
Article:
Gope, P. orcid.org/0000-0003-2786-0273 and Sikdar, B. (2019) Lightweight and privacy-friendly spatial data aggregation for secure power supply and demand management in smart grids. IEEE Transactions on Information Forensics and Security, 14 (6). pp. 1554-1566. ISSN 1556-6013

https://doi.org/10.1109/tifs.2018.2881730

Lightweight and Privacy-Friendly Spatial Data Aggregation for Secure Power Supply and Demand Management in Smart Grids

Prosanta Gope *Member, IEEE* and Biplab Sikdar, *Senior Member, IEEE*

***Abstract*—The concept of smart metering allows real-time measurement of power demand which in turn is expected to result in more efficient energy use and better load balancing. However, finely granular measurements reported by smart meters can lead to starkly increased exposure of sensitive information, including various personal attributes and activities. Even though several security solutions have been proposed in recent years to address this issue, most of the existing solutions are based on public-key cryptographic primitives such as homomorphic encryption, elliptic curve digital signature algorithms (ECDSA), etc. which are ill-suited for the resource constrained smart meters. On the other hand, to address the computational inefficiency issue, some masking-based solutions have been proposed. However, these schemes cannot ensure some of the imperative security properties such as consumer's privacy, sender authentication, etc. In this paper, we first propose a lightweight and privacy-friendly masking-based spatial data aggregation scheme for secure forecasting of power demand in smart grids. Our scheme only uses lightweight cryptographic primitives such as hash functions, exclusive-OR operations, etc. Subsequently, we propose a secure billing solution for smart grids. As compared to existing solutions, our scheme is simple and can ensure better privacy protection and computational efficiency, which are essential for smart grids.**

***Index Terms*—Privacy, spatial data aggregation, smart grids**

I. Introduction

Smart grids are expected to enhance the efficiency of current power grids by using advanced digital information and communication technology. The combined volatility of both power supply and power demand is a growing problem that needs to be solved by the smart grids. Unlike water or gas, electricity is not easily or economically storeable in large quantities with current technologies. Therefore power grids are required to maintain a balance between power production and demand across short time scales. To ensure smart load balancing between production and demand, the deployment of smart meters is being pursued by many countries. The smart meters measure and report power consumption on a regular basis. This feature facilitates better power consumption monitoring, control and prediction, which in turn results in cost savings to both the power suppliers and consumers, as well as an immense reduction in the carbon dioxide emissions to the atmosphere [1]. To achieve these objectives, grid operators require effective *spatial data aggregation schemes*, where an aggregator periodically aggregates the electricity consumption of a group of consumers in a geographical region, each equipped with a smart meter. This consumption data may be used by the utility to dynamically update its prices in order to implement demand-side management. In addition, the aggregated consumption data may be used for supply-demand management, for example, by ramping production up or down as needed. These management features are particularly important in the context of increasing penetration of renewable energy sources (such as solar panels and wind turbines) in power grids across the world, given their inherent variability.

While smart meters offer some clear benefits, accurate and fine-grained measurements of household energy consumption trigger serious privacy concerns [2], [3]. In this regard, fine-grained smart meter data may reveal an user's presence/absence in his/her house, which electrical appliances they are using at any moment, or even their daily habits at home. Since the usage of smart meters is essential for better supply and demand management in smart grids, it is important to develop technologies that reconcile privacy with the desired utility and functionality of smart meters.

A. Related Work

Over the last decade, some interesting data aggregation schemes have been proposed under various settings (e.g., smart metering systems [4-25], and [34-38], vehicle-to-grid networks [39], and wireless sensors networks [40]). To tackle the privacy issues in smart grids, a number of research results have been proposed in recent years. These can be divided into two categories: public-key-based (such as homomorphic encryption based) schemes and masking-based schemes. We first consider the existing public-key-based schemes followed by the existing masking-based schemes, and elaborate on their strengths and weaknesses. In 2010, Garcia et al. [4] proposed a multi-party computation protocol that allows a number of smart meters in a locality to compute a partial aggregation of their data without revealing their individual measurements by taking advantage of Paillier homomorphic encryption [5]. However, this scheme lacks efficiency in terms of computation and communication overheads. In addition, due to the malleability of homomorphic encryption, this scheme is also vulnerable to data forgery attacks. Similarly, Lu et al. designed a privacy-preserving data aggregation protocol [6] using the Paillier homomorphic

P. Gope, is with Department of Computer Science, University of Hull, Cottingham Rd, Hull HU6 7RX, United Kingdom. (E-mail: prosanta.nitdgp@gmail.com/p.gope@hull.ac.uk)

B. Sikdar is with Department of Electrical and Computer Engineering, National University of Singapore, 21 Lower Kent Ridge Rd, Singapore 119077. (Email: bsikdar@nus.edu.sg)

Corresponding author: Prosanta Gope

cryptosystem, which also causes higher computation overhead on entities like smart meters. Liang et al. proposed a usage-based dynamic pricing scheme for smart grids [7] by using a fully homomorphic technique devised by Naehring et al. [8]. Fully homomorphic techniques are difficult to implement with current computing resources which limits the practical applicability of this scheme. Chia-Mu et al. introduced a ring signature based scheme to protect users' usage profile [9]. However, the computational cost increases with the size of the ring. Liu et al. have proposed an aggregation scheme based on blind signatures [10]. However, this scheme does not protect the privacy of the consumer's usage data profile. Also, Zhang et al. have proposed a self-certified signature scheme without considering usage data integrity and consumer's privacy [11]. Sui et al. have designed an incentive-based data aggregation scheme which is constructed with the assumption of an anonymity network, where the sources of usage reports are anonymous [12]. Therefore, it is hard to identify any smart meter or any communication failure. Besides, in this scheme the aggregator is unable to verify the legitimacy of the smart meter and the usage data, which may lead to forgery attacks. Li et al. introduced a hop-by-hop technique for data aggregation in smart grids [13], [14]. However, aggregation in the presence of node failures as well as the methodology to construct the aggregation tree have not been explicitly addressed. In addition, as shown in [15], the schemes presented in [13], [14] cannot ensure privacy during smart metering and reveal the identity of the users. Also, the public key signatures used in these schemes result in higher computation cost. A few more homomorphic encryption based schemes have been proposed in [15]-[19]. However, in these schemes, the smart meter is not authenticated during data aggregation. Thus, a dishonest or fake smart meter may falsify the data, leading to inaccurate aggregation result. Besides, these schemes does not support consumer's privacy. Also, the data aggregation schemes presented in [17] and [18] do not ensure data integrity. In [19] Jo et al. proposed two data aggregation schemes based on Paillier homomorphic encryption and elliptic curve digital signature algorithms (ECDSA), respectively. While their ECDSA-based scheme can ensure sender authentication, the usage report transmitted by each smart meter SM_i reveals it's identity ID_{SM_i}, which is fixed for all transactions. Therefore, an adversary can easily identify if the usage data is from the same consumer's end and link ID_{SM_i} to an actual user. Thus, the scheme presented in [19] cannot ensure anonymity of a consumer. Apart from the schemes above, a few more public-key-based data aggregation protocols have been introduced in recent years [30], [31]. In [30] a discrete logarithm problem (DLP)-based data aggregation scheme is introduced, in which the authors allows a substation to access private data using a shared key. Hence, this scheme cannot ensure strong privacy. Abdullah et al. proposed a lattice based homomorphic data aggregation scheme [31]. However, lattice-based encryption systems incur a huge computational and communication cost. Besides, the scheme presented in [31] does not support individual sender authentication and consumer's privacy features. Hence, the security solution in [31] is not suitable for smart grids. Koo et al. [34] investigated some of the state-of-the-art data aggregation schemes in smart grids, where they found that most of the existing solutions cannot ensure authenticity of the metering data and consumer's privacy. Vahedi et al. [35] recently proposed a privacy preserving data aggregation scheme for smart grids using ECDSA. Even though their scheme can guarantee authentication of the source and integrity of the usage data. However, in their scheme consumers reveal their identity, and hence cannot ensure consumer's privacy. The solution in [36] preserves the usage privacy thanks to the property of bilinear pairings. However, this solution incurs high communication and computational overheads due to the generation and distribution of keys and of the encrypted measurements of each involved entity. Additional overheads come from random public parameters that should be signed to ensure their integrity and authenticity.

Table I
NOTATIONS AND CRYPTOGRAPHIC FUNCTIONS

Symbol	Definition
PS	Power Supplier
HAN	Home area network
SM	Smart meter
TPA	Third-party aggregator
PID_i	Pseudo identity of SM_i
TID_i	Temporary identity of SM_i
ID_{SM_i}	Identity of smart meter SM_i
ID_A	Identity of the TPA
k_i	Secret key of the SM_i
K_{as}	Shared secret key between TPA and PS
kh_i	Shared integrity key between SM_i and TPA
$E_k[x]$	Plaintext x encrypted using key k

Next, we consider the existing masking-based data aggregation schemes. In 2011, Kursawe et al. suggested a set of masking-based schemes for privacy in smart grids [20]. In their schemes, the authors utilized the concept of Decisional Diffie-Hellman (DDH) groups and bilinear mapping for checking the correctness of the shared masking value, which are computationally expensive and ill-suited for resource constrained smart meters. Besides, it lacks the ability to deal with a dishonest or malicious smart meter that falsifies usage data in order to make the final aggregation result incorrect. Also, during data aggregation, the smart meters reveals their identity, which is fixed for all sessions. Thus, this scheme cannot ensure consumer's privacy, where an adversary can easily comprehend and target a specific user and reveal the consumer's activity through human-factor-aware data aggregation (HDA) attacks [25]. , Shi et al. presented a method that combines masking and distributed differential privacy using noise [21]. However, in their scheme the aggregator can do partial decryption of the meter reading. Hence, the scheme presented in [21] cannot ensure the desired privacy. In [22], Danezis et al. proposed another masking-based scheme where their objective was to examine the usage of complex functions on smart meters by splitting them into Boolean circuits. However, their approach costs more computing rounds and also negatively affects the bandwidth and latency. Moreover, in this scheme each smart

meter sends its usage data without any integrity protection. Furthermore, the schemes presented in [21] and [22] cannot guarantee sender authentication and the consumer's privacy. Recently, Knirsch et al. proposed a masking-based approach for data aggregation [23]. Their scheme utilizes the concept of homomorphic hashing for checking the correctness of the shared secrets. However, this construction has a couple of issues. First, it is complicated to implement and computationally expensive to compute. Second, it cannot ensure security of the hashed data, and an attacker can compute the original message block by taking the logarithm of the hash for that block. Third, it can be shown that the data aggregation scheme presented in [23] is vulnerable to collusion attacks. In this case, when the aggregator (DC in [23]) colludes with a smart meter SM_2, then the aggregator can know the usage data of another smart meter SM_1, which is a serious privacy issue. Apart from [20]-[23], a masking-based multi-hop data aggregation scheme has been proposed in [24]. During data aggregation, each smart meter needs to select n proxies and add masking values to meter readings. Proxies remove these masking values to obtain an aggregated reading. However, this scheme is difficult to implement in practice, cannot ensure the integrity protection of the usage report, and does not provide sender authentication. Recently, Baloglu et al. proposed a solution [37] that combines masking using noise with Decisional Diffie-Hellman (DDH) based encryption, where DDH is used for the encryption of the noisy usage data. In their scheme, metering data is transmitted to at least two aggregators to maintain the integrity and to increase the reliability. This solution incurs high setup and communication cost. Moreover, the proposed solution in [37] cannot ensure sender authentication and consumer's privacy properties.

Motivation: Even though several solutions have been proposed for privacy-preserving data aggregation in smart grids, most of the existing works are based on computationally expensive operations such as homomorphic encryption etc. These are not suitable for resource constrained smart meters, which typically have limited computational capability. For example, a smart meter from Atmel's family with ARM Cortex-M4 processor can provide a maximum CPU speed of 780 MHz [33]. As such, this smart meter may not be suitable to perform any computationally expensive operations. Also, since smart grid systems are mostly operated in a large scale, computationally expensive operations may impair the efficiency of the system. Besides, homomorphic encryption based differential privacy does not guarantee the correct summation result [23]. On the other hand, existing masking-based approaches suffer from the following weaknesses:

- In existing masking-based approaches, a smart meter is not authenticated during data aggregation. In other words, the identity and the legitimacy of the smart meters are not verified. Consequently, a dishonest or fake smart meter may falsify the data, which will cause an inaccurate aggregated result.
- In existing masking-based schemes, computationally expensive operations such as DDH group and Bilinear mapping, or homomorphic hashing are used for verifying the correctness of the masking secrets, which are not suitable for the resource-limited smart meters.
- None of the existing schemes (including homomorphic-encryption-based schemes) ensure anonymity of the consumer. In this case, each smart meter reveals it's identity, which is fixed for all transactions. Therefore, an adversary can easily understand that the usage data is from the same consumer's end and reveal the consumer's activity through HDA attacks [25].

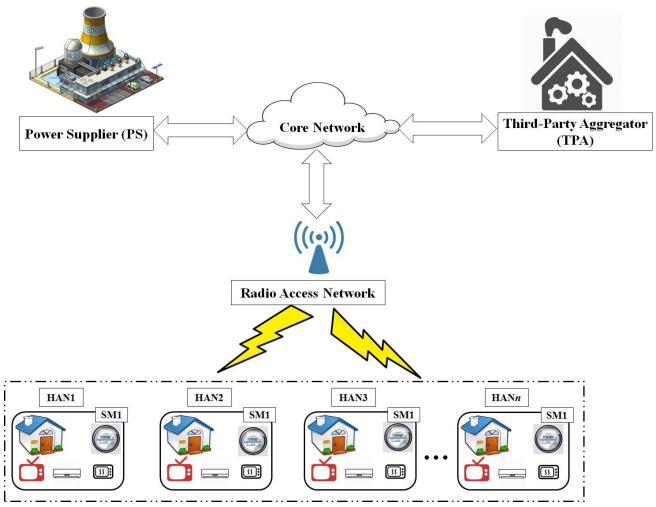

Figure 1. System model for smart grid metering.

B. Our Contribution

This paper first proposes a spatial data aggregation scheme which provides up-to-date and accurate aggregated consumption information to the power grid about any group of consumers. Subsequently, we propose a secure billing solution. In this regard, we only utilize lightweight cryptographic primitives like one-way hash functions, exclusive-or operations, etc. In the proposed scheme, no information about the individual consumers is disclosed. However, the power grid can still monitor the total amount of power needed by its customers situated in a specific region or locality, without compromising the privacy of any individual customer.

The rest of the paper is organized as follows. In Section II, we explain the underlying smart grid model, security goals, and the preliminaries that are relevant to this article. In Section III, we present the proposed spatial data aggregation scheme with a secure billing solution for smart grids. Security of the proposed scheme is analyzed in Section IV. In Section V, we formally analyze the privacy of our proposed scheme. A discussion on the performance of the proposed scheme is given in Section VI. Finally, the conclusion is drawn is Section VII. The symbols and cryptographic functions used in this paper are defined in Table I.

II. System and Adversary Model and Security Goals

In this section, we first briefly describe the network architecture of the proposed privacy-preserving data aggregation

mechanisms for smart grids and also present the adversary model. Subsequently, we define the security goals of the proposed scheme.

A. System Model

Figure 1 shows the system model considered in the paper for smart grid metering, which also forms the foundation of the proposed data aggregation scheme. Our system model consists of four major entities: a power supplier (PS), a third-party aggregator (TPA) employed by the power supplier, a set of smart meters (SMs), and numerous home area networks (HANs). In our system model, the PS is responsible for the distribution of electricity to each HAN. The TPA periodically aggregates the electricity consumption of a group of HANs in a geographical region, and provides the data to the PS. The PS may use this data to adjust its electricity prices for demand-side management and also to provide appropriate feedback to its power generating stations or suppliers. In this way, the TPA plays a crucial role in maintaining the balance between power production and demand. Each HAN is composed of a SM and a set of home appliances. The SM sends its periodic readings to the TPA through an in-home network (e.g. WiFi). The TPA and PS communicate through the public Internet (a cellular network based Internet access is shown as an example in Figure 1).

B. Adversary Model

In our adversary model we consider the PS as a trusted organization (e.g. owned by the government, such as Singapore Power in Singapore and National Grid in United Kingdom). On the other hand, the TPA is owned by a private company whose main responsibility is to assist the PS. Therefore, in our system model we consider the TPA as a honest-but-curious entity, who may want to know the consumption data of each HAN and subsequently may try to sell the usage information to another company, e.g. for marketing materials for home appliances. On the other hand, here we assume that various elements inside the core network may also act as adversaries and be interested in private details of the power consumption of each HAN. A compromised network and its various elements (like a router or a switch) may alter or fabricate the meter's consumption data. Hence, any communication through the network may not be secure. Also, any SM may be the adversary and be interested to know the consumption data of another SM from a different HAN. An outside attacker may also try to impersonate as a legitimate entity (e.g. a SM or the TPA) to send data under its name. For instance, a dishonest or fake SM could falsify the data for causing inaccurate aggregation result. In addition, the outside attacker may eavesdrop on the network transmission media for obtaining the power consumption data and may also try to alter or re-transmit them.

C. Security Goals

- **Authentication:** Before aggregating any data, the TPA needs to authenticate each SM. This will prevent any inaccurate aggregation results. Similarly, before obtaining any relevant information from the PS through the insecure public communication channel, the TPA needs to authenticate the PS.
- **Usage Data Confidentiality:** The secrecy of the end-to-end communication during meter data collection is vital. Therefore, the electricity consumption data is required to be kept secret from any third party for protecting the privacy of the customer. In this regard, even if an outside or an inside adversary like other SMs from different HANs or the TPA obtains the messages with electricity consumption information, then he/she should not be able to comprehend the encrypted message.
- **Usage Data Integrity:** The TPA should be able to verify the integrity of the data received from the SM of each HAN. Similarly, the TPA needs to verify the integrity of the relevant information received from the PS during data aggregation.
- **Consumer Privacy:** The TPA should not be able to know the real identity of a HAN user. Only the PS should have the ability to know a consumer's real identity. In addition, after eavesdropping the usage data, an outside adversary should not be able to comprehend that the data is from a particular consumer or if two meter readings are from the same user.

III. Proposed Privacy-Friendly Spatial Data Aggregation Scheme

In this section, we propose our privacy-friendly spatial data aggregation scheme for smart grids, which consists of two phases: *authenticated initialization* and *data aggregation*. Assume that there are n HANs in a locality which obtain their power supply from the PS. In the *authenticated initialization phase*, smart meter SM_i and the aggregator TPA prove their legitimacy to the PS and subsequently establish a key kh_i and a set of temporary identities between them. Besides, this phase also helps both the SM_i and the TPA to update their secret key kh_i and establish a new set of temporary identities. In the *data aggregation phase*, the TPA periodically aggregates the electricity consumption of a group of HANs in a locality without knowing the power consumption of each individual HAN.

A. Authenticated Initialization

Consider the scenario where the PS is interested in the aggregated power consumption of a group of n HANs that are its customers. During the installation of the smart meter SM_i of each household HAN_i, the PS randomly generates a pseudo identity PID_i and a secret key k_i using the meter's pseudo random number generator (PRNG) and assigns them to SM_i. The PS records PID_i and k_i for future communication with SM_i. This phase of the proposed scheme consists of the following steps:

Step AU1: The smart meter SM_i generates a random number N_s and calculates $V_0 = h(PID_i||N_s||k_i)$. Then, SM_i composes a message $\mathrm{M}_{\mathrm{A}_1} : \{PID_i, N_s, V_0\}$ and sends it to the TPA.

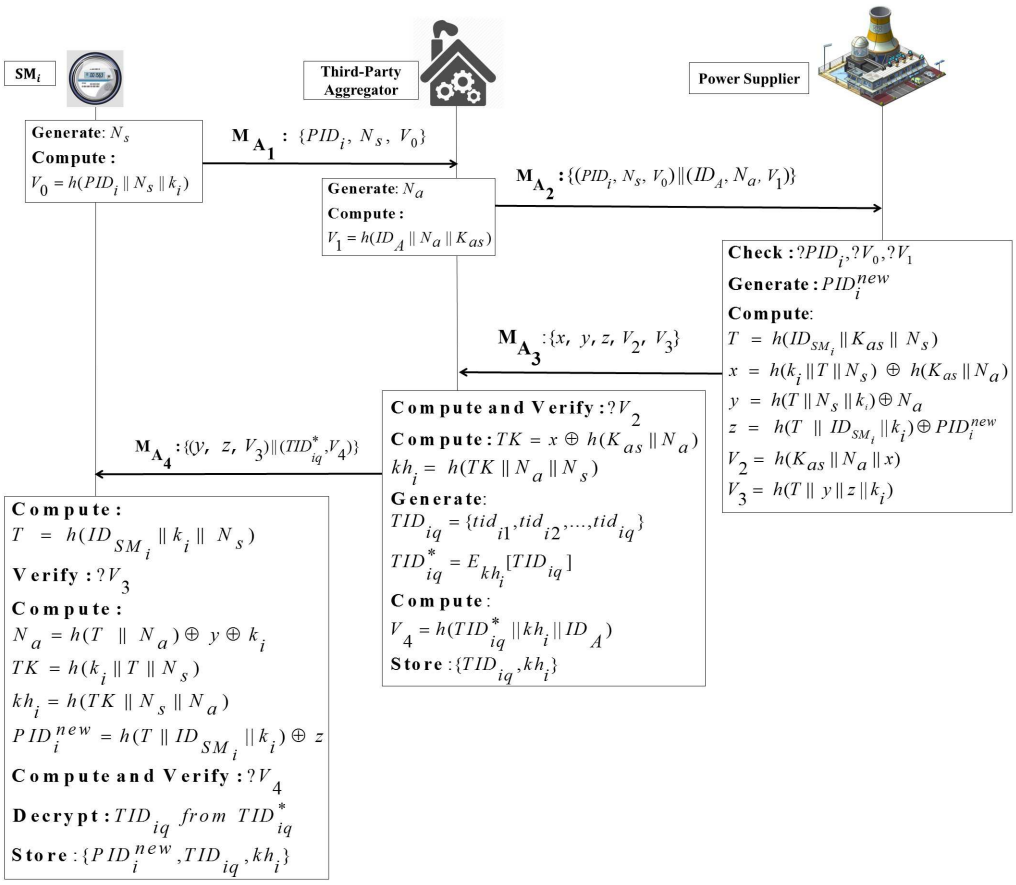

Figure 2. Authenticated initialization process.

Step AU2: After receiving the request message $\mathrm{M_{A_1}}$, the TPA generates a nonce N_a and calculates a hash-integrity output $V_1 = h(ID_A||N_a||K_{as})$. Then the TPA composes a message $\mathrm{M_{A_2}} : \{(PID_i, N_s, V_0)||(ID_A, N_q, V_1)\}$ and sends it to the PS.

Step AU3: Upon receiving message $\mathrm{M_{A_2}}$, the PS tries to map the identity PID_i to the real identity of a user, and then computes and verifies V_0 and V_1. If the verification is successful then the PS generates a new pseudo identity PID_i^{new} and computes $T = h(ID_{SM_i}||k_i||N_s)$, $x = h(k_i||T||N_s) \oplus h(K_{as}||N_a)$, $y = h(T||N_s||k_i) \oplus N_a$, $z = h(T||ID_{SM_i}||k_i) \oplus PID_i{}^{new}$, $V_2 = h(K_{as}||N_a||x)$, and $V_3 = h(T||y||z||k_i)$. After that, the PS composes a response message $\mathrm{M_{A_3}} : \{x, y, z, V_2, V_3\}$ and sends it to the TPA.

Step AU4: After receiving the response message $\mathrm{M_{A_3}}$, the TPA first computes and validates the parameter V_2. If the validation is successful, then the TPA first calculates $TK = x \oplus h(K_{as}||N_a)$ and $kh_i = h(TK||N_a||N_s)$. Then the TPA generates a set of temporary identities $TID_{iq} = \{tid_{i1}, tid_{i2}, \cdots, tid_{iq}\}$ and derives $TID^*_{iq} = E_{kh_i}[TID_{iq}]$ and $V_4 = h(TID_{iq}||kh_i||ID_A)$. Finally, the TPA composes a response message $\mathrm{M_{A_4}} : \{(y, z, V_3)||(TID^*_{iq}||V_4)\}$ and sends it to SM_i.

Step AU5: Upon receiving message $\mathrm{M_{A_4}}$, SM_i first computes $T = h(ID_{SM_i}||k_i||N_s)$ and verifies V_3. If the verification is successful, then SM_i derives $N_a = h(T||N_s||k_i) \oplus y$, $TK = h(k_i||T||N_s)$, $kh_i = h(TK||N_a||N_s)$, and $PID_i^{new} = h(T||ID_{SM_i}||k_i) \oplus z$. Hereafter, SM_i decrypts TID_{iq} from TID^*_{iq} and stores $\{TID_{iq}, kh_i\}$ for data aggregation. The details of this phase are depicted in Figure 2.

B. Data Aggregation

Our data aggregation process consists of the following steps:

Step AG1: To maintain proper balance between power production and demand, the power supplier PS periodically (say, every 1 or 2 hours) needs to know the electricity consumption of the group of *n* HANs. In order to do that, for each time interval T_j, the PS picks a set of *n* random integers $R_j = \{r_1, r_2, \cdots, r_n\}$ from a cryptographic pseudo random number generator that fully exploits the range $\{0, \cdots, d-1\}$ of a uniform distribution, where $d \gg \sum_{i=1}^{n} M_i$, where M_i is the meter reading of SM_i. Hereafter, the PS selects a random integer $r_i \in R_j$ for each smart meter SM_i and calculates $\Delta_i = E_{k_i}[ID_{SM_i}||r_i||k_i||T_j]$ and $H_i = h(\Delta_i||k_i||T_j)$. In this way, for all smart meters of n HANs, the PS derives $(\Delta^*, H^*) = \{(\Delta_1, H_1), (\Delta_2, H_2), \cdots, (\Delta_n, H_n)\}$. Hereafter, the PS generates a timestamp t_{ps} and calculates $R_{Sum} = \sum_{i=1}^{n}(r_i \mod d)$, $\Delta_{PS} = E_{K_{as}}(R_{Sum}||T_j)$, and $H_{PS} = h(\Delta_{PS}||K_{as}||t_{ps})$, and sends $\mathfrak{R}^* = \{(\Delta^*, H^*)||(R_{Sum}, \Delta_{PS}, t_{ps})\}$ to the TPA for using at pre-

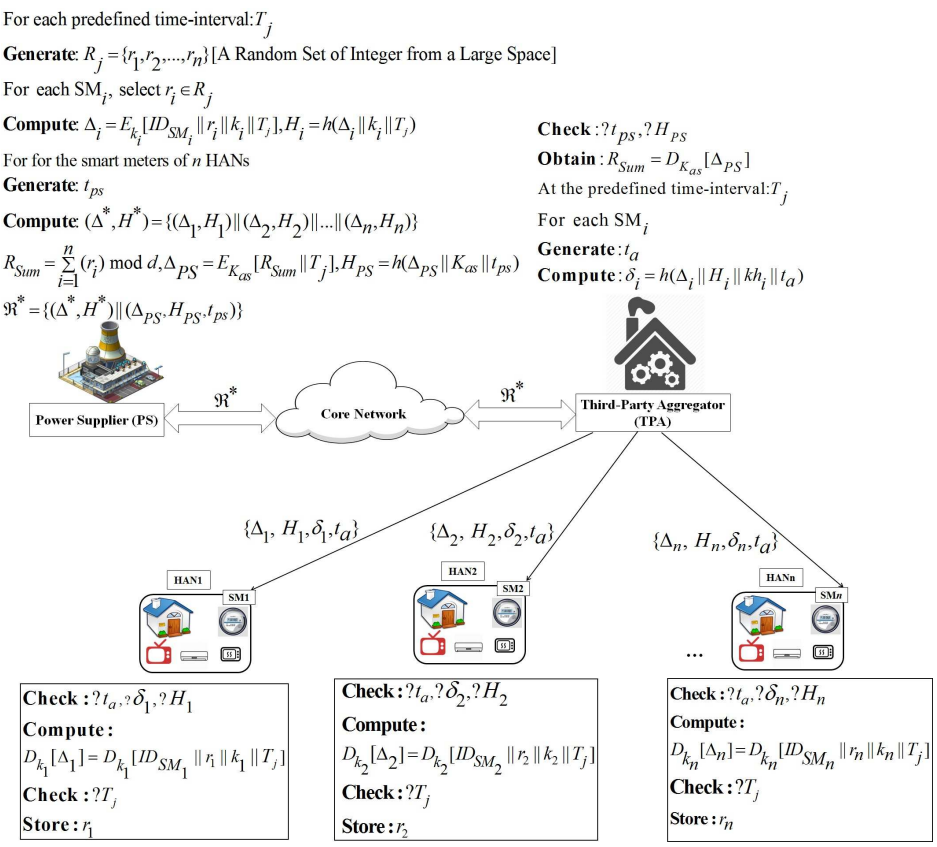

Figure 3. Pictorial representation of step AG1 of the proposed data aggregation scheme.

defined time interval T_j. After receiving message $\Re^*$, the TPA first checks t_{ps} and H_{PS}, and then obtains R_{Sum} from Δ_{PS}. Subsequently, at the time interval T_j the TPA generates a timestamp t_a and for each smart meter SM_i, it computes $\delta_i = h(\Delta_i || H_i || kh_i || t_a)$ and finally distributes $(\Delta_i, H_i, \delta_i, t_a)$ to each smart meter SM_i. Upon receiving $(\Delta_i, H_i, \delta_i, t_a)$, smart meter SM_i first checks t_a, δ_i, H_i, and then decrypts Δ_i and verifies the time interval T_j. This verification prevents the TPA from repeatedly using the same Δ_i for two different time intervals. If the verification is successful, SM_i decrypts Δ_i and obtains the random integer r_i. Details of this step are shown in Figure 3.

Step AG2: After obtaining the random integer r_i, SM_i generates a timestamp t_i and selects an unused temporary identity $tid_{ij} \in TID_{iq}$ and calculates its blinded measurement $X_i = M_i + r_i \mod d$, computes $H_i = h(X_1 || kh_i || t_i)$, composes a message $\{tid_{ij}, X_i, H_i, t_i\}$, and sends it to the TPA. Finally, SM_i deletes tid_{ij} from TID_{iq}. Once all the temporary identities are used up, SM_i needs to ask for a new set from the TPA. In that case, SM_i and the TPA execute the *authenticated initialization phase* again. Now, upon receiving the usage data from each smart meter, the TPA first locates and validates the temporary identity tid_{ij} and key-hash integrity output H_i. If the validation is successful, the TPA computes $\sum_{i=1}^{n} X_i$ (i.e., $\sum_{i=1}^{n} M_i + \sum_{i=1}^{n} r_i \mod d - R_{Sum} = \sum_{i=1}^{n} M_i$). In this way, the TPA obtains the aggregated power consumption data of n HANs. Details of this step are shown in Figure 4. Note that if any check in the aforementioned steps is invalid, then this phase of the proposed scheme is aborted. Also, to ensure more efficient performance of the above data aggregation scheme, the PS can pre-compute $\Re^* = \{(\Delta^*, H^*) || (R_{Sum}, \Delta_{PS})\}$ for several sessions and send them to the TPA. In this way, we can expedite the data aggregation process. Now, for the correctness of our protocol, all the smart meters need to participate during the data aggregation process. To avoid the failure report problem (i.e. the absence of reports when a smart meter fails), the TPA needs to do ping tests with the smart meters on a regular basis. In case the TPA does not receive any response from smart meter SM_i, it informs the PS to take necessary actions. In this context, the PS first abstains from creating any r_i for that particular smart meter and then initiates technical support steps to resolve the issue.

Privacy Enhancement Under Collusion: In the system model considered so far, the PS is assumed to be a trusted entity (e.g. owned by the government). However, this assumption may not be valid for all scenarios. In this context, if the PS colludes with the the TPA, then the TPA will be able to know the individual measurements of the smart meters. However, this issue can be easily addressed with a few changes to the proposed scheme. In this regard, some changes are required in step AG2 of the proposed scheme. In this new scenario, we assume that each smart meter can directly communicate with its neighboring smart meters and each meter SM_i has a secure line with its adjacent neighbor SM_{i+1}. Now, after obtaining the masking value r_i from the TPA, each smart meter SM_i picks a random number s_i (called its "share")

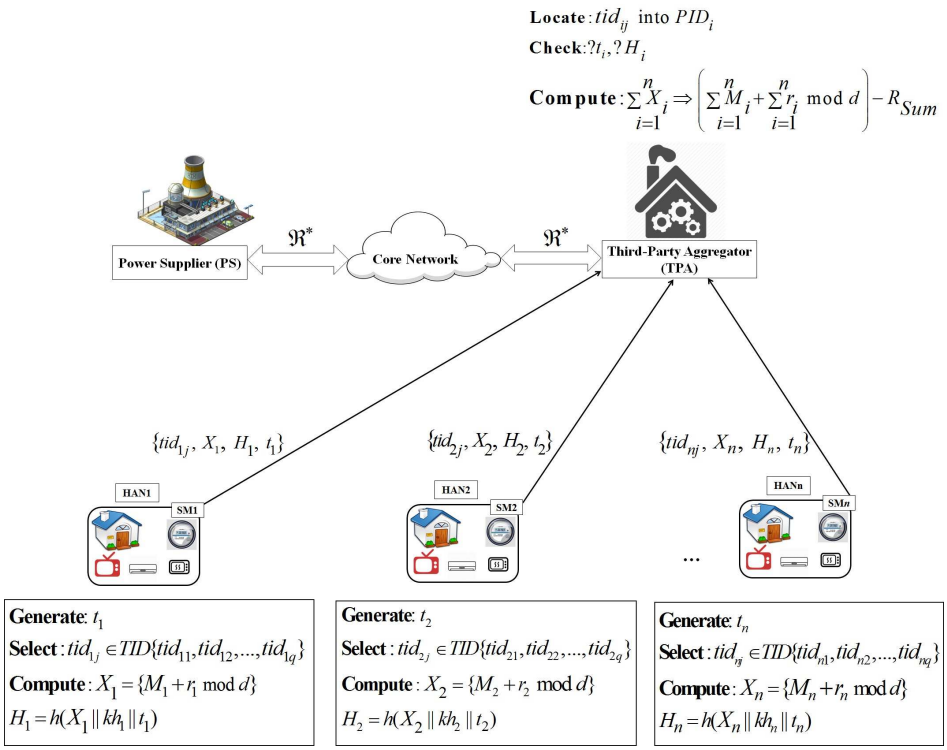

Figure 4. Pictorial representation of step AG2 of the proposed data aggregation scheme.

from a large space and adds this share to its measurement value M_i yielding X_i, i.e., $X_i = M_i + s_i + r_i \mod d$, which is then sent directly to the TPA. Additionally, SM_i adds s_i to the accumulated share value S_{i-1} that it has received from SM_{i-1} and calculates $S_i = s_i + S_{i-1}$. SM_i then sends S_i to its next adjacent neighbor SM_{i+1} through the secure line. This continues up to the last smart meter SM_n which computes $S_n = S_{n-1} + s_n$, which equals $\sum_{i=1}^{n} s_i$. Finally, SM_n encrypts S_n and the timestamp t_n with kh_n and sends it to the TPA. The TPA computes $\sum_{i=1}^{n} X_i - (R_{sum} + S_n)$ yielding $\sum_{i=1}^{n} M_i$, which gives the desired aggregated load. The details of the revised step AG2 are depicted in Figure 5.

C. Secure Billing

We assume that each smart meter SM_i maintains a parameter β_i for billing. Initially, during meter installation, the value of β_i is set to 0. Now, for each time interval T_j, when SM_i sends its blinded measurement X_i to the TPA for spatial data aggregation, then SM_i also updates $\beta_i = M_i^j + \beta_i$ and stores β_i in its memory, where M_i^j denotes the meter reading of SM_i at time T_j. Finally, at the end of the month (or any desired interval), SM_i generates a timestamp t and selects an unused temporary identity tid_i and then computes $E_{k_i}[\beta_i||PID_i]$, $\nu_i = h(E_{k_i}[\beta_i||PID_i]||k_i||t)$, and composes a message $Bill_i = \{\text{“}Billing\text{”}, tid_i, E_{k_i}[\beta_i||PID_i], \nu_i, t\}$ and sends it to the TPA. After receiving the message $Bill_i$, the TPA first finds PID_i corresponding to tid_i and then composes a message $Bill_i^* = \{\text{“}Billing\text{”}, E_{K_{as}}[PID_i], E_{k_i}[\beta_i||PID_i], \nu_i, t\}$ and then sends it to the PS. Upon receiving $Bill_i^*$, the PS first decrypts $E_{K_{as}}[PID_i]$ and then checks the timestamp t and ν_i. If they are valid then the PS decrypts and obtains β_i for PID_i. Then the PS defines an acknowledgment $\mathcal{ACK}_i$ and generates a timestamp t^* and a valid key-hash response $\lambda = h(\mathcal{ACK}_i||k_i||t^*)$, and subsequently sends $(\mathcal{ACK}_i, \lambda, t^*)$ to SM_i through the TPA. When SM_i receives the acknowledgment $\mathcal{ACK}_i$, it first checks the timestamp t^* and and the key-hash response λ. If they are valid, then SM_i informs its owner and sets β_i to 0; otherwise, it requests the PS for the acknowledgment.

IV. Security Analysis

This section demonstrates that the proposed scheme ensures all the security goals listed in Section II.

A. Accomplishment of Authentication

In the *authenticated initialization phase* of the proposed scheme, the PS authenticates SM_i by verifying the pseudo identity PID_i and the parameter V_0 in request message $\mathrm{M}_{\mathrm{A}_2}$, where only a legitimate SM_i can generate the valid key-hash output V_0. The PS authenticates the TPA by using the request parameter V_1, which must be equal to $h(ID_A||N_a||K_{as})$. On the other hand, both SM_i and the TPA authenticate the PS by using the response parameters V_2 and V_3, respectively. Now, in the *data aggregation phase* of the proposed scheme, before aggregating the usage data, the TPA authenticates each smart meter SM_i by using the timestamp t_i and the response H_i. Moreover, in this phase of the proposed scheme, the TPA authenticates the PS by using hash-response parameter H_{PS}. Furthermore, in the proposed data aggregation scheme, if an adversary tries to perform any replay attempt, the receiving end can easily comprehend such activities by using

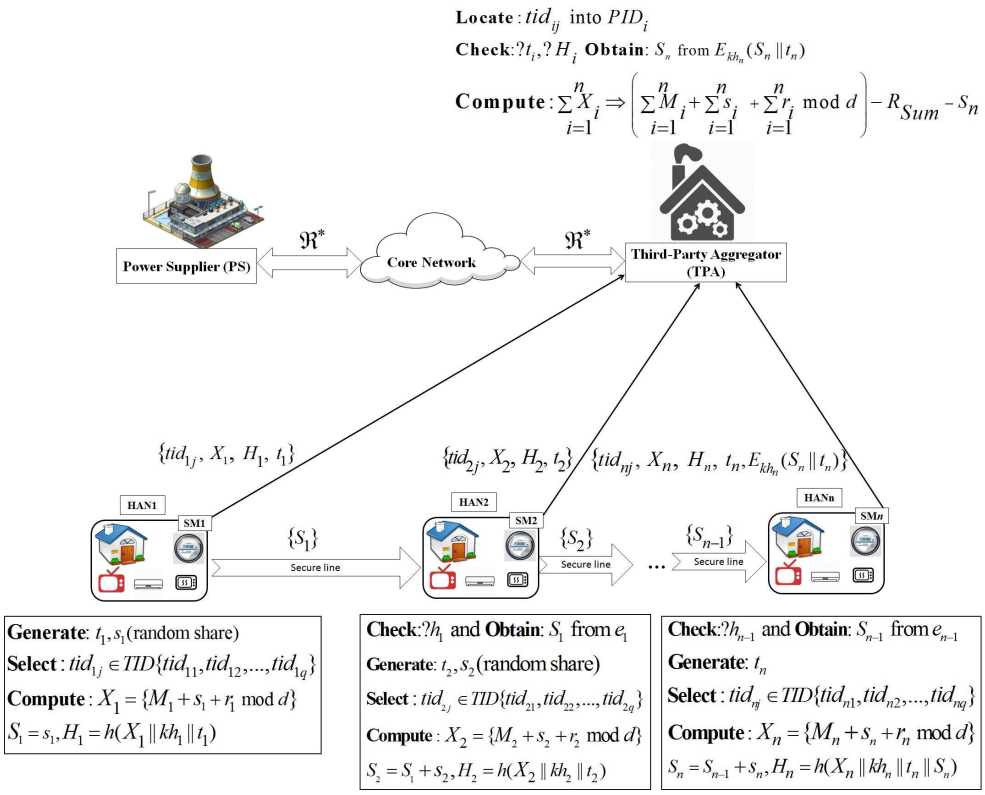

Figure 5. Pictorial presentation of the revised step AG2 for addressing collusion attacks between PS and TPA

the timestamps t_i, t_a, and t_{ps}. On the other hand, during the billing phase, if an adversary (even TPA) attempts to forge the billing value β_i, then the PS will be able to detect that by using the key-hash response ν_i. In this way, the proposed scheme can detect any forgery attacks..

B. Accomplishment of Secure Key-Establishment

In the *authenticated initialization phase* of the proposed data aggregation scheme, each smart meter SM_i and the TPA need to securely establish a key kh_i, which will protect against dishonest or fake smart meters from falsifying data. In this regard, only the legitimate TPA who knows the secret key K_{as} can calculate $TK = x \oplus h(K_{as}||N_a)$ and $kh_i = h(TK||N_a||N_s)$. Similarly, only an authentic smart meter SM_i with the installed secret key k_i can derive $T = h(ID_{SM_i}||k_i||N_s)$, $TK = h(k_i||T||N_s)$, and $kh_i = h(TK||N_a||N_s)$. Therefore, the security of the shared secret key kh_i depends on the secrecy of the keys K_{as} and k_i, where it is assumed that the respective entities (PS, TPA, and SM_i) will not reveal their shared secret to anyone.

C. Accomplishment of Usage Data Confidentiality

The amount electricity usage in HAN_i is blinded with the random integer r_i from a large space, i.e., $X_i = M_i + r_i \bmod d$. Therefore, the TPA can only know the blinded measurement of each SM_i. Besides, after calculating $\sum_{i=1}^{n} X_i - R_{Sum}$, the TPA can only obtain the aggregated spatial (i.e., a group of HANs') usage data. This prevents analysis of a single customer's data. Also, since the masking integer r_i is chosen randomly, even if the usage of electricity for two consumers is the same, an adversary (even the TPA) cannot comprehend it from the blinded measurements. Thus, the pattern of the electricity consumption is protected from detection by any eavesdropper. Furthermore, for ensuring privacy under collusion attacks between the PS and the TPA, each smart meter SM_i uses a random share s_i for obtaining X_i, i.e., $X_i = M_i + s_i + r_i \bmod d$. Therefore, even if the PS and the TPA collude, since they do not know the value of s_i, they will not be able to obtain the desired M_i. Hence, the proposed scheme can ensure privacy even under collusion attacks.

D. Accomplishment of Usage Data Integrity

In the proposed scheme, before doing data aggregation, the TPA first checks whether it has received the same data as that was sent by each smart meter SM_i. For that, the TPA computes $H_i^* = h(X_i||kh_i||t_i)$ and checks whether H_i^* is equal to H_i or not. This approach facilitates the detection of any manipulation of the aggregated usage data during communication.

E. Accomplishment of Consumer Privacy

In the proposed data aggregation scheme, no one except for the PS can gain knowledge of any private information such as the real-identity of a HAN user. The TPA only knows the pseudo identity PID_i of an user based on which it accumulates the reading of each smart meter SM_i. We also note that while sending the usage data, SM_i is not allowed to use the same temporary identity tid_{ij} twice. No one except the TPA can recognize that. Therefore, an outsider cannot guess whether

the usage data for two consecutive sessions are from the same HAN. This approach of the proposed scheme is useful for achieving privacy against eavesdropper (PAE) [32].

V. Formal Privacy Analysis

In this section, we formally analyze the privacy of the proposed scheme.

A. Privacy Model

We now consider Ouafi and Phan's privacy model [29]. In this model, attacker $\mathcal{A}$ can eavesdrop on all the channels between the smart meters and TPA and he/she can also perform any active or passive attacks. $\mathcal{A}$ is allowed to run the following queries:

- **Execute** $(\mathcal{M}, \mathcal{TPA}, i)$: This query represents the passive attacks. In this context, the attacker can eavesdrop all the transmitted messages between the smart meter $\mathcal{M}$ and the aggregator $\mathcal{TPA}$ in the i-th session. Consequently, the attacker obtains all the exchanged data between the $\mathcal{TPA}$ and meter $\mathcal{M}$.
- **Send**(U, V, m, i): This query models the active attacks in the system. In this query, attacker $\mathcal{A}$ has the permission to impersonate an aggregator U in the i-th session, and forwards a message m to a smart meter V. Besides, the attacker has the permission to block the exchanged message m between the smart meter and the aggregator.
- **Corrupt**$(\mathcal{M}, K)$: In this query, the attacker $\mathcal{A}$ has the permission to access secret information K stored in the smart meter's memory.
- **Test**$(\mathcal{M}_0, \mathcal{M}_1, i)$: This query is the only query that does not correspond to any of $\mathcal{A}$'s abilities or any real-world event. This query allows to define the indistinguishability-based notion of untraceable privacy. If the party has accepted and is being asked a Test query, then depending on a randomly chosen bit $b \in \{0, 1\}$, $\mathcal{A}$ is given $\mathcal{M}_b$ from the set $\{\mathcal{M}_0, \mathcal{M}_1\}$. Informally, $\mathcal{A}$ succeeds if it can guess the bit b. In order for the notion to be meaningful, a **Test** session must be fresh in the sense of Definition 2.

Definition 1 (Partnership and Session Completion): An aggregator instance $\mathcal{TPA}_j$ and a meter instance $\mathcal{M}_i$ are partners if, and only if, both have output Accept($\mathcal{M}_i$) and Accept($\mathcal{TPA}_j$), respectively, signifying the completion of the protocol session.

Definition 2 (Freshness): A party instance is fresh at the end of execution if, and only if (i) it has output Accept with or without a partner instance and (ii) both the instance and its partner instance (if such a partner exists) have not been sent a Corrupt query.

Definition 3 (Indistinguishable Privacy (INDPriv)): It is defined using the game $\mathcal{G}$ played between a malicious adversary $\mathcal{A}$ and a collection of smart meters and reader and aggregator instances. $\mathcal{A}$ runs the game $\mathcal{G}$ whose setting is as follows.

- ***Learning phase*:** $\mathcal{A}$ is able to send any **Execute** and **Send** query and interact with the aggregator $\mathcal{TPA}$ and smart meter $\mathcal{M}_0$ and $\mathcal{M}_1$ that is chosen randomly.
- ***Challenge phase***: The attacker selects two meters $\mathcal{M}_0$ and $\mathcal{M}_1$ and forwards a Test query $(\mathcal{M}_0, \mathcal{M}_1, i)$ to challenger $\mathcal{C}$. After that, $\mathcal{C}$ randomly selects $b \in \{0, 1\}$ and the attacker determines the meter $\mathcal{M}_b \in \{\mathcal{M}_0, \mathcal{M}_1\}$ using *Execute* and *Send* queries.
- ***Guess phase:*** The attacker $\mathcal{A}$ finishes the game $\mathcal{G}$ and outputs a bit $b' \in \{0, 1\}$ as guess of b. The success of attacker $\mathcal{A}$ in the game $\mathcal{G}$ and consequently breaking the security of INDPriv is quantified via $\mathcal{A}$'s advantage in recognizing whether attacker $\mathcal{A}$ received $\mathcal{M}_0$ or $\mathcal{M}_1$, and is denoted by $Adv_{\mathcal{A}}^{INDPriv}(k) = |\Pr b^{'} = b] - 1/2|$, where k is a security parameter.

Proposition 1: *The proposed scheme satisfies Indistinguishable Privacy.*

Proof. In the proposed scheme, each meter reading is masked with a new random integer r_j. Besides, the temporary identity TID changes in each session. Therefore, it is difficult for an adversary to perform any traceability attack by performing the following phases:

- ***Learning phase:***: In the j-th round, the attacker $\mathcal{A}$ sends an *Execute* query $(\mathcal{TPA}, \mathcal{M}_0, j)$ and obtains the parameters $\{tid_j^{\mathcal{M}_0}, X_{0,j}^{\mathcal{M}_0}, H_{0,j}\}$.
- ***Challenge phase:*** $\mathcal{A}$ selects two meters $\mathcal{M}_0$ and $\mathcal{M}_1$ and sends a Test query $(\mathcal{M}_0, \mathcal{M}_1, j+1)$. Next, according to the randomly chosen bit $b \in \{0, 1\}$, the attacker is given a meter $\mathcal{M}_b \in \{\mathcal{M}_0, \mathcal{M}_1\}$. After that the attacker $\mathcal{A}$ sends an Execute query $(\mathcal{TPA}, \mathcal{M}_b, j+1)$ and obtains $\{tid_{j+1}^{\mathcal{M}_b}, X_{0,j+1}^{\mathcal{M}_b}, H_{0,j+1}\}$.
- ***Guess phase:*** In the *Learning phase* the meter $\mathcal{M}_0$ updates its masking secret r_j. Therefore, for the two subsequent sessions j and $j+1$ the parameter $(X_{0,j}^{\mathcal{M}_0}, X_{0,j+1}^{\mathcal{M}_b})$ and $(H_{0,j}^{\mathcal{M}_0}, H_{0,j+1}^{\mathcal{M}_b})$ are calculated as follows: $X_{0,j}^{\mathcal{M}_0} = \mathcal{M}_{0,j} + r_{\mathcal{M}_0,j} \mod d$, $X_{0,j+1}^{\mathcal{M}_b} = \mathcal{M}_{b,j+1} + r_{\mathcal{M}_b,j+1} \mod d$, $H_{0,j}^{\mathcal{M}_0} = h(X_{0,j}^{\mathcal{M}_0} || kh_{\mathcal{M}_0} || t_{\mathcal{M}_0,j})$, and $H_{0,j+1}^{\mathcal{M}_b} = h(X_{0,j+1}^{\mathcal{M}_b} || kh_{\mathcal{M}_b} || t_{\mathcal{M}_b,j+1})$. Since $r_{\mathcal{M}_0,j} \neq r_{\mathcal{M}_b,j+1}$, $t_{\mathcal{M}_0,j} \neq t_{\mathcal{M}_b,j+1}$, $tid_j^{\mathcal{M}_0} \neq tid_{j+1}^{\mathcal{M}_b}$, and $h(\cdot)$ is an ϵ-secure pseudorandom function, the adversary thus needs to make a random guess. In this context, the advantage of the adversary at recognizing $\mathcal{M}_0$ or $\mathcal{M}_1$ can be denoted by $Adv_{\mathcal{A}}^{INDPriv}(k) = |\Pr[b^{'} = b] - 1/2| \leq \epsilon$.

VI. Performance Analysis and Comparison

The objective of the proposed spatial data aggregation scheme is not only to fulfill several security requirements for smart meters, but also to ensure that the computational overhead during the data aggregation process is reasonable. To manifest the advantages of the proposed scheme, in this section, we first compare the performance of the proposed scheme with the following previously proposed non-masking-based data aggregation schemes for smart grids: [4], [5], [10], [11], [12] [18], [19], [35], and [36]. Table II shows the security properties that each scheme supports, and we can see that the proposed scheme and the schemes presented in [4], [5], [11], [12], [18], [19], [35], and [36] can guarantee data confidentiality but [10] cannot. In this regard, during data aggregation, the scheme presented in [10] reveals the consumer's usage profile to the aggregator and the outside

Table II
PERFORMANCE BENCHMARKING BASED ON SECURITY PROPERTIES WITH RESPECT TO NON-MASKING-BASED SOLUTIONS

Scheme	Data Confidentiality	Data Integrity	Sender Authentication	Consumer's Privacy
Garcia et al. [4]	Yes	No	No	No
Liang et al. [5]	Yes	No	No	No
Liu et al. [10]	No	Yes	No	No
Zhang et al. [11]	Yes	No	Yes	No
Sui et al. [12]	Yes	No	No	Yes
Wang et al. [18]	Yes	Yes	No	No
Jo et al. [19]	Yes	Yes	Partial	No
Vahedi et al. [35]	Yes	Yes	Yes	No
Zhang et al. [36]	Yes	Yes	Yes	No
Proposed Scheme	Yes	Yes	Yes	Yes

Table III
COMPUTATION COST OF DIFFERENT CRYPTOGRAPHIC OPERATIONS

Machine	Paillier Encryption [4][5][18][19]	Paillier Decryption [4][5][18][19]	Pairing Time [10][12][36]	ECDSA Signature Generation [11][19][35]	ECDSA Signature Verification [11][19][35]	SHA-256 [Proposed Scheme]
2.60 GHz CPU (Operating as TPA /PS)	18.62 ms	31.45 ms	161.82 ms	23.81 ms	17.56 ms	0.12 ms
798 MHz CPU (Operating as SM)	89.70 ms	152.6 ms	685.3 ms	837.92 ms	768.20 ms	0.43 ms

Table IV
VARIATION IN AGGREGATION TIME FOR VARIOUS NUMBER OF SMS

Schemes	Number of Smart Meters	Aggregation Time
Paillier-based Schemes ([4][5][18][19])	50	1570 ms
	80	2175 ms
	120	3290 ms
Pairing-based Schemes ([10][12][36])	50	8091 ms
	80	12945 ms
	120	19418 ms
ECDSA-based Schemes ([11][19][35])	50	878 ms
	80	1404 ms
	120	2107 ms
Proposed Scheme	50	6.28 ms
	80	9.98 ms
	120	14.83 ms

adversary. Table II also shows that the some of these schemes do not support data integrity during data aggregation. This allows an adversary to alter the usage data and cause an incorrect aggregated result without the aggregator detecting it. Next, from Table II we see that most of these non-masking-based schemes do not support sender authentication. As a result, a dishonest or fake smart meter may falsify the usage data, leading to an incorrect aggregation outcome. In [19], the authors have proposed two data aggregation schemes, only one of which ensures sender authentication (discussed in Section IA). Hence, [19] only partially supports sender authentication. Furthermore, almost all of the existing schemes allow the transmission of the identity of a smart meter in plain-text. As a result, an outside attacker can target the smart meter of a particular HAN and reveal the consumer's behavior through HDA attacks. Hence, these schemes cannot ensure a consumer's privacy. Sui et al. [12] have considered this issue and designed their scheme with the assumption of an anonymity network. However, their scheme needs to bear the additional computational overhead for establishing such a network.

To show the effectiveness of our proposed scheme with

Table V
PERFORMANCE BENCHMARKING BASED ON SECURITY PROPERTIES WITH RESPECT TO MASKING-BASED SOLUTIONS

Scheme	Data Confidentiality	Data Integrity	Sender Authentication	Consumer's Privacy
Kursawe et al. [20]	Yes	Yes	No	No
Shi et al. [21]	No	Yes	No	No
Danezis et al. [22]	Yes	No	No	No
Knirsch et al. [23]	Partial	Yes	No	No
Baloglu et al. [37]	Yes	Yes	No	No
Proposed Scheme	Yes	Yes	Yes	Yes

Table VI
PERFORMANCE COMPARISON BASED ON METHODOLOGIES WITH RESPECT TO MASKING-BASED SOLUTIONS

Scheme	Masking Method	Utilized Method for Masking Data Integrity
Kursawe et al. [20]	Addition of random values	Decisional Bilinear Diffie-Hellman
Shi et al. [21]	Noisy Statistics	Decisional Diffie-Hellman
Danezis et al. [22]	Complex function with Boolean circuits	-
Knirsch et al. [23]	Addition of random shares (generated from PRNG)	Homomorphic hashing
Baloglu et al. [37]	Noise and DDH Encryption	Decisional Diffie-Hellman
Proposed Scheme	Addition of random values (generated from PRNG)	Normal secure non-collision hash function

respect to the existing non-masking-based schemes, we conducted simulations of the cryptographic operations used by various schemes on an Ubuntu 12.04 virtual machine with an Intel Core i5-4300 dual-core 2.60 GHz CPU (operating as the TPA or the PS as per the scheme). To simulate a smart meter, we used a single-core 798 MHz CPU with 256 MB of RAM, which reflects the capabilities of real smart meters. The simulations used the JPBC library Pbc-0.5.14 [26], JCE [27], and the Paillier library libpaillier-0.8 [28] to evaluate the execution time of different cryptographic operations. Table III shows the computation time of the cryptographic operations for 768 bits of data. From Table III, we can see that SHA-256 leads to significantly lower computation cost as compared to other primitives, and hence is better suited for resource constrained smart meters. Table IV shows the variation in the aggregation time for different numbers of SMs in the proposed scheme, Paillier homomorphic encryption-based schemes ([4], [5], [18] and [19]), pairing-based schemes ([10], [12]), [36]) and ECDSA signature-based schemes ([11], [19]), [35]). It can be seen from Table IV that the aggregation time is significantly lower in the proposed scheme as compared to the others.

Next, we compare the performance of the proposed scheme with existing masking-based aggregation schemes for smart grids: [20], [21], [22], [23], and [37]. Table V shows the security properties that each scheme supports, and we can see that our proposed scheme and the schemes presented in [20], [22], [37] can guarantee data confidentiality, while the schemes presented in [21] and [23] cannot. In [21], the TPA is allowed to know the individual meter readings (discussed in Section IA). The scheme presented in [23] can only ensure data confidentiality when the aggregator (DC in [23]) does not collude with a smart meter (discussed in Section IA). Hence, we say that the scheme presented in [23] can partially ensure data confidentiality. Table V also shows that although most of the masking-based schemes (except [22]) can ensure data integrity, they do not authenticate the sender (smart meter) during the data aggregation process. Consequently, a dishonest or fake smart meter may falsify the data, leading to an inaccurate aggregated result. Moreover, similar to the existing non-masking-based schemes, the schemes presented in [20], [21], [22], [23], [37] allow the smart meters to send their identity in plain-text. Hence, these schemes cannot ensure security under HDA attacks. On the other hand, the proposed scheme ensures data integrity and sender authentication through one-way non-collision hash functions. In addition, smart meters use their temporary identities during the data aggregation process, and meters are not allowed to use a temporary identity more than once. Thus, an attacker cannot isolate or identify the information from any specific HAN.

Next, we compare the existing masking-based schemes and our scheme with respect to masking and data integrity methodologies. From Table VI, we can see that both the proposed scheme and the scheme presented in [23] use the same approach of masking, where the masking random values are generated from a cryptographic pseudo random number generator. Table VI also shows that the proposed scheme uses the normal secure one-way hash-function for verifying the masking data integrity. On the other hand, existing masking-based schemes use computationally expensive operations such as decisional bilinear Diffie-Hellman, homomorphic hashing etc. for the same purpose, which result in high computational overhead on resource limited smart meters. Now, in order to analyze the performance of the proposed scheme in terms of computation cost more comprehensively, we compare it to Knirsch et al.'s scheme [23]. In this context, we conduct simulations of the cryptographic operations used in [23] (such as homomorphic hashing) on the same platform that we used for the performance evaluation of the non-masking-based schemes: an Ubuntu 12.04 virtual machine with an Intel Core i5-4300 dual-core 2.60 GHz CPU (operating as the TPA) and

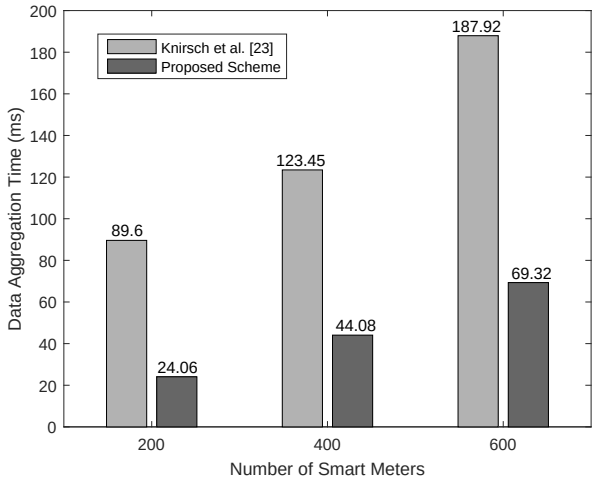

Figure 6. Variation of aggregation time in terms of number of SMs

a single core 798 MHz CPU and 256 MB of RAM (operating as a smart meter). The simulations used the GNU MP library and the JCE library [27] to evaluate the execution time of the cryptographic operations such as homomorphic hashing, and naive one-way non-collision hash function (SHA-256). Figure 6 shows the aggregation time as a function of the number of SMs for the proposed scheme and [23] for aggregating 768 bits of data. In this regard, for ensuring data integrity support of 768 bits of data for 200 smart meters, the homomorphic hashing used in [23] takes 88.83 ms and naive one-way non-collision hash function (used in our proposed scheme) takes only 23.29 ms. The masking time is 0.77 ms for both schemes. It can be seen from Figure 6 that the aggregation time is significantly lower in the proposed scheme as compared to others. Hence, the proposed scheme can be used for efficient data aggregation in smart grids.

A. Computation Cost During Billing

In our billing phase, each smart meter needs to perform one symmetric-key encryption (AES-CBC) and one hash operation, which takes an additional $0.79 + 0.43 = 1.22$ ms. The TPA needs to perform one symmetric-key encryption and one symmetric-key decryption operation, which takes $0.31 + 0.42 = 0.73$ ms. The PS needs to perform one symmetric-key decryption and two hash operations which take $0.42 + 2 \times 0.12 = 0.66$ ms. Therefore, the overall computation cost for the billing phase is 2.61 ms.

B. Complexity of the Proposed Scheme

Table VII shows a detailed analysis of the complexity for our data aggregation scheme where N denotes the number of smart meters. The complexity value is given for both the smart meters and the TPA. Here, the operations conducted in each smart meter is of constant complexity and the smart meters can do their operations in parallel. For the TPA, the complexity increases linearly with the number of smart meters. Generally, while smart meters have limited computational capabilities, the TPA will have powerful computational resources, and thus the proposed scheme scales well with the size of the grid.

Table VII
COMPLEXITY FOR ONE ROUND OF SPATIAL DATA AGGREGATION

	SM_i	TPA
Addition	$\mathcal{O}(1)$	$\mathcal{O}(N)$
Hash	$\mathcal{O}(1)$	$\mathcal{O}(N)$
Messages in	$\mathcal{O}(1)$	$\mathcal{O}(N)$
Messages out	$\mathcal{O}(1)$	$\mathcal{O}(1)$

VII. CONCLUSION

This paper proposed a lightweight and privacy-friendly spatial data aggregation scheme for securely obtaining the power demand in a smart grid. Security of the proposed scheme is analyzed to confirm its robustness against known attacks. In addition, the privacy of individual meter readings is analyzed under the honest-but-curious adversary model. Performance analysis of the proposed scheme with existing data aggregation schemes shows that the proposed scheme has significantly lower computational cost as compared to other approaches and is hence the best option for smart grid environments.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Corporate Laboratory@University Scheme, National University of Singapore, and Singapore Telecommunications Ltd.

REFERENCES

[1] G. Wood, M. Newborough, "Dynamic energy-consumption indicators for domestic appliances: Environment, behavior and design," *Elsevier Energy Build*. 35(8), 821–841, 2003.

[2] P. Gope, and B. Sikdar, "Privacy-Aware Authenticated Key Agreement Scheme for Secure Smart Grid Communication," *IEEE Transactions on Smart Grid* DOI:10.1109/TSG.2018.2844403, 2018.

[3] A. Cavoukian, J. Polonetsky, C. Wolf, "SmartPrivacy for the smart grid: Embedding privacy into the design of electricity conservation," *Identity Inf. Soc*. 3(2), pp. 275–294, 2013.

[4] F.D. Garcï¿œa, B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," *in Proc. Workshop on Security and Trust Management*, 2010. pp 226–23, Athens, Greece.

[5] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," *in Proc. EUROCRYPT*, pp. 223–238, 1999.

[6] R. Lu, X. Liang, X.L Li, and X. Shen, "Eppa: an efficient and privacy-preserving aggregation scheme for secure smart grid communications,". *IEEE Trans. Parallel Distrib. Syst.* 23(9), 1621–1631, 2012.

[7] X. Liang, X. Li, R. Lu, X. Lin and X. Shen, UDP: usage based dynamic pricing with privacy perservation for smart grid, *IEEE Trans. Smart Grid*, 2013.

[8] M. Naehrig, K. Lauter and V. Vaikuntanathan, "Can homomorphic encryption be practical?", *in Proc. ACM Cloud Computing Security Workshop*, pp. 113-124, 2011.

[9] Y. Chia-Mu, C.-Y. Chen, S.-Y. Kuo, H.-C. Chao, "Privacy-preserving power request in smart grid networks," *IEEE Syst. J.*, 8(2), pp. 441–449, 2014.

[10] X. Liu, Y. Zhang, B. Wang, H. Wang, "An anonymous data aggregation scheme for smart grid systems,". *Secur. Commun. Netw.*, 7(3), 602–610, 2014.

[11] J. Zhang, L. Liu, Y. Cui, Z. Chen, " SP 2 DAS: self-certified PKC-based privacy-preserving data aggregation scheme in smart grid,". *Int. J. Distrib. Sens. Netw.*, pp. 1–11, 2013.

[12] Z. Sui, A. Alyousef, H. de Meer, "IAA: incentive-based anonymous authentication scheme in smart grids.," *In: Tiropanis, T.*, Vakali, A., Sartori, L., Burnap, P. (eds.) Internet Science. LNCS, vol. 9089, pp. 133–144. Springer, Heidelberg (2015).
[13] F. Li, B. Luo, "Preserving data integrity for smart grid data aggregation," *in Proc. IEEE SmartGridComm*, Tainan, Taiwan, pp. 366–371, 2012.
[14] F. Li, B. Luo, P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," *in Proc. IEEE SmartGridComm*, Gaithersburg, USA, pp. 327–332, 2010.
[15] C. Rottondi, M. Savi, G. Verticale, C. Kraub, "Mitigation of Peer-to-Peer Overlay Attacks in the Automatic Metering Infrastructure of Smart Grids," *Secur. Commun. Netw.*, 8, pp. 343–359, 2015.
[16] S. Cho, H. Li, H., B. Choi, "PALDA: Efficient Privacy-Preserving Authentication for Lossless Data Aggregation in Smarg Grids," *in Proc. IEEE SmartGridComm*, Venice, Italy, pp. 914–919, 2014.
[17] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, X. Shen, "EPPDR: An Efficient Privacy-Preserving Demand Response Scheme with Adaptive Key Evolution in Smart Grid," *IEEE Trans. Parallel Distrib. Syst.*, 25, 2053–2064, 2014.
[18] X-F Wang, Y. Mu, R-M Chen, "An efficient privacy-preserving aggregation and billing protocol for smart-grid," *Secur. Commun. Netw.*, 9, pp. 4536–4547, 2016.
[19] H. J. Jo, I. S. Kim and D. H. Lee, "Efficient and Privacy-Preserving Metering Protocols for Smart Grid Systems," in *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1732-1742, doi: 10.1109/TSG.2015.2449278, May 2016.
[20] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart grid,"*in Proc. Privacy Enhanced Technology Symposium*, pp. 175–191, 2011.
[21] E. Shi, R. Chow, T.-h. H. Chan, D. Song, and E. Rieffel, "Privacyp-reserving aggregation of time-series data," *in Proc. NDSS Symposium* 2011.
[22] G. Danezis, C. Fournet, M. Kohlweiss, and S. Zanella-Beguelin, "Smart ´ Meter Aggregation via Secret-sharing," *in Proc. ACM Workshop on Smart Energy Grid Security*, pp. 75–80, 2013.
[23] F. Knirsch et al. "Error-resilient Masking Approaches for Privacy Preserving Data Aggregation," *IEEE Trans. Smart Grid,* DOI 10.1109/TSG.2016.2630803, 2016.
[24] H. Mohammed et al. ""Efficient Privacy-Preserving Data Collection Scheme for Smart Grid AMI Networks"," in *in Proc. IEEE GLOBECOM*, Washington DC, USA, 2016.
[25] W. Jia, H. Zhu, Z. Cao, X. Dong, and C. Xiao, "Human-Factor-Aware privacy-preserving aggregation in smart-grid," *IEEE Systems Journal*, vol. 8(2), 2014.
[26] Pbc library. Tech. rep. http://crypto.standford.edu/pbc/, (accessed on 16 April 2017).
[27] Oracle Technology Network, Java Cryptography Architecture (JCA). [Online]. Available: http://docs.oracle.com/javase/6/docs/technotes/guides/crypto/CrypoSpec.html, (accessed on Apr. 20, 2017).
[28] libpaillier-0.8. Tech. rep. http://hms.isi.jhu.edu/acsc/libpaillier/ (accessed on 16 April 2017).
[29] K. Ouafi and R. C.-W. Phan, Privacy of recent RFID authentication protocols, in: Information Security Practic and Experience, Springer, pp. 263-277, 2008.
[30] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. Shen, "A lightweight message authentication scheme for smart grid communications," *IEEE Trans. Smart Grid,* vol. 2, no. 4, pp. 675–685, Dec. 2011.
[31] A. Abdallah; X. Shen, "A Lightweight Lattice-based Homomorphic Privacy-Preserving Data Aggregation Scheme for Smart Grid," in *IEEE Transactions on Smart Grid*, vol.PP, no.99, pp.1-1 doi: 10.1109/TSG.2016.2553647, 2016.
[32] P. Gope, J. Lee, and T~Q. S. Quek, "Lightweight and Practical Anonymous Authentication Protocol for RFID Systems Using Physically Unclonable Functions," *IEEE Transactions on Information Forensics and Security,* vol. 13(11), pp. 2831-2843, 2018.
[33] Atmel's family of smart power meters. http://www.atmel.com/products/smart-energy/power-metering/ (accessed on 28 May 2017).
[34] D. Koo, Y. Shin, J. Hur, "Privacy-Preserving Aggregation and Authentication of Multi-Source Smart Meters in a Smart Grid System," *Applied Sciences*, vol. 7, pp. 1007, 2017.
[35] E. Vahedi, M. Bayat, M. Pakravan, M. Aref, "Secure ECC-based privacy preserving data aggregation scheme for smart grids," *Computer Networks*, vol. 129, no. 1, pp. 28-36, 2017.
[36] Y. Zhang, J. Zhao, D. Zheng, "Efficient and privacy-aware power injection over AMI and smart grid slice in future 5G networks," *Mob. Inf. Syst.*, 2017.
[37] U. Baloglu and Y. Demir, "Lightweight privacy-preserving data aggregation scheme for smart grid metering infrastructure protection," *International Journal of Critical Infrastructure Protection*, vol. 22, pp. 16-24, September 2018.
[38] P. Gope, and B. Sikdar, "An Efficient Data Aggregation Scheme for Privacy-Friendly Dynamic Pricing-based Billing and Demand-Response Management in Smart Grids," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3126-3135, 2018.
[39] W. Han Y.Xiao,"Privacy Preservation for V2G Networks in Smart Grid: A Survey," *Computer Communications*, vol. 91, pp. 17-28, 2016.
[40] S. Ozdemir Y. Xiao, "Secure Data Aggregation in Wireless Sensor Networks: A Comprehensive Overview," *Computer Networks*, vol. 53, pp. 2022-2037, 2009.

Prosanta Gope (M'18) received the M.Tech. degree in computer science and engineering from the National Institute of Technology (NIT), Durgapur, India, in 2009, and the PhD degree in computer science and information engineering from National Cheng Kung University (NCKU), Tainan, Taiwan, in 2015. He is currently working as a Lecturer in the department of Computer Science (Cyber Security) at the University of Hull. Prior to this, Dr. Gope was working as a Research Fellow in the department of Computer Science at National University of Singapore (NUS). His research interests include lightweight authentication, authenticated encryption, access control system, security in mobile communication and cloud computing, lightweight security solutions for smart grid and hardware security of the IoT devices. He has authored over 50 peer-reviewed articles in several reputable international journals and conferences, and has four filed patents. He received the Distinguished Ph.D. Scholar Award in 2014 from the National Cheng Kung University, Tainan, Taiwan. He currently serves as an Associate Editor of the IEEE Internet of Things Journal, IEEE SENSORS JOURNAL, the SECURITY AND COMMUNICATION NETWORKS and the MOBILE INFORMATION SYSTEMS JOURNAL.

Biplab Sikdar (S'98-M'02-SM'09) received the B.Tech. degree in electronics and communication engineering from North Eastern Hill University, Shillong, India, in 1996, the M.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 1998, and the Ph.D. degree in electrical engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 2001. He was on the faculty of Rensselaer Polytechnic Institute from 2001 to 2013, first as an Assistant and then as an Associate Professor. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His research interests include computer networks, and security for IoT and cyber physical systems. Dr. Sikdar is a member of Eta Kappa Nu and Tau Beta Pi. He served as an Associate Editor for the IEEE Transactions on Communications from 2007 to 2012. He currently serves as an Associate Editor for the IEEE Transactions on Mobile Computing.