An Efficient Privacy-Friendly Hop-by-Hop Data Aggregation Scheme for Smart Grids

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

***Abstract*—The fine-grained and large number of measurements collected by smart meters can be used to reconstruct consumer behavior, and thus their widespread deployment for the modernization of electricity distribution networks has been associated with privacy concerns. This paper proposes an efficient and privacy-friendly hop-by-hop data aggregation scheme and a billing solution for smart grid systems. In our approach, hop-by-hop communication is utilized for transmitting usage reports of the smart meters. From the outcome of the security and performance analyses we can argue that our proposed scheme is secure and computationally more efficient, as compared to the other solutions.**

***Index Terms*—Privacy, security, hop-by-hop, data aggregation, computational efficiency, smart grids.**

I. Introduction

The last few decades have seen the emergence of various technologies that have facilitated the modernization of electricity grids into smart grids. A smart grid can be regarded as a modern grid system that aims to provide high capacity, efficiency, and reliability by combining emerging cyber-physical technologies into current electricity networks. The Internet of Things (IoT) serves as an enabling technology for smart grids with devices in the grid operating as connected objects [1-2]. A large number of devices/sensors autonomously report their information to the grid infrastructure by using information and communication technology (ICT). Thus, smart grids can be viewed as a modernization of power grids with advanced information and communication infrastructure. However, this interconnection of grid technology with ICT leads to various security and privacy issues [3]. One of the primary goals of smart grids is to improve the utilization of resources, particularly in view of the volatility of both the power supply and demand. Thus, these grids require smart devices that monitor the power consumption and generation in the grid and can report their measurements in real-time over a network. To ensure reliable and cost-effective demand response management between the consumers and the generators in a power grid, a smart grid utilizes the smart metering infrastructure. Fine granular readings of power consumption generated by the metering infrastructure are transmitted to the power supplier and/or the grid operator. These power consumption profiles are used to enable a precise prediction of power demand that can be used for managing power production. However, these profiles also allow the creation of a usage profile of a specific person, household, or a company, and that may cause several privacy issues. Such profiles can be analyzed for extracting the personal behavior of users or to evaluate the business activity in enterprises. For instance, a long-term analysis of the consumers' data can reveal private information related to their daily routines that can be used by an outside adversary or a third-party company to deduce the consumers' living habit and lifestyle, which may introduce serious privacy issues. Therefore measures have to be taken to ensure the required level of privacy by considering all the aforesaid issues.

This article proposes an efficient and privacy-friendly data aggregation scheme for smart grids by using hop-by-hop communication. The proposed scheme only uses computationally inexpensive operations such as hash operations. Thus, the proposed scheme is well suited for the resource limited smart meters. The rest of the paper is organized as follows. The related work and motivation for the paper is described in Section II. In Section III, we present the underlying system model for smart grids and the security goals of the proposed scheme. In Section IV, we present the proposed scheme. A discussion on the security and the performance evaluation of the proposed scheme is presented in Section V. In Section VI, we formally analyze the privacy of the proposed scheme. Finally, the conclusions are presented in Section VII. In Table I, we define all the important symbols and cryptographic operations which are frequently used in this paper.

II. Related Work and Motivation

In recent years, numerous privacy-friendly data aggregation schemes have been proposed for addressing various privacy issues in smart grids (as mentioned above). For instance, Lu et al. introduced a data aggregation scheme [4] by using Paillier encryption [5]. However, this results in high computational overhead on the resource limited smart-meters. Subsequently, Liang et al. proposed a new scheme using the concept of fully homomorphic encryption [6]. However, Naehring et al. have shown that a fully homomorphic technique is hard to implement [7]. Therefore, the scheme presented in [6] is regarded as an unrealistic one. Yu et al. proposed a new scheme to protect an individual's usage profile [8] by using ring signatures. However, in their scheme, if the size of the ring is increased, then the computational cost of the proposed scheme will also become higher. Liu et al. [9] proposed an aggregation scheme based on blind signatures [9]. However, this scheme cannot protect the privacy of the consumers'

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

usage data profile [10]. Zhang et al. proposed a signature-based scheme [10] and Sui et al. designed an incentive-based data aggregation scheme for smart grids [11]. Both of these schemes are designed with the assumption of an anonymous network which can hide sources of usage reports. These schemes, however, do not support data integrity of the usage data. Alharbi et al. [32] proposed a data aggregation scheme using the bilinear pairing technique. However, the scheme fails to ensure some of the important security properties such as data integrity and consumer's privacy. Li et al. proposed a new solution for data aggregation for smart grids. In this context, they utilize the concept of hop-by-hop communication [13]. However, the scheme presented in [13] does not support message authentication. As a consequence, a malicious smart meter or an adversary may try to falsify the data for causing an inaccurate aggregation outcome. Li and Luo have proposed a homomorphic signature algorithm [35] based on a short signature scheme using bilinear pairing, which can support data integrity. Yang and Li [36] proposed a anomaly detection scheme based on dynamic grouping and data re-encryption using ElGamal encryption scheme. However, in the above schemes, the smart meters reveal their identity. Therefore, the above schemes cannot ensure consumer's privacy. Recently, Knirsch presented a masking-based solution where the concept of homomorphic hashing is used for validating the shared secrets [14]. However, it can be shown that the data aggregation scheme presented in [14] is vulnerable to collusion attacks. In this case, if the aggregator (DC in [14]) 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. In [15], Mohammed et al. introduced another hop-by-hop data aggregation scheme, where during data aggregation, each smart meter has to select n proxies and add masking values to their meter readings. Proxies remove the masking values added to the meters' readings 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, which may cause an inaccurate aggregated result. In [36], a dynamic-pricing-based billing solution is proposed, where the aggregator (TPA) needs to verify the legitimacy of each smart meter before obtaining their billing data. In this case, the computational complexity will linearly increase with the number of smart meters. A similar problem can also be seen in [37]. Recently, Gope and Sikdar proposed an authenticated key-agreement scheme [38], which can detect any physical tampering attempt on the smart meter. However, this scheme does not address data aggregation and billing.

A. Problem Statement and Motivation

Conceive that there is a region/neighborhood with several apartment blocks or houses. Each apartment block or house has a set of one or more units and they are individually equipped with a smart meter. In order to maintain proper balance between power generation and power consumption, the grid requires to know the aggregated electricity usage data for the entire region on a regular basis. Now, for the correctness of the aggregated usage report, the aggregator has to verify the legitimacy of each individual smart meter and the integrity of their readings. However, this will result in a very large burden on the aggregator, especially when the aggregator needs to handle a large number of smart meters. To address this issue, Li et al. [13] and Mohammed et al. [15] proposed two possible frameworks where smart meters transmit the usage reports in a hop-by-hop way. However, the frameworks have several weaknesses. For instance, the solution proposed in [13] is malleable: given a cipher and a public key, an adversary can generate meaningful usage data. Consequently, a malicious smart meter may try to falsify the usage data, which will result a wrong aggregated output. Furthermore, in many of the existing works [16-21], it is assumed that there is a secure channel between the third-party aggregator and smart meters, which is a strong assumption.

Table I
NOTATIONS AND CRYPTOGRAPHIC FUNCTIONS

Symbol	Definition
PS	Power Supplier
HAN	Home are network
SM	Smart meter
TPA	Third-party aggregator
PID_i	Pseudo identity of SM_i
ID_{SM_i}	Identity of smart meter SM_i
$kh_{i,i+1}$	Shared integrity key between SM_i and SM_{i+1}
$E_k[x]$	Plaintext x encrypted using key k
Sign	Signing algorithm
Ver	Verification algorithm

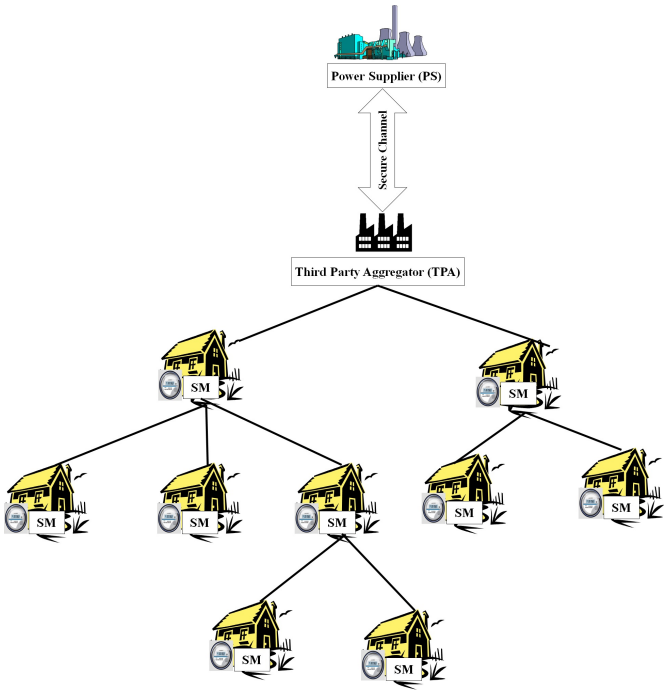

Figure 1. System model for the proposed hop-by-hop data aggregation scheme.

B. Our Contribution

In this paper, we propose an efficient and privacy-friendly solution for addressing all the above issues. In our approach, smart meters transmit usage reports through hop-by-hop communication. Even though some existing approaches can accomplish similar security features, our scheme has lower computational cost as shown by our performance analysis and experiments. Smart meters do not have to perform any computationally expensive operations (such as inefficient Paillier encryptions) during the data aggregation process. Hence, the proposed data aggregation scheme is suitable for the resource constrained devices in smart grids.

III. System and Adversarial Model, and Security Goals

In this section, we first define our system for privacy-friendly data aggregation in smart grids. Next, we present the underlying adversary model. This section concludes with the desired security goals of the proposed scheme.

A. System Model

In our system model, we adopt the concept of data aggregation trees [11], which supports the usage data to be transmitted by hop-by-hop communication, as shown in Fig. 1. The system model consists of three major entities: the power supplier (PS), the third-party aggregator (TPA), and a list of smart meters (SMs). Here, a smart meter is an electronic device that records the consumption of electric energy and communicates the information to the electricity supplier for monitoring and billing. Smart meters typically record energy hourly or more frequently, and report at least daily [33]. Smart meters enable two-way communication between the meter and the TPA. In the system model, the PS is responsible for arranging and supplying electricity to a list of home-area-networks (HANs). Each HAN is equipped with a SM. The TPA periodically aggregates the electricity consumption of a group of HANs in a locality and helps the owners of the HANs to adjust their consumption according to the current loading conditions (e.g., through demand side management) and also inform the current demand conditions to the PS in order to help with supply-demand management. The consumption information allows the PS to optimally control its dispatchable generation, and conduct short and long term trades in the energy market. In this way, the TPA plays a crucial role in balancing the power production and demand. In our system model, we assume that the TPA and the PS communicate through a secure channel. Each HAN is composed of a SM, which is assumed to be tamper-proof. The SMs form a tree topology and send the consumers' energy usage report to the TPA through hop-by-hop communication.

B. Adversarial Model

In our adversarial model the PS is considered as a trusted organization (e.g., operated by the government, such as Singapore Power in Singapore and National Grid in United Kingdom). On the other hand, in our adversarial model the TPA is considered as a semi-honest (i.e., honest-but-curious) entity, who is interested in obtaining the usage data of each HAN and subsequently may try to sell the usage information to another company, e.g., for marketing materials for home appliances. The TPA is operated by a private company whose main responsibility is to assist the PS. Besides, in our system model any SM can be the adversary and be interested in obtaining the usage report of another SM from a different HAN. An outsider may also try to pretend as a legitimate SM or the TPA to send data under its name.

C. Security Goals

- **Message Authentication:** In general, usage reports from each SM pass through the insecure wired and wireless links of the communication network. Therefore, before aggregating any data, the aggregator needs to validate whether the report has been received from a legitimate source or not. This will prevent any inaccurate aggregation result.
- **Usage Data Privacy:** Ensuring privacy in the end-to-end communication is an imperative security goal. For example, if an adversary can know the power consumption data from a HAN, then he/she can determine its occupancy. This information can be used by robbers to determine the best day or time to break into a home. Therefore, the electricity consumption data is required to be kept secret from any third party for protecting the privacy of the customer.
- **Usage Data Integrity:** To avoid any inaccurate data aggregation result, the TPA must validate the integrity of the usage report received from each SM of a HAN. On the other hand, during data aggregation, the TPA also needs to check the integrity of the relevant information received from the PS.

IV. Proposed Scheme

In this section, we propose our efficient and privacy-friendly data aggregation scheme for smart-grids, which consists of three phases: *initialization*, *hop-by-hop data aggregation*, and *secure billing*. In our tree model, each parent may have n children. The responsibility of the parent node is to accumulate and validate the usage reports of its child nodes and subsequently aggregate these usage data along with its own reading and send the aggregated result to its parent. In order to simplify the description, we assume that each SM (e.g., SM_i) only has one child (SM_{i-1}). However, it can be easily extended to other tree constructions. The proposed scheme requires the formation of a topology that constructs a tree topology with the TPA at the root and the SMs as children. Algorithms for tree construction have been widely investigated in literature, specially in the context of wireless sensor networks [29-30]. These algorithms may also be used for constructing the topology for the proposed scheme. The tree construction problem is in fact simpler in our case since the devices are static and do not have a power constraint (i.e., they are not battery powered). For example, in case of individual houses on a street with a SM in each house, the

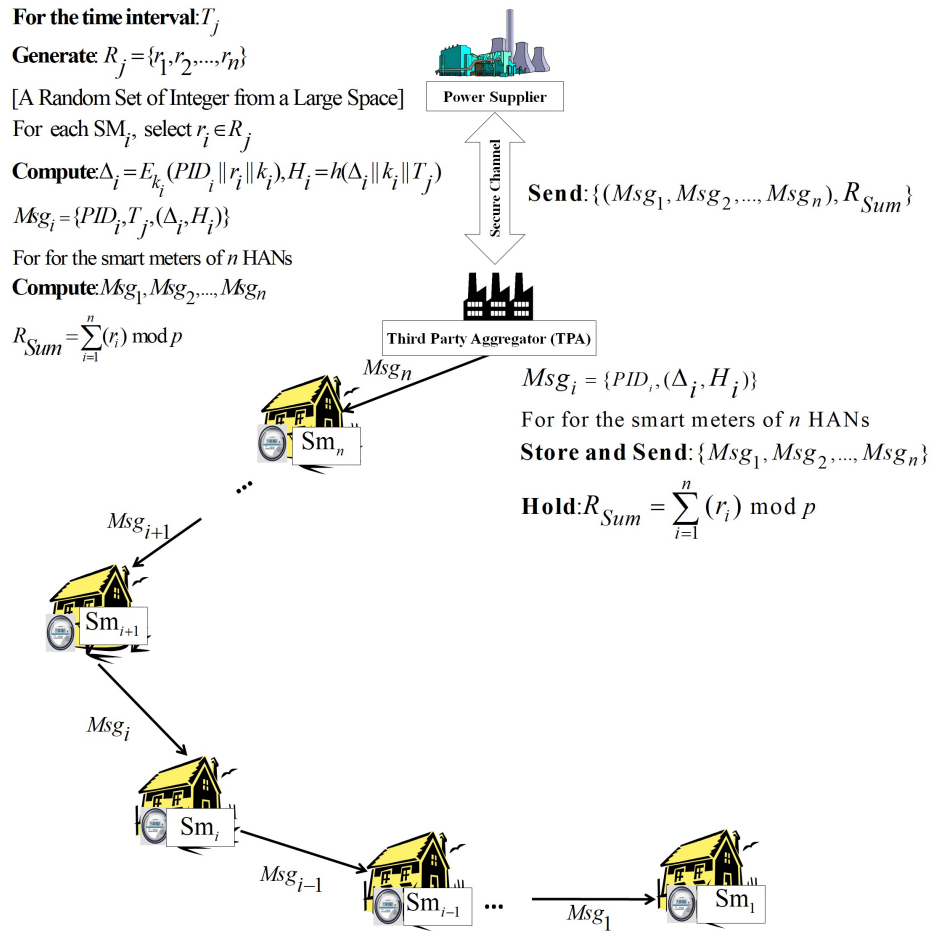

Figure 2. Step AG1 of the proposed data aggregation scheme.

algorithms from [29-30] may be directly applied without the energy constraints (or setting the energy conditions at each SM to be the same). Similarly, in apartment blocks with multiple floors and one or more apartments per floor, the first SM (or a randomly chosen SM) from a floor can act as the parent node for all the SMs in the floor immediately above it. Since existing techniques can be readily used for creating the hop-by-hop topology, we do not focus on it in this paper.

In the initialization phase of the proposed scheme, the interrelated SMs establish their common secret keys. After that, during the data aggregation process, the SMs send their usage reports on a regular basis (e.g., every 15-30 minutes) using the hop-by-hop data aggregation process.

A. Initialization

Conceive that there are n HANs in a locality that consume their electricity from the PS. At the time of meter installation of a home HAN_i, the PS randomly generates a pseudo identity PID_i and a secret key k_i and assigns them to the SM of HAN_i. Here, we also assume that each SM is equipped with a tamper-resistant black box [17]. The black box contains a key pair (PK, SK). Any other party has access to the public key PK. However, the secret key SK is stored within the black box and is never disclosed or changed. To ensure secure communication with its neighboring SMs, each smart meter SM_i executes a key establishment protocol (e.g., the protocol proposed in [22]) with its PK and SK. Consequently, the keys $kh_{i-1,i}$ and $kh_{i,i+1}$ are shared between SM_i and SM_{i-1}, and SM_i and SM_{i+1}, respectively. The smart meters can also update the keys shared with their neighbors by executing the key establishment protocol [22]. Now, the TPA generates the key pair (SK-TPA, PK-TPA) and publishes PK-TPA to others. Similarly, the PS generates the key pair (SK-PS, PK-PS) and publishes PK-PS to others. The TPA and the PS use their secret keys (SK-TPA, SK-PS) to generate signatures. Anyone who knows their public keys (PK-TPA, PK-PS) can verify the signature.

B. Hop-by-hop Data Aggregation Scheme

Our data aggregation scheme consists of the following steps:

Step AG1: To maintain balance between the power production and demand, the PS periodically (say, every 15 or 30 minutes) requires to obtain the usage reports 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 $\{0, \cdots, p-1\}$ of a uniform distribution, where $p \gg \sum_{i=1}^{n} M_i$, where M_i is the meter reading of SM_i. Next, for each smart meter SM_i, the PS picks the random integer $r_i \in R_j$, calculates $\Delta_i =$

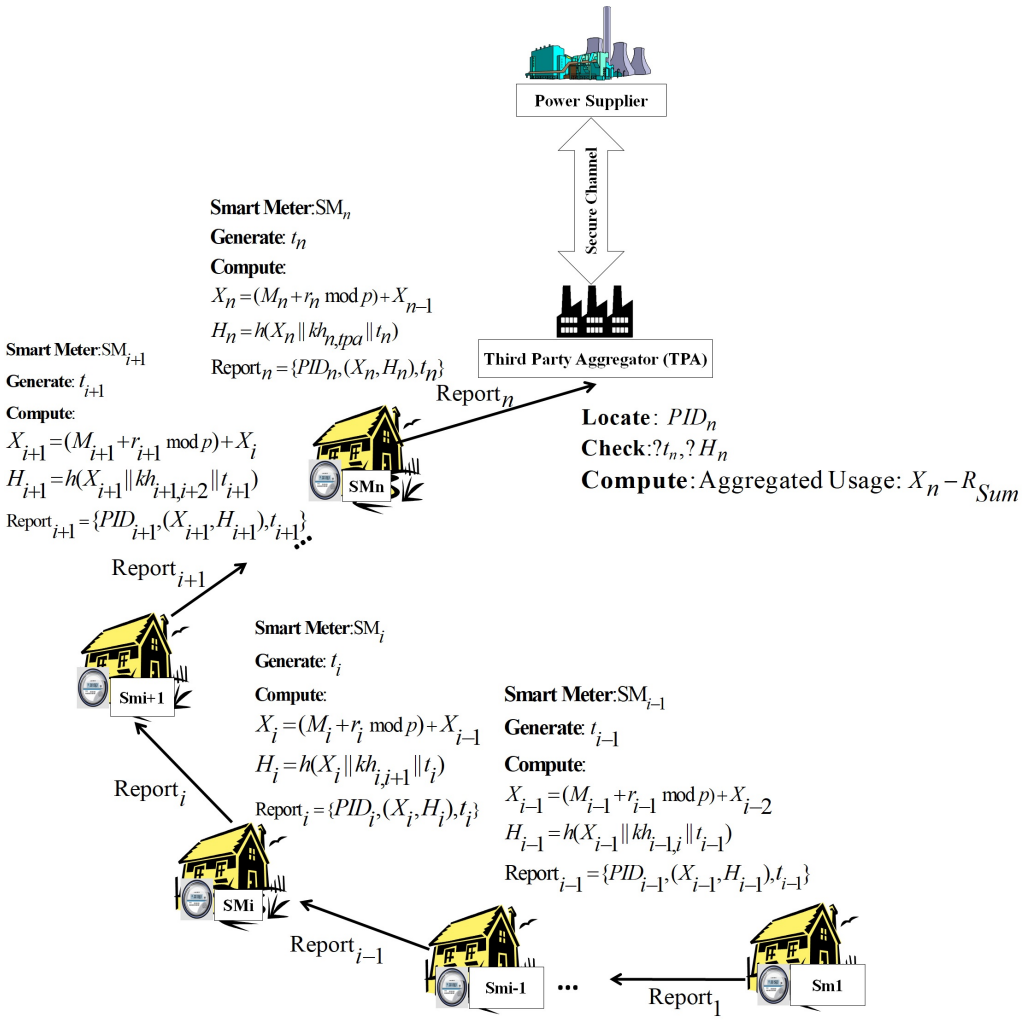

Figure 3. Step AG2 of the proposed data aggregation scheme.

$E_{k_i}(PID_{SM_i}||r_i||k_i)$ and $H_i = h(\Delta_i||k_i||T_j)$, and composes a message $Msg_i = \{PID_i, T_j, (\Delta_i, H_i)\}$. In this way, the PS derives $\{Msg_1, Msg_2, \cdots, Msg_n\}$ for each of the n smart meters. The PS also computes $R_{Sum} = \sum_{i=1}^{n}(r_i \mod p)$ and finally sends $\{Msg_1, Msg_2, \cdots, Msg_n\}$ and R_{Sum} to the TPA through the secure channel. Note that for more efficient operation of the above data aggregation scheme, the PS can pre-compute $\{Msg_1, Msg_2, \cdots, Msg_n\}$ and R_{Sum} for several sessions and send them to the TPA. Next, at time interval or session T_j, the TPA distributes $\{Msg_1, Msg_2, \cdots, Msg_n\}$ to each smart meter. In this regard, each smart meter SM_i helps its neighbors to obtain their respective messages. Next, upon receiving Msg_i, smart meter SM_i first checks the time interval T_j (for detecting replay attacks) and also computes and checks H_i. If the verification is successful, SM_i decrypts Δ_i and obtains the random integer r_i. The details of the step AG1 are depicted in Figure 2.

Step AG2: After obtaining the random integer r_i, SM_i generates a timestamp t_i and computes its blinded measurement $X_i = (M_i + r_i \mod p) + X_{i-1}$ and $H_i = h(X_i||kh_{i,i+1}||t_i)$, where X_{i-1} denotes the blinded measurement received from its neighbor SM_{i-1}. Next, SM_i composes $\mathrm{Report}_i = \{PID_i, X_i, H_i, t_i\}$ and sends it to its parent SM_{i+1}. Upon receipt of Report_i, smart meter SM_{i+1} first validates the time stamp t_i and also computes and validates H_i using the secret key $kh_{i,i+1}$. If the validation is successful, SM_{i+1} generates a timestamp t_{i+1} and calculates the blinded measurement $X_{i+1} = (M_{i+1} + r_{i+1} \mod p) + X_i$ and $H_{i+1} = h(X_{i+1}||kh_{i+1,i+2}||t_{i+1})$. Finally, smart meter SM_{i+1} composes $\mathrm{Report}_{i+1} = \{PID_{i+1}, X_{i+1}, H_{i+1}, t_{i+1}\}$ and sends it to its parent SM_{i+2}. Continuing in this way, upon receipt of usage report $\mathrm{Report}_n = \{PID_n, X_n, H_n, t_n\}$ from SM_n, the TPA first checks its validity. If it is valid, the TPA calculates $X_n - R_{Sum}$ to obtain the *aggregated usage* data of the n HANs. The details of step AG2 are depicted in Figure 3.

Next, the TPA sends the aggregated usage data to the PS. If the PS finds a mismatch between the energy production and consumption, it takes the necessary steps to increase production. In addition, the PS may employ demand-side management in order to modulate consumer behavior. Towards this end, the PS first composes the instructions or information for demand-side management (denoted by Γ) and conveys it to the TPA for dissemination to the consumers. Next, the

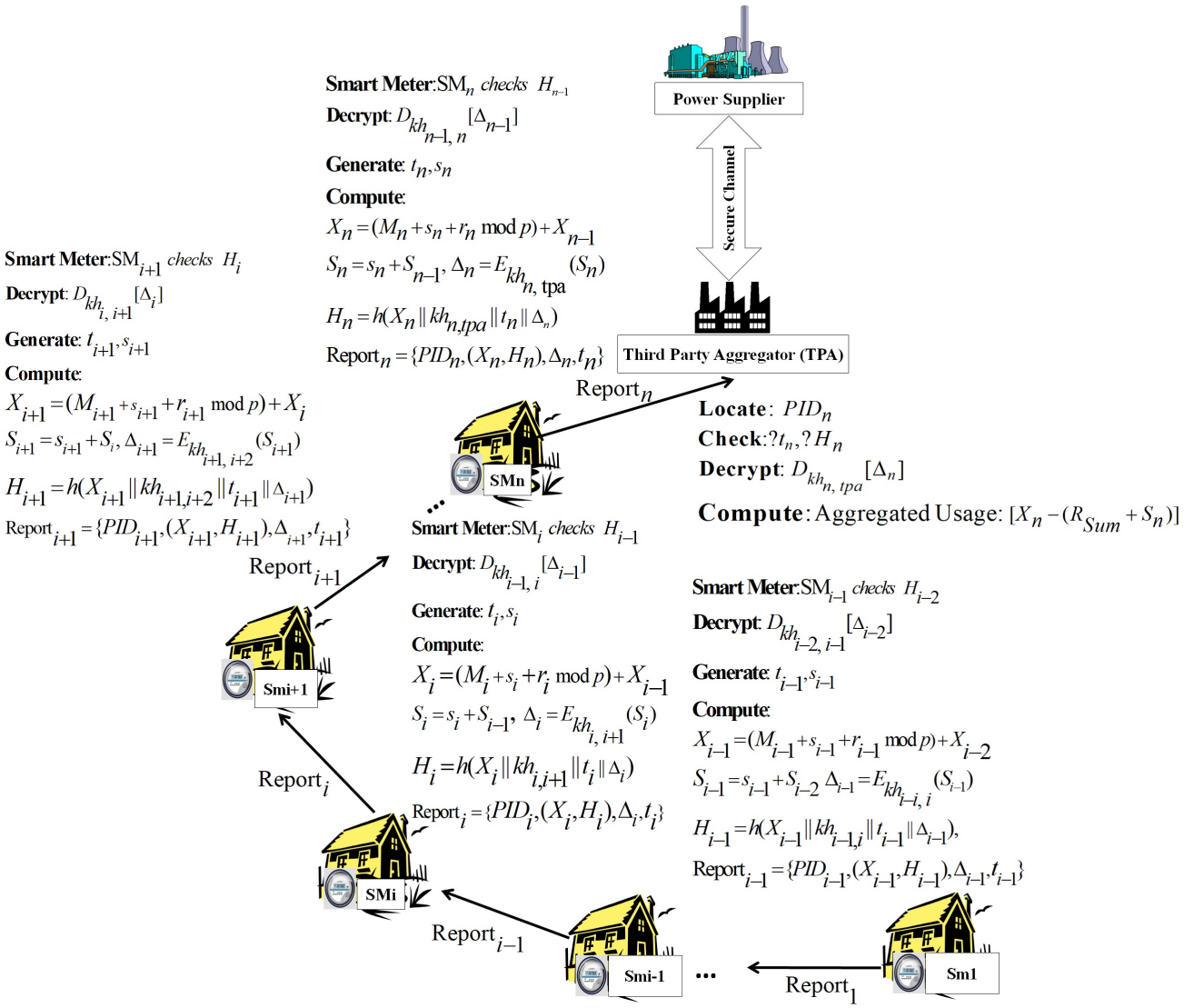

Figure 4. Revised step AG2 for addressing collusion attacks between PS and TPA.

TPA generates a valid signature $f = \textbf{Sign}(h(\Gamma, t), \text{SK-TPA})$ and subsequently broadcasts the instructions and the signature (Γ, f, t) to all the SMs. When a SM receives the usage instructions, it first checks the timestamp t and checks $\textbf{Ver}(h(\Gamma, t), f, \text{PK-TPA})$. If they are valid, then the SM informs its owner to adjust their usage; otherwise, it just ignores the instruction and signature. Here, **Sign** and **Ver** denote the signing algorithm and verification algorithm of a secure public key signature scheme [27-28].

It should be noted that 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), each smart meter needs to do ping tests with its neighbors at regular intervals. In case smart meter SM_i does not receive any response from its neighbor SM_{i-1}, then SM_i informs the PS with the help of the TPA through hop-by-hop communication. 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. Now, after obtaining the masking value r_i, each smart meter SM_i picks a random number s_i (called its "share") drawn uniformly from a cryptographic pseudo random number generator and then the SM_i adds this share to its measurement value M_i yielding X_i, i.e., $X_i = M_i + s_i + r_i \mod p$. Hereafter, SM_i adds s_i to the accumulated share value S_{i-1} that it has received from SM_{i-1} and calculates $S_i = s_i + S_{i-1}$ and then encrypts S_i, i.e., $\Delta_i = E_{kh_{i,i+1}}(S_i)$ and also computes $H_i = h(X_i||kh_{i,i+1}||t_i||\Delta_i)$, where t_i denotes the times-

tamp generated by SM_i. Next, smart meter SM_i composes a message $\mathrm{Report}_i = \{PID_i, (X_i, H_i), \Delta_i, t_i\}$ and sends it to its next adjacent neighbor SM_{i+1}. Upon arrival of Report_i, SM_{i+1} checks the timestamp t_i and the key-hash output H_i. After successfully validating these parameters, SM_{i+1} decrypts Δ_i and obtains S_i. After that, SM_{i+1} generates a timestamp t_{i+1} and a random share s_{i+1} and then adds s_{i+1} to the accumulated share value S_i that it has received from SM_i. 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 composes Report_n and sends it directly to the TPA. The TPA computes $[X_n - (R_{Sum} + S_n)]$ which gives the desired aggregated load. The details of the revised step AG2 are depicted in Figure 4.

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 its neighbor SM_{i+1}, SM_i also updates $\beta_i = M_i^j + \beta_i$ and stores β_i into 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 computes $E_{k_i}[\beta_i]$, $\nu_i = h(E_{k_i}[\beta_i]||k_i||t)$, and composes a message $Bill_i = \{PID_i, E_{k_i}[\beta_i], \nu_i, t\}$ and sends it to the PS through hop-by-hop communication via the TPA. Upon receiving $Bill_i$, the PS first checks the timestamp t and ν_i. If they are valid then the PS defines an acknowledgment $\mathcal{ACK}_i$ and generates a timestamp t^*, and a valid signature $\lambda = \mathbf{Sign}(h(\mathcal{ACK}_i, t^*), \text{SK-PS})$ and subsequently sends $(\mathcal{ACK}_i, \lambda, t^*)$ to SM_i through the TPA using hop-by-hop communication. When SM_i receives the acknowledgment $\mathcal{ACK}_i$, it first checks the timestamp t^* and $\mathbf{Ver}(h(\mathcal{ACK}_i, t), \lambda, \text{PK-PS})$. If they are valid, then SM_i informs its owner and sets β_i to 0; otherwise, it requests the PS for the acknowledgment.

V. Discussion

In this section, we first analyze the security of proposed solution. In this regard, we consider all the security goals that we listed in Section II. Subsequently, we demonstrate that our proposed solution incurs reasonable computational overhead that is acceptable even for the resource constrained entities in smart grids.

A. Security Analysis

- **Message Authentication:** In the proposed hop-by-hop data aggregation scheme, when smart meter SM_i receives the usage report from its neighbor SM_{i-1}, then SM_i first checks the timestamp t_{i-1}. If the timestamp is valid then SM_i computes $H^*_{i-1} = h(X_{i-1}||kh_{i-1,i}||t_{i-1})$ and verifies whether H^*_{i-1} is equal to H_{i-1} or not. If they are equal, then SM_{i-1} passes the authentication process. In this way, all the smart meters authenticate their neighbor before aggregation and finally when the TPA receives the usage report from its neighbor SM_n, the TPA computes $H^*_n = h(X_{i-1}||kh_{n,tpa}||t_n)$ and checks whether H^*_n is equal to H_n or not. If so, the TPA calculates $X_n - R_{Sum}$ and obtains the *aggregated usage* data of the n HANs. In this way, the proposed hop-by-hop data aggregation scheme ensures the message authentication property. Furthermore, in the proposed scheme, if an attacker attempts to execute a replay attack, then the receiving end will be able to comprehend such activities by using the timestamps.
- **Usage Data Confidentiality:** The amount of electricity usage in each HAN (e.g., HAN_i) is blinded with a random integer (r_i). Hence, when the neighbor aggregator (HAN_{i+1}) receives the usage report Report_i, it can only see the blinded value of the usage data of the HAN. Similarly, when the TPA calculates $X_n - R_{Sum}$, it can only know the summation of the usage data of a group of HANs. In this way, we protect the details of the electricity consumption of each HAN. On the other hand, in the *secure billing phase*, the aggregated usage data for a month is encrypted $(E_{k_i}[\beta_i])$ and sent to the PS. Therefore, no one except the PS can know the value of β_i.
- **Usage Data Integrity:** In the proposed scheme, when a smart meter SM_i receives the usage report from its neighbor SM_{i-1}, then it checks whether the message it has received is the same as that sent by SM_{i-1}. In this regard, SM_i computes $H^*_{i-1} = h(X_{i-1}||kh_{i-1,i}||t_{i-1})$ by using the shared secret key $kh_{i-1,i}$ and then verifies whether H^*_{i-1} is equal to H_{i-1} or not. This approach helps to detect any manipulation of the aggregated usage data. Furthermore, in the *secure billing phase* we ensure the integrity of the aggregated usage data for a month, β_i, by using the parameter ν_i. Therefore, if an adversary attempts to tamper with $E_{k_i}[\beta_i]$, then the PS can comprehend that.

B. Performance Evaluation

To manifest the advantages of our proposed hop-by-hop data aggregation scheme, we now compare the proposed scheme with other existing hop-by-hop data aggregation schemes [13] and [15] and a recently proposed lightweight masking based scheme [14] for smart grids. We also show that the proposed scheme is efficient and well suited for the computationally resource-limited smart grid devices (such as smart meters). Now, for analyzing the performance of the proposed scheme, particularly on the security front, our scheme has been compared with the protocols of [13], [14], and [15] (shown in Table II), by considering all the security goals listed in Section II. From Table II we see that the proposed scheme can ensure *all* the security goals listed in Section II. In contrast, the hop-by-hop data aggregation schemes presented in [13] and [15] and the masking based scheme presented in [14] cannot ensure *message authentication*. Hence, a dishonest or fake smart-meter can falsify the data, which will cause an inaccurate aggregation result. In addition, the scheme presented in [14] is vulnerable to collusion attacks. Furthermore, unlike the proposed scheme, the schemes presented in [13], [14] and [15] do not support billing.

Table II
PERFORMANCE BENCHMARKING BASED ON SECURITY PROPERTIES

Security Properties	Li et al. [13]	Knirsch et al.[14]	Mohammed. et al. [15]	Proposed Scheme
Message Authentication	No	No	No	Yes
Usage Data Confidentiality	Yes	Yes	Yes	Yes
Usage Data Integrity	Yes	Yes	No	Yes
Resilience Against Collision Attacks	Yes	No	Yes	Yes
Billing Support	No	No	No	Yes

Next, we compare the aggregation time of the schemes, which includes the computation and the report transmission time. In [13], all SMs need to do homomorphic encryption, where the data encryption computations of the SMs can be calculated in parallel. We assume that the number of SMs is n, the communication time between any two neighbor SMs (denoted by T_c) is fixed, and the height of the aggregation tree is no less than $\log_q n$, where q denotes the number of children of each node in the tree. Based on these assumptions, the hop-by-hop data aggregation scheme presented in [13] has an aggregation time of $T(q) = T_h + T_c \log_q n$, where T_h denotes the homomorphic computation time. In [15], a smart meter needs to decrypt all the reports received from its children. Therefore, the hop-by-hop data aggregation scheme presented in [15] has an aggregation time of $T(q) = qT_{dec} + T_c \log_q n$. On the other hand, the aggregation time in the proposed hop-by-hop data aggregation scheme is $T(q) = T_{dec} + (qT_a + T_c) \log_q n$, where T_{dec} denotes the AES decryption computation time in Step AG1, and T_a denotes the message authentication and hash operation time.

Now, for analyzing the performance of the proposed scheme with respect to [13], [14] and [15], we conducted simulations of the cryptographic operations in all the 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 SP as per the scheme). To simulate a smart meter, we used a single core 798 MHz CPU and 256 MB of RAM, which is similar to the computational capability of real SMs [26]. The simulations used the JCE library [23] and the Paillier library libpaillier-0.8 [24] to evaluate the execution time of different cryptographic operations used in the proposed scheme, [13], [14] and [15]. The form of the usage report of SM_i is $\{PID_i, X_i, H_i, t_i\}$ whose size is taken to be 576 bits. Based on the simulations, the mean values of T_h, T_a, and T_{dec} are 22.69 ms, 0.0167 ms, and 0.021 ms, respectively. Next, for communication cost we consider the data rate of WIMAX presented in [25]. We set the communication rate to 2 Mbps and the maximum number of children for each SM as 3. With these parameters, the proposed scheme has an aggregation time of 5.32 ms while the schemes presented in [13], [14] and [15] have an aggregation time of 11.58 ms, 16.24 ms and 7.96 ms, respectively. The variation in the aggregation time as a function of the number of SMs for the proposed scheme, [13], [14], and [15] is shown in Fig. 5. From Fig. 5, we can see that the aggregation time of the proposed scheme is lower as compared to [13], [14] and [15]. Overall, we argue that the performance of the proposed hop-by-hop data aggregation scheme is better than that of [13], [14] and [15], and hence it is more suitable for smart grid security.

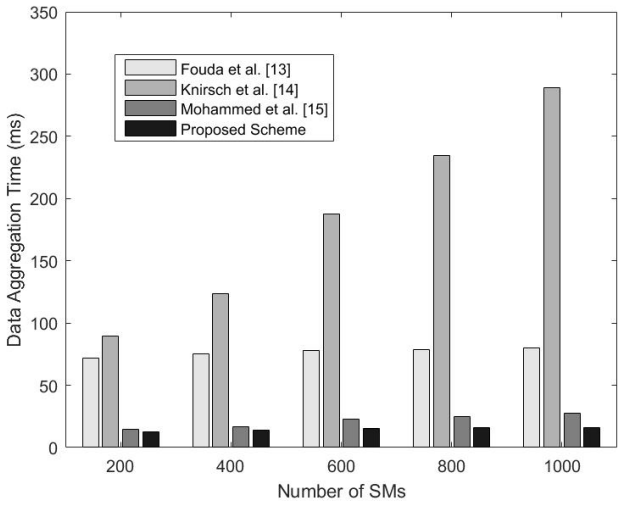

Figure 5. Variation of aggregation time in terms of number of SMs.

VI. 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 [31] for formally analyzing the privacy of our scheme. In this model, attacker $\mathcal{A}$ is allowed to eavesdrop on all the channels between the smart meters and the aggregator and then he/she is also allowed to perform any active or passive attack. $\mathcal{A}$ is allowed to run the following queries:

- **Execute**$(\mathcal{M}, \mathcal{AG}, i)$: This query denotes the passive attacks. In this regard, the attacker can eavesdrop on all the transmitted messages between the smart meter $\mathcal{M}$ and the aggregator $\mathcal{AG}$ in the i-th session. Here, $\mathcal{AG}$ can be another smart meter (acting as an aggregator) or the TPA. Consequently, the attacker can obtain all the exchanged data between the $\mathcal{AG}$ and meter $\mathcal{M}$.
- **Send**(U, V, m, i): This query denotes the modeling of the active attacks in the system, where the 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 is allowed to block the exchanged message m between the smart meter and the aggregator.

- **Corrupt**$(\mathcal{M}, K)$: In this query, the attacker $\mathcal{A}$ is allowed to access secret information K stored in the smart meter's memory.
- **Test**$(\mathcal{M}_0, \mathcal{M}_1, i)$: This query permits us 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 completion of the session): An aggregator instance $\mathcal{AG}_j$ and a meter instance $\mathcal{M}_i$ are partners if, and only if, both have output Accept($\mathcal{M}_i$) and Accept($\mathcal{AG}_j$), 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 $\mathcal{AG}$ 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. 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 $(\mathcal{AG}, \mathcal{M}_0, j)$ and obtains the parameters $\{X_{0,j}^{\mathcal{M}_0}, H_{0,j}\}$.
- ***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 $(\mathcal{AG}, \mathcal{M}_b, j+1)$ and obtains $\{X_{0,j+1}^{\mathcal{M}_b}, H_{0,j+1}\}$.
- ***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 parameters $(X_{0,j}^{\mathcal{M}_0}, X_{0,j+1}^{\mathcal{M}_b})$ and $(H_{0,j}^{\mathcal{M}_0}, H_{0,j+1}^{\mathcal{M}_b})$ are calculated as follows: $X_{0,j}^{\mathcal{M}_0} = \mathcal{M}_{0,j} + r_{\mathcal{M}_0,j} \mod p$, $X_{0,j+1}^{\mathcal{M}_b} = \mathcal{M}_{b,j+1} + r_{\mathcal{M}_b,j+1} \mod p$, $H_{0,j}^{\mathcal{M}_0} = h(X_{0,j}^{\mathcal{M}_0} || kh_{\mathcal{M}_0} || t_{\mathcal{M}_0,j})$, 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 $r_{\mathcal{M}_0,j} \neq r_{\mathcal{M}_b,j+1}$, 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 $\mathcal{M}_\mathbf{0}$ or $\mathcal{M}_1$ can be denoted by $Adv_{\mathcal{A}}^{INDPriv}(k) = |\Pr[b^{'} = b] - 1/2| \leq \epsilon$.

VII. Conclusion

In this paper we developed a novel *hop-by-hop* data aggregation scheme for smart grids. The usage reports are aggregated according to the aggregation tree. Security analysis shows that the proposed scheme satisfies all the desired requirements. Computation and communication analyses show that the proposed scheme has better performance than existing hop-by-hop data aggregation schemes. Therefore, it can be argued that the proposed scheme is efficient and more suitable for smart grid security.

Acknowledgment

This research 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] Y. Li, X. Cheng and Y.Cao, "Smart Choice for the Smart Grid: Narrowband Internet of Things (NB-IoT)," *IEEE Internet of Things Journal*, DOI: 10.1109/JIOT.2017.2781251, 2017.

[2] Z. Guan et al., "Achieving Efficient and Secure Data Acquisition for Cloud-Supported Internet of Things in Smart Grid," *IEEE Internet of Things Journal*, vol. 4(6), pp. 1934-1944, December 2017.

[3] U.S. NIST, "Guidelines for smart grid cyber security," NIST IR-7628, Aug. 2010, available at: http://csrc.nist.gov/publications/PubsNISTIRs.html#NIST-IR-7628.

[4] 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.* vol. 23(9), pp. 1621–1631, 2012.

[5] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," *in Proc. EUROCRYPT,* pp. 223–228, Prague, Czech Republic, 1999.

[6] X. Liang, X. Li, R. Lu, X. Lin and X. Shen, "UDP: usage based dynamic pricing with privacy preservation for smart grid," *IEEE Transactions on Smart Grid*, vol. 4(1), pp. 141-150, March 2013.

[7] M. Naehrig, K. Lauter and V. Vaikuntanathan, "Can homomorphic encryption be practical?," *In Proc. the 3rd ACM Cloud Computing Security Workshop*, pp. 113-124, 2011.

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

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

[10] 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*. 2013, 1–11.

[11] 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)

[12] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart grid,"*in Proc. Privacy Enhanced Technology Symposium*, pp. 175–191, 2011.

[13] F. Li, B. Luo, P. Liu, "Secure information aggregation for smart grids using homomorphic encryption,"*in First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Gaithersburg, USA, 4–6 October, pp. 327–332. IEEE (2010).

[14] F. Knirsch et al. "Error-resilient Masking Approaches for Privacy Preserving Data Aggregation," *IEEE Trans. Smart Grid,* DOI 10.1109/TSG.2016.2630803, 2016.

[15] H. Mohammed et al. ""Efficient Privacy-Preserving Data Collection Scheme for Smart Grid AMI Networks"," *in Proc. IEEE GLOBECOM*, Washington DC, USA, 2016.

[16] X. Dong, J. Zhou, K. Alharbi, X. Lin and Z. Cao, "An ElGamal-based efficient and privacy-preserving data aggregation scheme for smart grid," *2014 IEEE Global Communications Conference*, Austin, TX, 2014, pp. 4720-4725. doi: 10.1109/GLOCOM.2014.7037553.

[17] C. Castelluccia, A.C.-F. Chan, E. Mykletun and G. Tsudik, "Efficient and provably secure aggregatioin of encrypted data in wireless sensor networks, " *ACM Trans. Sen. Netw.* vol. 5(20): 1-36, 2009.

[18] Nico Saputro, Kemal Akkaya, "Performance evaluation of Smart Grid data aggregation via homomorphic encryption," *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2945-2950, 2012, ISSN 1525-3511.

[19] J. Won, Chris Y. T. Ma, David K. Y. Yau, Nageswara S. V. Rao, "Privacy-Assured Aggregation Protocol for Smart Metering: A Proactive Fault-Tolerant Approach," *IEEE/ACM Transactions on Networking*, vol. 24, pp. 1661-1674, 2016, ISSN 1063-6692.

[20] C-M Chen, Y-H Lin, Ya-Ching Lin, H-Min Sun, "RCDA: Recoverable Concealed Data Aggregation for Data Integrity in Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, pp. 727-734, 2012, ISSN 1045-9219.

[21] S. Stamm, N.P. Sheppard, R. Safavi-Naini, "Implementing trusted terminals with a, SITDRM," *Electr. Not. Theoret. Comput. Sci*. 197(1), 73–85 (2008).

[22] Diffie, W., Hellman, M.E."New directions in cryptography," *IEEE Trans. Inf. Theory* 22(6), 644–654 (1976). IEEE.

[23] Oracle Technology Network. Java Cryptography Architecture (JCA). [Online]. Available: http://docs.oracle.com/javase/6/docs/technotes/guides/crypto/CrypoSpec.html, accessed Apr. 20, 2017.

[24] libpaillier-0.8. Tech. rep. http://hms.isi.jhu.edu/acsc/libpaillier/ (accessed on 16 April 2017).

[25] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, S., C. Buccella, C. Cecati, G.P. Hancke, "Smart grid technologies: communication technologies and standards," *IEEE Trans. Industr. Inf.* vol. 7(4), 529–539 (2011)

[26] Atmel's family of smart power meters. http://www.atmel.com/products/smart-energy/power-metering/ (accessed on 28 May 2017).

[27] R. L. Rivest, A. Shamir, L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems,"*Comm. ACM,* vol. 21(2) pp. 120-126, 1978.

[28] G. J. Popek, C. S. Kline "ncryption Protocols Public-Key Algorithms and Digital Signatures in Computer Networks,"*in Foundations of Secure Computation,* pp. 133–153, 1999.

[29] Y. Wu, S. Fahmy and N. Shroff, "On the construction of a maximum-lifetime data gathering tree in sensor networks: NP-completeness and approximation algorithm," *Proc. IEEE INFOCOM,* pp. 356-360, Phoenix, AZ, April 2008.

[30] H. Tan, I. Korpeoglu and I. Stojmenovi, "Computing localized power-efficient data aggregation trees for sensor networks," *IEEE Transactions on Parallel and Distributed Systems,* vol. 22, no. 3, pp. 489-500, March 2011.

[31] K. Ouafi and R. C.-W. Phan, Privacy of recent RFID authentication protocols, in: Information Security Practice and Experience, Springer, pp. 263-277, 2008.

[32] K. Alharbi, X. Lin, "LPDA: A lightweight privacy-preserving data aggregation scheme for smart grid," *WCSP 2012*, 2012.

[33] "Federal Energy Regulatory Commission Assessment of Demand Response & Advanced Metering," *FERC.gov. Retrieved 16 January 2018.*.

[34] F. Li, B. Luo, "Preserving data integrity for smart grid data aggregation, " *3rd IEEE International Conference on Smart Grid Communications* , 2012.

[35] L. Yang, F. Li, "Detecting false data injection in smart grid in-network aggregation, " *4th IEEE International Conference on Smart Grid Communications*, 2013.

[36] 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.

[37] P. Gope, and B. Sikdar, "Lightweight and Privacy-Friendly Spatial Data Aggregation for Secure Power Supply and Demand Management in Smart-Grids," *IEEE Transactions on Information Forensics & Security*, DOI:10.1109/TIFS.2018.2881730, 2018.

[38] 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.

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, UK. 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 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.