. 3 (2016): 1304-1313.

[3] Song, Haoyu, Sarang Dharmapurikar, Jonathan Turner, and John Lockwood. "Fast hash table lookup using extended Bloom Filter: an aid to network processing." ACM SIGCOMM Computer Communication Review 35, no. 4 (2005): 181-192.

[4] Boneh, Dan, and Hovav Shacham. "Group signatures with verifier-local revocation." In Proceedings of the 11th ACM conference on Computer and communications security, pp. 168-177. ACM, 2004.

[5] Gross, George, and Francisco D. Galiana. "Short-term load forecasting." Proceedings of the IEEE 75, no. 12 (1987): 1558-1573.

[6] Gerwig, Carola. "Short term load forecasting for residential buildingsAn extensive literature review." In Intelligent Decision Technologies, pp. 181-193. Springer, Cham, 2015.

[7] Miracl, Multiprecision integer and rational arithmetic c/c++ library.

[8] Efthymiou, Costas, and Georgios Kalogridis. "Smart grid privacy via anonymization of smart metering data." In First IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 238-243, 2010.

[9] Dimitriou, Tassos, and Ghassan Karame. "Privacy-friendly tasking and trading of energy in smart grids." In 28th Annual ACM Symposium on Applied Computing, pp. 652-659, 2013.

[10] Ford, Vitaly, Ambareen Siraj, and Mohammad Ashiqur Rahman. "Secure and efficient protection of consumer privacy in Advanced Metering Infrastructure supporting fine-grained data analysis." Journal of Computer and System Sciences 83, no. 1 (2017): 84-100.

[11] M. Ambrosin, H. Hosseini, K. Mandal, M. Conti, and R. Poovendran. "Despicable me (ter): Anonymous and fine-grained metering data reporting with dishonest meters." In IEEE Conference on Communications and Network Security (CNS), pp. 163-171, 2016.

[12] Lyu, Lingjuan, Karthik Nandakumar, Benjamin Rubinstein, Jiong Jin, Justin Bedo, and Marimuthu Palaniswami. "PPFA: Privacy Preserving Fog-enabled Aggregation in Smart Grid." IEEE Transactions on Industrial Informatics (2018).