Journal of Computer and Knowledge Engineering, Vol. 6, No. 2, 2023. (57-66)

Ferdowsi University of Mashhad

Journal of Computer and Knowledge Engineering

https://cke.um.ac.ir

Information and Communication Technology Association of Iran

A Lightweighted Secure Scheme for Data Aggregation in Large-Scale IoT-Based Smart Grids

Research Article

Mohammad J. Abdolmaleki[1], Amanj Khorramian[2], Mohammad Fathi[3]

DOI: 10.22067/cke.2024.82734.1080

Abstract. With the emergence of IoT devices, data aggregation in the area of smart grids can be implemented based on IoT networks. However, the communication and computation resources of IoT devices are limited so it is not possible to apply conventional Internet protocols directly. On the other hand, gathering data from smart meters in the advanced metering infrastructure faces challenges such as privacy-preserving and heavy-loaded authentication and aggregation schemes. In this paper, we propose an improved lightweight, secure, and privacy-preserving scheme for aggregating data of smart meters in large-scale IoT-based smart grids. The proposed scheme adopts light-weight operations of cryptography such as exclusive-OR, hash, and concatenation functions. In comparison with the schemes in the literature, the analysis and simulation results show that the proposed scheme satisfies the same security levels, while at the same time burdens lower computation and communication overheads. This observation makes the proposed scheme more suitable to be employed in large-scale and IoT-based smart grids for data aggregation**.**

Keywords: Index Terms— internet of things, smart grids, large-scale networks, light-weight security, data aggregation.

1. Introduction

The Internet of Things (IoT) is a set of smart devices communicating with each other through wired or wireless channels in order to achieve a specific goal [1]. IoT consists of four general environments as Internet of Vehicles (IoV), Internet of Sensors (IoS) or wireless sensor networks (WSN), machine-to-machine (M2M) communications, and Internet of energy (IoE) which is called smart grid (SG) [2].

The growth in population faces the world with challenges of supplying the energy needed by industrial units, offices, houses, etc. Moreover, environmental issues arise due to excessive consumption of fossil fuels. Therefore, it sounds necessary to supply some part of the energy demand from other resources. In order to mitigate these challenges, a smart network is needed to coordinate both suppliers and consumers. The aim of SG is to achieve a trade-off between the supply and demand of electrical energy and also to optimize the consumed energy through power and demand-side management programs.

Using digital information and communication infrastructures, SG establishes a platform to integrate consumers and energy sources such as renewable energies and power plants. Using the aforementioned infrastructure, SG can also provide two-way communication between the power supplier and consumers. In other words, unlike traditional distribution systems, the power supplier can also send data to the consumer. This feature can sometimes cause the consumer to act as a supplier and generate electricity for other consumers. This happens when the supplied energy is more than the energy needed for a consumer [3].

The structure of SG consists of several operators and devices including maintenance personnel, security officers, advanced metering infrastructure (AMI), data aggregator (DA), and intelligent electronic devices. A typical structure of SG data aggregation is shown in Figure 1. Here, AMI's task is to measure and report the amount of consumed energy by each customer smartly using a smart meter (SM). The reported data of SMs is then collected by a trusted third-party DA and then forwarded to the power supplier (PS). Finally, PS generates the power according to the reports of the SMs. This cycle occurs periodically in order to monitor, control, and predict the amount of power consumption, which respectively eventuates cost reduction for both the PS and customers [4].

Data aggregation in SGs faces a number of challenges. One of the most important challenges is posed by the limited processing and communication resources of SMs. In order to address this challenge, protocols of the SG infrastructure have to be designed computationally and communicationally lightweighted while at the same time providing the security of the connections [5]. Another challenge is to preserve the privacy of costomers, for which the measured data is transmitted throughout the network. In this paper, a lightweighted secure scheme for data aggregation in SG is proposed.

A. Related Work

So far, different schemes have been proposed for data aggregation in various networks. In 2018, Zhang et al. [6] proposed a light-weight privacy-preserving data aggregation for resource-constraint edge terminals and edge computing systems that uses online/offline digital

* Manuscript received: 2023 June1, Revised, 2023October7, Accepted, 2023 November7.

[1] MSc Student, Department of Electrical Engineering, University of Kurdistan, Sanandaj, Iran.

[2] Assistant Professor, Department of Computer Engineering and Information Technology, University of Kurdistan, Sanandaj, Iran.

[3] Corresponding author. Associate Professor, Department of Electrical Engineering, University of Kurdistan, Sanandaj, Iran

Email: mfathi@uok.ac.ir.

signature, Paillier homomorphic cryptosystem, and double trapdoor Chameleon hash function. Their scheme provides data confidentiality and keeps the privacy of the network users from the control center and also edge server. Shen et al. [7] proposed a privacy-preserving cube-data aggregation scheme for electricity consumption in smart grids. The proposed scheme is based on Horner's Rule and Paillier cryptosystem. In 2018, Zheng He et al. [8] presented a privacy-preserving multi-functional data aggregation without a trusted third party in smart grids. They have used Paillier homomorphic cryptography to design their scheme. Likewise, Lu et al. [9] proposed a privacy-preserving data aggregation protocol that uses the Paillier homomorphic cryptosystem which also causes higher computation overhead. Li et al. [10] proposed a privacy-preserving multi-subset scheme for data aggregation in smart grids. They have used the Paillier homomorphic cryptosystem in order to prevent the access of trusted third-party aggregators to the private information of consumers. In 2019, Chen et al. [11] proposed a scheme for aggregating the data of power consumed by an SM which uses the Paillier homomorphic cryptosystem. This scheme enables the power supplier to achieve the whole data consumption of SMs, while it has no access to the data of individual SMs.

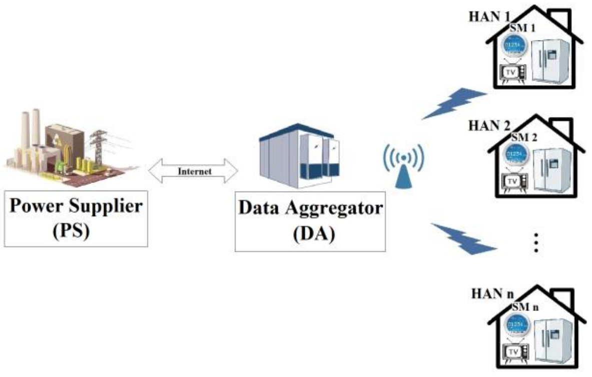

Figure 1. Network model of SG data aggregation

Jo et al. [12] have proposed efficient and privacy-preserving metering protocols for SGs. Their proposed protocols are based on bilinear mapping, hard problem, and the Paillier homomorphic encryption. In 2016, Abdallah et al. [13] presented a lightweight privacy-preserving electricity consumption aggregation scheme that employs a lightweight lattice-based homomorphic cryptosystem. In their proposed scheme, smart domestic facilities aggregate their readings without involving the SM. In [14], a data aggregation scheme is designed based on the discrete logarithm problem, in which a substation has access to the private data of consumers.

In [15], Jo et al. presented two data aggregation schemes based on Paillier homomorphic encryption and elliptic curve digital signature algorithms (ECDSA), respectively. Their schemes cannot warrant the anonymity of the consumers and are also computationally heavy. In 2017, Vahedi et al. [16] designed a privacy-preserving data aggregation scheme for smart grids based on ECDSA. Their scheme cannot assure the privacy of the customers. In 2018, Liu et al. [17] presented a practical privacy-preserving method for aggregating data that exploits EC-ElGamal to encrypt the data of SM. In their approach, although there is no trusted third party to aggregate the data, the control center is assumed to be honest but curious.

He et al. [18] proposed an efficient and privacy-preserving data aggregation scheme for SGs against internal adversaries. They have proposed the scheme using the Boneh-Goh-Nissim public key cryptography method. Unlike most proposed schemes for data aggregation, their proposed scheme does not use bilinear pairing. In 2015, Abdallah et al. [19] presented a lightweight security and privacy-preserving scheme for costumer-side networks which uses *n*th degree truncated polynomial ring units (NTRU) cryptosystem. Their scheme is based on forecasting the electricity demand for a cluster of houses in the same residential areas. Mustafa et al. [20] have proposed a secure and privacy-preserving protocol for smart metering operational data collection. In their proposed protocol, power suppliers and grid operators are allowed to collect the consumption data of SMs securely while SMs' privacy is protected. Their protocol uses Multiparty Computations as the underlying cryptographic primitive. In 2016, Knirsch et al. [21] proposed an approach for privacy-preserving data aggregation based on symmetric cryptographic systems and advanced encryption standard (AES). Their proposed scheme prevents error occurrence in the aggregation process. Gope et al. [22] in 2018 proposed a lightweight and privacy-friendly spatial data aggregation for secure power supply and demand management in smart grids. Their scheme uses only light-weight cryptographic primitives like hash function, XOR, etc.

Recent work has also focused on secure aggregation protocols optimized for federated learning systems. So et al. [25] proposed LightSecAgg, a lightweight and versatile secure aggregation protocol for federated learning that encodes/decodes aggregate mask values instead of dropped user masks. This reduces overhead with increasing dropped users compared to prior art. Experiments showed significantly reduced training time for diverse models and datasets. In other work, Zhang et al. [26] proposed a lightweight multi-dimensional data aggregation scheme for IoT using the Chinese Remainder Theorem and simple additive encryption. It has lower communication and storage than homomorphic encryption schemes while still ensuring security properties like confidentiality, integrity, and robustness. Additional relevant work includes the two aggregation schemes by Qian et al. [27] using lightweight homomorphic encryption resilient to quantum attacks, and the edge-assisted aggregation scheme by Wu et al. [28] using Paillier encryption and virtual name-based verification to reduce overhead.

B. Preliminaries

PRNG: Pseudo-random number generator is a function which its algorithm generates a sequence of numbers with attributes like really random numbers. Its algorithm starts generating numbers with an initial value called the PRNG seed. By having a seed, the PRNG creates a sequence of numbers that is unique to the seed and can be generated one by one or even a whole sequence at a time. Furthermore, numbers generated in a one-by-one procedure are the same as those generated in a sequence with the same specified order.

C. *Contribution*

In this paper, firstly, a procedure to achieve mutual authentication between PS, DA, and SMs is presented. Then a data aggregation scheme is proposed to aggregate the data of power consumption of the SMs periodically. In this regard, lightweight cryptographic operations such as Exclusive-OR (XOR), PRNG, one-way non-collision hash, and concatenation functions are employed. The motivation behind the proposed scheme in this paper is to address the vulnerability in the similar scheme proposed in [22]. The scheme in [22] employs PRNG to generate a sequence of random numbers with a given seed in PS and then transmit these numbers via the network to the SMs. If a sequence of random numbers is sniffed in the network by an attacker, then it would be possible to detect the adopted seed and determine the next numbers. This certainly results in information disclosure. Besides, transmitting the numbers to SMs burdens high computation and communication overhead in the network.

To address the aforementioned vulnerability in [22], in this paper, the same random numbers are generated simultaneously in the PS and individual SMs. In other words, rather than transmitting random numbers sequentially in the network, these numbers are generated with the same seeds at both PS and SMs. Therefore, having this approach causes significant improvements in security regards due to the inaccessibility of the random numbers and the seed by an intruder. Moreover, computation and communication overheads are highly decreased due to the generation of random numbers at both sides of the network.

In the proposed scheme, the personal information of each consumer is private and unachievable by the others. Nevertheless, the SG is able to monitor the summation of the consumed power in the network. The security of the proposed scheme is evaluated in terms of authentication, secure key establishment, data confidentiality, data integrity, and consumer privacy. Finally, the communication and computation overhead of this scheme is evaluated and compared with those in the literature.

In this work, rather than transmitting random numbers from the power supplier to smart meters, we generate matching random numbers simultaneously at both entities using pre-shared seeds. This enhances security by eliminating the transmission of random numbers that could be intercepted. Unlike the prior scheme that transmits pseudorandom numbers over the network, the proposed one generates matching random numbers at the supplier and meter sides using shared seeds. The scheme requires fewer hash function evaluations, encryptions/decryptions, and concatenations compared to the existing protocols in the literature. Our scheme is suitable for resource-constrained IoT devices and can support large-scale smart grid networks.

2. System Model

In this section, the network and the opponent models are explained. The notations and cryptographic functions which are used throughout this paper are shown in Table I.

A. *Network Model*

The network model considered in this paper for data aggregation consists of PS, DA, home area networks (HAN), and SMs, as shown in Figure 1. The PS distributes the electricity to all HANs. The DA’s responsibility is to aggregate the value of power consumption of each SM periodically and then forward it to the PS. The PS takes advantage of reported data to give suitable feedback to the power-generating stations and suppliers. This feedback helps suppliers to make a balance between demand and supply, i.e., how much energy should be generated and distributed over the network. Any fault in reporting the data to the PS may cause wasted energy or lack of supply. Therefore, DA as a trusted third party has a vital role in sustaining the balance between the supply and demand of energy. Each HAN includes a number of electric devices handled by an SM. The SM measures the amount of electricity consumed by the devices and then reports it to the DA using a local network (e.g. WLAN or a cellular network). The PS and DA are connected together using the public Internet (ADSL, cellular network, and so on).

Table 1 Notations and Cryptographic Operations in LWS-DA

Symbols	Definition
SM	Smart Meter
DA	Data Aggregator
PS	Power Supplier
HAN	Home Area Network
s_i	Seed of smart meter SM$_i$
ID_{SM_i}	Real identity of smart meter SM$_i$
ID_A	Identity of data aggregator
PID_i	Pseudo identity of SM$_i$
TID_i	Temporary Identity of SM$_i$
k_i	Secret key between PS and SM$_i$
K_{as}	Secret pre-shared key between DA and PS
kh_i	Shared key between SM$_i$ and DA
$E_k[x]$	Plaintext x encrypted using key k
h	Hashing function
$\oplus$	Bitwise Exclusive-OR (XOR)
$\|\|$	Concatenation

B. *Opponent Model*

It is assumed that the PS is an honest partner as it is owned by the government. Moreover, the DA may be owned by a private company to cooperate with the PS. Hence, the DA is assumed as a truthful but curious entity that may be tempted to acquire the data of power consumption by each HAN in the motivation of selling the usage information to other companies. Also, it is assumed that any element on the network connections between DA and PS may behave as an adversary and attempt to perceive the private information of each HAN. Furthermore, there may be an SM that is interested in accessing the data consumption of other SMs from other HANs. Attacks from outside of network are also likely and assumed. For instance, an illegal user may impersonate itself as legal entity such as SM or even DA.

3. Proposed Scheme

To mitigate the challenges of privacy-preserving and high computation and communication overheads of data aggregation schemes in the literature, here, in this section, we propose a light-weighted secure data aggregation (LWS-DA) scheme for SGs. The proposed scheme is illustrated in two phases: the *authentication phase* and the *data aggregation phase*. A set of n SMs, indexed by i, are considered which are supplied by the PS. In the authentication phase, the PS ascertains the identity of each SM$_i$ and DA. Then it proves its legality to both of them, using an encryption key kh_i and a set of temporary identities that can be updated and generated between the SM and the DA.

In the data aggregation phase, the DA collects the aggregate consumption of HANs, without knowing the exact power consumption of each HAN, as explained in subsection B.

In the proposed LWS-DA scheme, a time period consisting of one authentication interval followed by m data aggregation timeintervals are assumed. In other words, for n SMs in a period, one authentication process is done and then the data aggregation process is done for m time intervals. After the mth round of data aggregation, a new period begins. Therefore, the authentication process must be done in order to continue the data aggregation of the next rounds.

A. Authentication Phase

In this phase, we adopt the authentication method used in [22] with a number of modifications, shown in Figure 2. The modifications, shown in red-dashed rectangles, are to define and forward parameters required in the aggregation phase. The following are steps done in the Authentication phase.

Step AUTH1: Each SM$_i$, with the pseudo-identity PID_i and pre-shared secret key k_i with the PS, generates a random number N_s as its nonce. It then computes the hash-output $V_0 = h(PID_i \parallel N_s \parallel k_i)$ by concatenating its pseudo-identity, nonce, and its pre-shared secret key, where $h(.)$ is a hash function. After that, it creates the message M$_{AU1}$: $\{PID_i, N_s, V_0\}$ and sends it to the DA.

Step AUTH2: On receiving the messages M$_{AU1}$ from each SM, the DA generates its own nonce N_a. Similar to SMs, DA concatenates its real identity ID_A, the nonce N_a, and its pre-shared secret key K_{as} with PS to compute the hash-output $V_1 = h(ID_A \parallel N_a \parallel K_{as})$. Then, the DA sends message M$_{AU2}$: {M$_{AU1}$ ∥ (ID_A, N_a, V_1)} to the PS.

Step AUTH3: When PS receives the message M$_{AU2}$, firstly, it calculates two has-outputs V_0 and V_1, and then compares the values of PID_i, V_0 and V_1 of itself with those that are achieved from message M$_{AU2}$. If all three values in the comparison are matched, then the PS generates PID_i^{new}, which is a new pseudo-identity for each SM$_i$. Thereafter, PS calculates the set of seeds $s_i = PID_i^{new} \oplus k_i$ for SMs in order to generate random numbers in the aggregation phase to the encrypt energy consumption of individual SMs. Then, PS computes parameters $T = h(ID_{SM_i} \parallel k_i \parallel N_s)$, $x = h(k_i \parallel T \parallel N_s) \oplus h(K_{as} \parallel N_a)$, $y = h(T \parallel N_s \parallel k_i) \oplus N_a$, $z = h(T \parallel ID_{SM_i} \parallel k_i) \oplus PID_i^{new}$, $V_2 = h(K_{as} \parallel N_a \parallel x)$ and $V_3 = h(T \parallel y \parallel z \parallel k_i)$, and stores all generated seeds as $S = \{s_1, s_2, \dots, s_n\}$. Details of these parameters are given in Figure 2. Finally, the PS sends message M$_{AU3}$: $\{x, y, z, V_2, V_3\}$ to the DA.

Step AUTH4: Upon receiving message M$_{AU3}$, the DA calculates the hash-output V_2 and compares it with V_2 received from the PS. If the comparison is verified, DA computes $TK = x \oplus h(K_{as} \parallel N_a)$ and generates the secret key $kh_i = h(TK \parallel N_a \parallel N_s)$ between DA and each SM$_i$. Then, it generates a set of temporary identities $TID_{i,m} = \{tid_{i,1}, tid_{i,2}, \dots, tid_{i,m}\}$ to be used in the aggregation time intervals. These identities are encrypted with keys kh_i as $TID_{i,m}^* = E_{kh_i}[TID_{i,m}]$. The DA finally computes $V_4 = h(TID_{i,m}^* \parallel kh_i \parallel ID_A)$, and after storing TID_{im} and kh_i, sends message M$_{AU4}$: $\{(y, z, V_3) \parallel (TID_{i,m}^*, V_4)\}$ to each SM$_i$.

Step AUTH5: On receiving M$_{AU4}$, each SM$_i$ computes $T = h\left(ID_{SM_i} \parallel k_i \parallel N_s\right)$ and then calculates V$_3$ to verify if V$_3$ received from DA is verified. If the verification is succeeded, it computes $N_a = h(T \parallel N_s \parallel k_i) \oplus y$, $TK = h(k_i \parallel T \parallel N_s)$, $kh_i = h(TK \parallel N_s \parallel N_a)$, $PID_i^{new} = h(T \parallel ID_{SM_i} \parallel k_i)$ and the seed $s_i = PID_i^{new} \oplus k_i$. Then, it calculates V_4 and compares it with V_4 received from DA. If the comparison passed, the SM$_i$ decrypts $TID_{i,m}^*$ in order to achieve $TID_{i,m}$, and stores $\{s_i, PID_i^{new}, TID_{i,m}, kh_i\}$ for data aggregation. Note that the seed s_i used in each SM$_i$ is the same as in the PS.

B. Data Aggregation Phase

This phase is done in two steps at the beginning of each time interval T_j. In step 1, as illustrated in Figure 3, the PS generates a set of random numbers $R_j = \{r_{i,j}\}_{i=1}^{n}$ at each time interval T_j using the adopted seeds in the authentication phase. Then, the PS computes the $\mathfrak{R}^*$ as shown in Figure 3, and sends it to the DA. The DA then generates a time-stamp t_a and also encrypts the time-interval and sends them together to all SMs. Hereafter, each SM decrypts the received message to verify the time interval and then generates its own random number $r_{i,j}$ using the seed s_i generated in the authentication phase. This completes the first step of this phase. The motivation behind generating random numbers R at both PS and SMs is to mitigate the vulnerability in the similar scheme proposed in [22], in which random numbers are only generated at PS and then sent to individual SMs in the network. If a sequence of these numbers is sniffed by an attacker, it is possible to detect the seeds and upcoming random numbers. This certainly results in information disclosure of the consumed energy of SMs. Moreover, as random numbers are not transmitted in the network, there is a gain of communication overhead in the proposed LWS-DA scheme.

After that, in step 2, as shown in Figure 4, each SM first generates a time-stamp t_i, then chooses a temporary identity $tid_{i,j}$ from the temporary identity set TID_i which was received from DA in the authentication phase. Then, using the measured value M_i as the amount of energy consumed by SM$_i$ and the pre-generated random number $r_{i,j}$, it computes a blinded value X_i by adding $r_{i,j}$ to M_i. Finally, SM$_i$ sends its temporary identity $tid_{i,j}$, the blinded value X_i, the hash-output H_i, and its time-stamp t_i to the DA. Upon receiving the data from SMs, the DA sums up all the X_is to send the total amount of consumed energy to the PS.

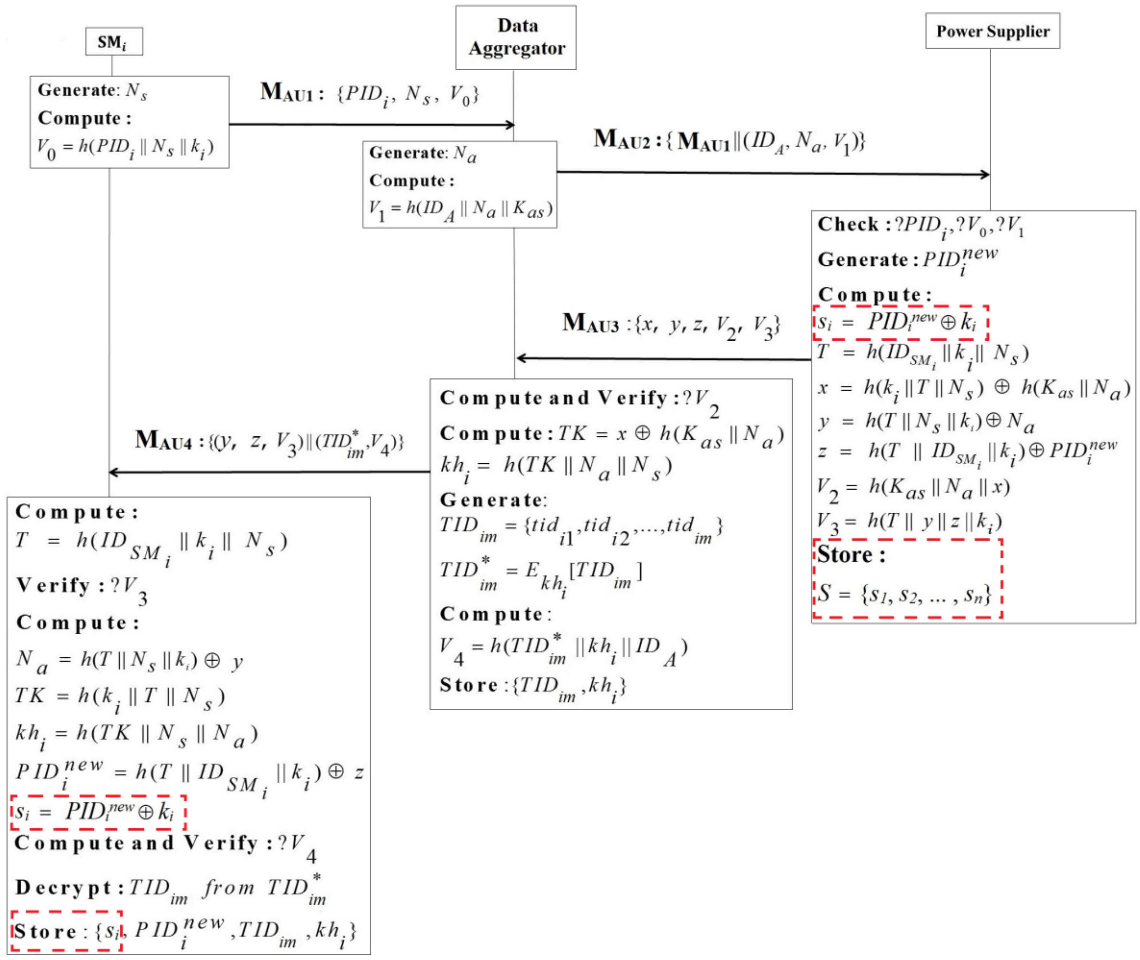
SM_i
Data Aggregator
Power Supplier
Generate: N_s
Compute : $V_0 = h(PID_i \| N_s \| k_i)$
M_{AU1} : $\{PID_i, N_s, V_0\}$
Generate: N_a
Compute : $V_1 = h(ID_A \| N_a \| K_{as})$
M_{AU2} : $\{M_{AU1} \| (ID_A, N_a, V_1)\}$
Check : $?PID_i, ?V_0, ?V_1$
Generate : PID_i^{new}
Compute :
$s_i = PID_i^{new} \oplus k_i$
$T = h(ID_{SM_i} \| k_i \| N_s)$
$x = h(k_i \| T \| N_s) \oplus h(K_{as} \| N_a)$
$y = h(T \| N_s \| k_i) \oplus N_a$
$z = h(T \| ID_{SM_i} \| k_i) \oplus PID_i^{new}$
$V_2 = h(K_{as} \| N_a \| x)$
$V_3 = h(T \| y \| z \| k_i)$
Store :
$S = \{s_1, s_2, \ldots, s_n\}$
M_{AU3} : $\{x, y, z, V_2, V_3\}$
Compute and Verify : $?V_2$
Compute : $TK = x \oplus h(K_{as} \| N_a)$
$kh_i = h(TK \| N_a \| N_s)$
Generate:
$TID_{im} = \{tid_{i1}, tid_{i2}, \ldots, tid_{im}\}$
$TID_{im}^* = E_{kh_i}[TID_{im}]$
Compute:
$V_4 = h(TID_{im}^* \| kh_i \| ID_A)$
Store : $\{TID_{im}, kh_i\}$
M_{AU4} : $\{(y, z, V_3) \| (TID_{im}^*, V_4)\}$
Compute :
$T = h(ID_{SM_i} \| k_i \| N_s)$
Verify : $?V_3$
Compute :
$N_a = h(T \| N_s \| k_i) \oplus y$
$TK = h(k_i \| T \| N_s)$
$kh_i = h(TK \| N_s \| N_a)$
$PID_i^{new} = h(T \| ID_{SM_i} \| k_i) \oplus z$
$s_i = PID_i^{new} \oplus k_i$
Compute and Verify : $?V_4$
Decrypt : TID_{im} from TID_{im}^*
Store : $\{s_i, PID_i^{new}, TID_{im}, kh_i\}$

Figure 2. Modified Authentication Phase (modifications are shown in red rectangles)

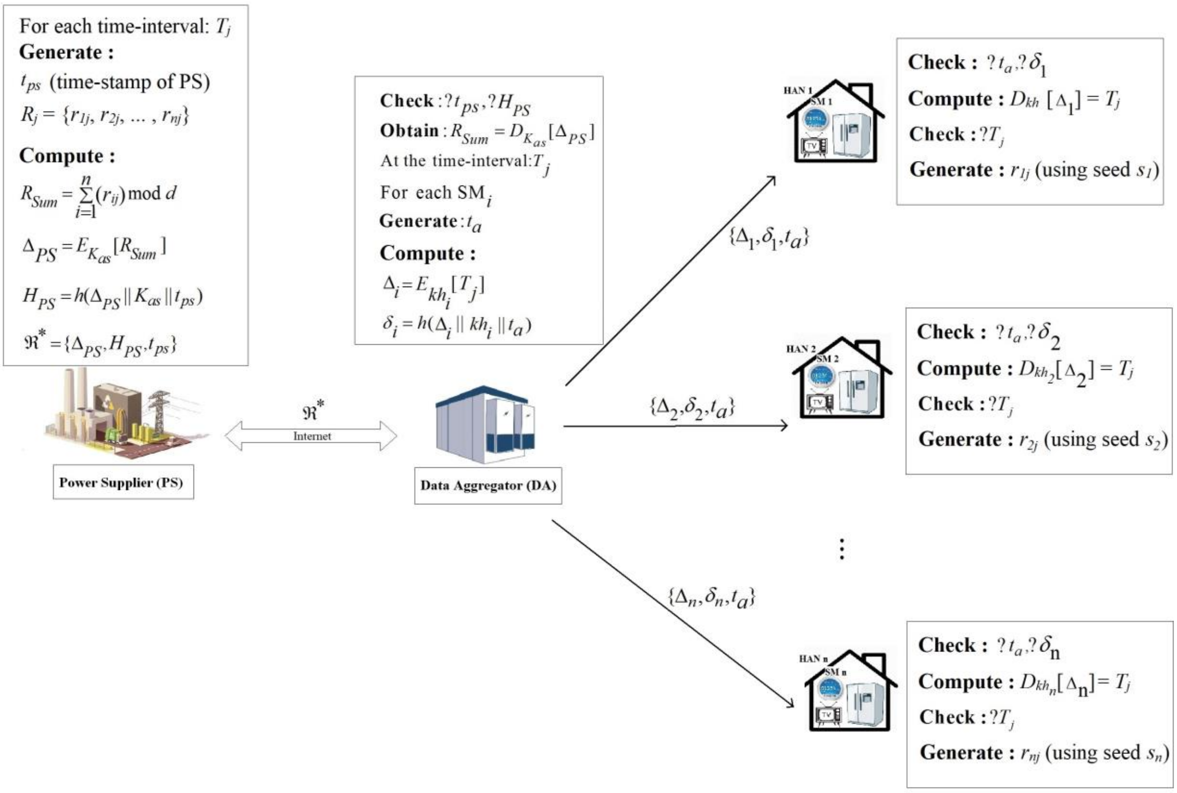
For each time-interval: T_j
Generate :
t_{ps} (time-stamp of PS)
$R_j = \{r_{1j}, r_{2j}, \ldots, r_{nj}\}$
Compute :
$R_{Sum} = \sum_{i=1}^{n} (r_{ij}) \bmod d$
$\Delta_{PS} = E_{K_{as}}[R_{Sum}]$
$H_{PS} = h(\Delta_{PS} \| K_{as} \| t_{ps})$
$\mathfrak{R}^* = \{\Delta_{PS}, H_{PS}, t_{ps}\}$
Power Supplier (PS)
$\mathfrak{R}^*$
Internet
Data Aggregator (DA)
Check : $?t_{ps}, ?H_{PS}$
Obtain : $R_{Sum} = D_{K_{as}}[\Delta_{PS}]$
At the time-interval: T_j
For each SM_i
Generate : t_a
Compute :
$\Delta_i = E_{kh_i}[T_j]$
$\delta_i = h(\Delta_i \| kh_i \| t_a)$
$\{\Delta_1, \delta_1, t_a\}$
$\{\Delta_2, \delta_2, t_a\}$
$\{\Delta_n, \delta_n, t_a\}$
HAN 1
SM 1
HAN 2
SM 2
HAN n
SM n
TV
Check : $?t_a, ?\delta_1$
Compute : $D_{kh}[\Delta_1] = T_j$
Check : $?T_j$
Generate : r_{1j} (using seed s_1)
Check : $?t_a, ?\delta_2$
Compute : $D_{kh_2}[\Delta_2] = T_j$
Check : $?T_j$
Generate : r_{2j} (using seed s_2)
Check : $?t_a, ?\delta_n$
Compute : $D_{kh_n}[\Delta_n] = T_j$
Check : $?T_j$
Generate : r_{nj} (using seed s_n)

Figure 3. Step 1 in LWS-DA data aggregation phase

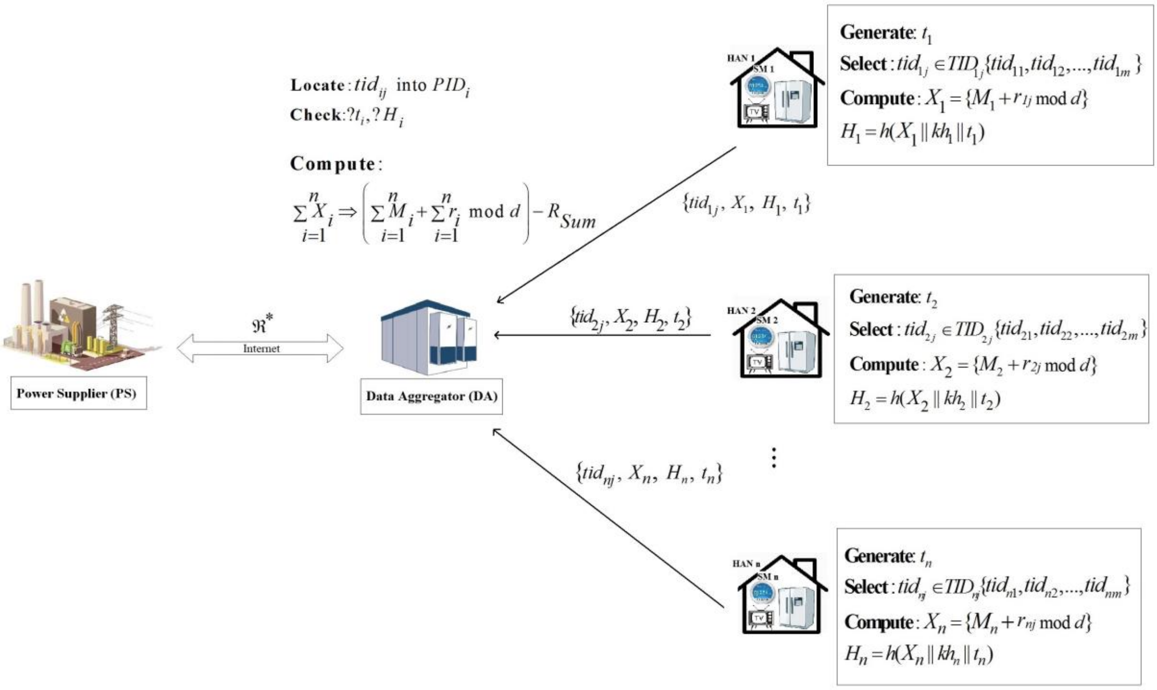
Locate: tid_{ij} into PID_i
Check: $?t_i, ?H_i$
Compute:
$\sum_{i=1}^{n} X_i \Rightarrow \left(\sum_{i=1}^{n} M_i + \sum_{i=1}^{n} r_i \bmod d\right) - R_{Sum}$
$\{tid_{1j}, X_1, H_1, t_1\}$
$\{tid_{2j}, X_2, H_2, t_2\}$
$\{tid_{nj}, X_n, H_n, t_n\}$
$\Re^*$
Internet
Power Supplier (PS)
Data Aggregator (DA)
HAN 1
SM 1
TV
HAN 2
SM 2
HAN n
SM n
Generate: t_1
Select: $tid_{1j} \in TID_{1j}\{tid_{11}, tid_{12}, ..., tid_{1m}\}$
Compute: $X_1 = \{M_1 + r_{1j} \bmod d\}$
$H_1 = h(X_1 \| kh_1 \| t_1)$
Generate: t_2
Select: $tid_{2j} \in TID_{2j}\{tid_{21}, tid_{22}, ..., tid_{2m}\}$
Compute: $X_2 = \{M_2 + r_{2j} \bmod d\}$
$H_2 = h(X_2 \| kh_2 \| t_2)$
Generate: t_n
Select: $tid_{nj} \in TID_{nj}\{tid_{n1}, tid_{n2}, ..., tid_{nm}\}$
Compute: $X_n = \{M_n + r_{nj} \bmod d\}$
$H_n = h(X_n \| kh_n \| t_n)$

Figure 4. Step 2 in LWS-DA data aggregation phase

4. Security Maintained

The security analysis of the former scheme has been discussed in detail [22]. The modifications made by the proposed LWS-DA scheme do not degrade the security characteristics, as investigated in the following.

A. *Authentication*

As seen in Figure 2, the PS authenticates each SM$_i$ by checking its *PID$_i$* as well as the hash-output *V$_0$*, where only a legal SM has access to them. Also, the PS authenticates the DA by verifying the identity of the DA and the hash-output *V$_1$* which only the legal DA is able to generate. The SM$_i$ authenticates the PS by checking the value of hash-output *V*$_2$ and also the DA authenticates the PS by checking *V*$_3$. Therefore, each entity in the considered network authenticates the other two entities.

B. *Secure Key-Establishment*

As presented in Figure 2, there is no key exchange between the entities of the network. In particular, the keys k_i, K_{as} and kh_i are not transmitted between entities on the communication channels. The key kh_i is generated in DA and SM$_i$ individually. Therefore, the keys remain secret.

C. *Data Confidentiality*

The data on consumers' energy usage must be secret from others except for the SM$_i$ itself and the PS. Therefore, the value metered by each SM$_i$ is blinded with a random number $r_{i,j}$ which is from a long-enough range, i.e., $X_i = M_i + r_{i,j}$ mod d. Therefore, the DA has access only to the blinded measurement X_i and cannot perceive the exact amount of consumed power by each SM. So transmitted data remains confidential.

D. *Data Integrity*

The DA has the option to check the integrity of the information received from the SM of every HAN. Also, DA needs to confirm the respectability of the pertinent data received from the PS during the information gathering. Using a one-way and non-collision hash function to transmit message hash, data integrity is also preserved.

E. ***Consumer Privacy***

As discussed in the proposed scheme, only the PS can map the pseudo-identity *PID$_i$* of SM$_i$ to its real identity ID_{SM_i}. Therefore, the DA or an illegal entity cannot access the identity of consumers.

5. Security Analysis

Suppose an adversary, denoted as **A**, possesses the ability to intercept all forms of communication. This includes the capacity to replay, alter, eliminate, and rearrange messages. Furthermore, assume **A** has the ability to query the random oracle **h**, which produces strings of 256 bits. The security of the LWS-DA scheme is analyzed in the following.

A. ***Session Key Secrecy***

To compute kh_i, **A** must compute $TK = x \oplus h(K_{as} \parallel N_a)$. **A** can obtain N_a, but not K_{as} since we assumed PS is trusted. **A** can query **h** to try to find an input that outputs x. But this will fail except with negligible probability of 2^{-256} since **h** is a random oracle. Therefore, **A** cannot compute TK or kh_i except with a negligible probability. Thus, **A** cannot compute any session key kh_i.

B. ***Entity Authentication***

SM_i is assured of DA's identity because only DA can produce $E_{kh_i}[TID_{i,m}]$ which can be correctly decrypted by kh_i and is known only by DA and SM_i. DA is assured of SM_i's identity because SM_i demonstrates knowledge of k_i by producing correct $h(T \parallel y \parallel z \parallel k_i)$ which DA can verify using values from PS. Thus, SM_i and DA are assured of each other's identities.

C. ***Data Confidentiality***

A seees only the blinded meter readings $X_i = \{M_i + r_{ij} \bmod d\}$. Without knowing r_{ij}, **A** cannot compute the actual consumption M_i. The value of r_{ij} is generated using seed s_i which we assumed **A** cannot obtain. Therefore, **A** cannot learn the smart meter readings except by guessing r_i which will succeed only with a negligible probability of 2^{-256}. Thus, **A** cannot learn any smart meter's power consumption.

D. ***Data Integrity***

A cannot tamper with the aggregated energy consumption data, because it cannot forge the signatures of SMs on their blinded values. The signatures are based on the secret keys which we assumed **A** cannot obtain. The signatures are also verifiable by the DA using values from PS.

E. ***Privacy of SMs***

A cannot learn the real identities of the SMs, because the SMs use temporary identities in the data aggregation phase. The temporary identities are encrypted with the session keys kh_i which we assumed **A** cannot compute. The temporary identities are also indistinguishable from random values. Thus, **A** cannot learn the real identities of the SMs.

Therefore, the LWS-DA scheme is secure against the attacks mentioned above if the random oracle is secure.

6. Performance Evaluation

Performance evaluation of the proposed LWS-DA scheme along with evaluating computation and communication overheads are done in this section. Firstly, in subsection *A*, the comparison of computation overhead between LWS-DA scheme and the scheme in [22] is analyzed. Then, in subsection *B*, the communication overhead comparison between the two mentioned schemes is done, and finally, the result of simulations which shows the differences between the two mentioned schemes in terms of running time is illustrated in subsection *C*.

In the aforementioned comparisons, key values, identities, and random numbers are assumed to be of length 256-bits. Moreover, time-stamps, time interval, and R_{sum} are of lengths 64, 8 and $\lceil Log_2^n \rceil$+256 bits respectively, where n is the number of SMs. Calculations are done for a time period consisting of $m = 24$ time intervals. In other words, the data aggregation process of n SMs occurs for m times. After the mth round, the authentication process is repeated in order to continue the data aggregation in the next round.

Table 2 Computation Overhead of The Proposed LWS-DA Scheme VS. [22]

Phase / Operation		Aggregation-step1 (one period)					
		[Gope et al.] Scheme			**LWS-DA Scheme**		
		Number of Operations		**Input length**	**Number of Operations**		**Input length**
Hashing	**H₃**	$24(n+3)$	$24(n+1)$	1040	$24(n+3)$	$24(n+1)$	328
			24×2	$584 + l^*$		24×2	$584 + l^*$
	H₄	$24(n+1)$		1352	$-$		
	Tot.	$24(2n+4)$		$-$	$24(n+3)$		$-$
Encryption	**E₁**	$-$			$24n$		8
	E₂	24		$264 + l^*$	24		$264 + l^*$
	E₄	$24n$		776	$-$		
	Tot.	$24(n+1)$		$-$	$24(n+1)$		$-$
Decryption		24×2	24	$264 + l^*$	24×2	24	$264 + l^*$
			24	776		24	8
Concatenation		$24(9n+3)$			$24(2n+4)$		

[*] $l = \lceil Log_2^n \rceil$

This paper compares the proposed LWS-DA scheme to Gope et al.'s prior scheme in [22] in terms of communication overhead and overall efficiency. The comparison is comprehensive and robust, covering all the key performance indicators and metrics. The scheme in [22] was selected as a benchmark because it is similar to LWS-DA in that it uses lightweight cryptographic primitives for secure data aggregation in smart grid networks. While several other related schemes have been recently proposed, a further detailed comparison with each of these schemes is not essential to demonstrate the advantages of LWS-DA. The in-depth analysis of security properties and the experimental results validate that LWS-DA accomplishes the key objectives of an efficient and secure data aggregation scheme suitable for resource-constrained smart grid environments. The insights from the comparison sufficiently highlight the technical contributions of LWS-DA without requiring additional comparative analyses at this stage, since the existing scheme, being a well-established and widely accepted benchmark in the field, provides a solid basis for comparison. Furthermore, the methodology used for this comparison is rigorous, ensuring that all relevant aspects are considered. Therefore, additional comparisons with other schemes, including recent works, are not essential. The results obtained from this comparison are sufficient to draw reliable conclusions and make informed decisions about the performance and efficiency of the proposed work. This eliminates the need for further comparisons, allowing us to focus on enhancing and implementing the proposed scheme.

A. Computation Overhead Comparison

Table II shows the comparison of the computation overhead of [22] and LWS-DA scheme in terms of hashing, encryption, decryption, and concatenation operations. The comparisons are done by calculating the number of input bits for each operation and also the number of usages of each operation. It is worth noting that the index of each operation indicates the number of concatenated input arguments and the parameter n is the number of SMs. For the sake of simplicity, we only focus on the items with different values.

Hashing. As shown in Table II, both schemes use $24(n + 3)$ three-input hashing functions (H_3), in which $24(n + 1)$ functions burden the input length of 1040 bits for the scheme in [22] while this value is 328 bits in our scheme, that is less than a third of the scheme in [22]. Moreover, the scheme in [22] uses $24(n + 1)$ four-input hashing functions (H_4) with input length of 1352 bits, while there is no usage of this operation in our scheme LWS-DA. This yields a significant improvement in the overall computational complexity. Figure 5 shows the number of input-bits of hash functions (H_3 and H_4) in LWS-DA compared to the scheme of [22] versus the number of SMs. While it seems that the length of input bits grows exponentially in [22], this length increases linearly in LWS-DA.

Encryption. In comparison between the schemes, it is apparent that $24n$ number of four-input encryption function (E_4) employed in [22] with an input length of 776 bits is replaced with $24n$ number of one-input encryption function (E_1) in LWS-DA with an input length of 8 bits. This is absolutely a significant improvement in terms of computation. This can also be seen in Figure 6 which shows the number of input bits of encryption functions for both schemes in terms of the number of SMs. Great outperformance of LWS-DA is observed.

Decryption. Considering the concurrency of computations in SMs, we only take into account the computational overhead of the decryption operation for one SM. Hence, as seen in Table II, both schemes use 24×2 decryption operations, in which 24 operations in [22] have an input length of 776 bits while this value is 8 bits in LWS-DA scheme.

Concatenation. The proposed scheme in [22] uses $24(9n + 3)$ number of concatenations, whereas our LWS-DA scheme uses $24(2n + 4)$ concatenations. The improvement becomes significant when the number of SMs becomes large, as observed in Figure 7.

Considering the aforementioned observations, our proposed scheme uses a smaller number of hash functions, encryption functions, decryption functions, and concatenation operations with less input bit-length than that in [22]. This certainly results in a light-weighted data aggregation scheme.

B. Communication Overhead Comparison

Table III shows the difference between the scheme presented in [22] and LWS-DA scheme in terms of communication overhead. As seen in Table III, in the scheme of [22], the number of bits transmitted by a message from PS to DA in step 1 of the data aggregation phase is increasing linearly with respect to the number of SMs. However, in the proposed scheme, the complexity is logarithmic. In other words, as the number of SMs increases, our proposed scheme transmits fewer bits than the scheme proposed in [22], asymptotically. One reason for this improvement is due to generating random numbers simultaneously at both PS and SMs, rather than transmitting them over the network.

Table 3 Communication Overhead of The Proposed LWS-DA Scheme VS. [22]

Scheme **Phase**		**[Gope et al.] Scheme (bits)**	**LWS-DA Scheme (bits)**
Aggregation-step 1	**PS to DA**	$(n \times 1032) + l^* + 584$	$l^* + 584$
	DA to SM$_i$	1352	328

C. Simulation Results

According to the modification that led to the achieved results towards the overhead reduction, we anticipate a significant reduction in the running time of the data aggregation scheme. To address this point, we conduct simulations using Python 3.7 in order to experimentally show the effectiveness of our proposed LWS-DA scheme in comparison with [22].

As noted in [23], at the same security level, the time performance of AES-CBC encryption method is more than other methods such as Paillier, RSA, ECC-based and, so on. On the other hand, since there is no key exchange between the components of the network in both schemes, we use symmetric cryptographies in comparison. Therefore, for the simulation of both schemes, we adopt the AES-CBC encryption with a 256-bits key size which is a lightweight cryptographic method. Moreover, all hashing operations are done with SHA-256 which is a non-collision one-way hash function [24]. Figure 8 shows the difference between LWS-DA and the former schemes in terms of aggregation time for a duration of one period including 24 time intervals. As seen, while the running time of [22] grows exponentially with the number of SMs, our scheme demonstrates a linear increase with a gentle slope.

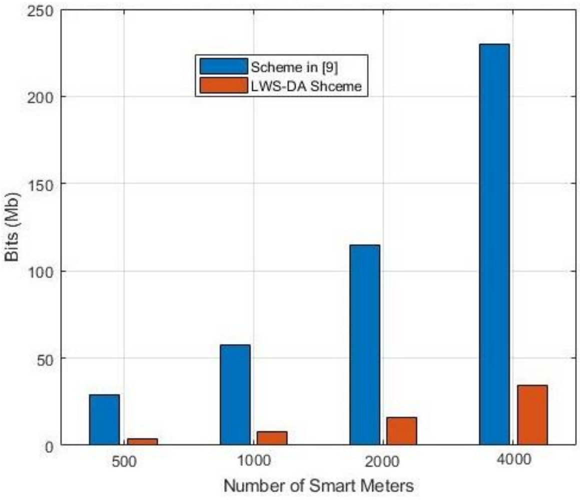

Figure 5. Number of input-bits of hash functions versus number of SMs.

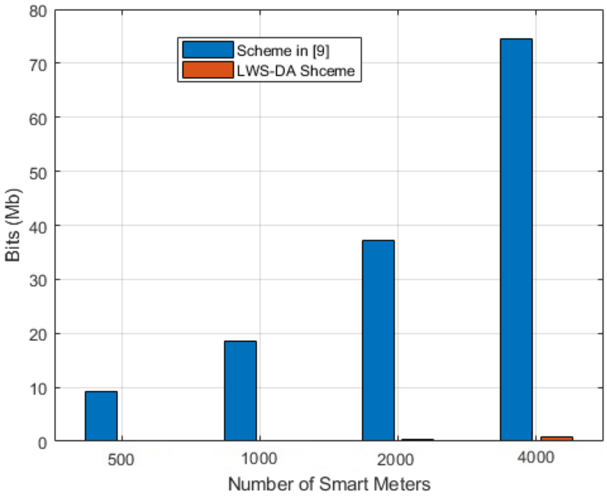

Figure 6. Number of input-bits of encryption functions versus number of SMs.

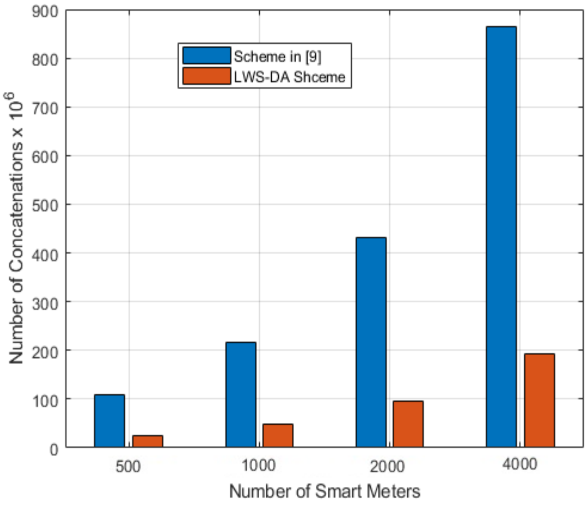

Figure 7. Number of concatenations versus number of SMs

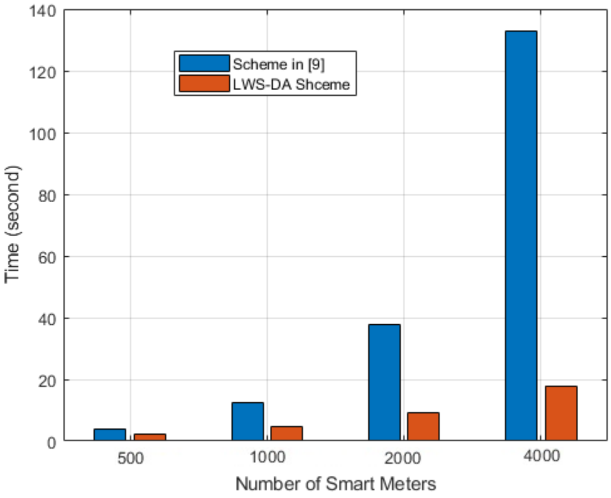

Figure 8. Aggregation time in terms of the number of SMs

7. Conclusion

In this paper, an improved light-weight secure scheme for data aggregation in large-scale IoT-based smart grids is proposed. Firstly, the security of the scheme is discussed and then the effectiveness of computation and communication overheads are demonstrated. The performance analysis shows that the proposed scheme becomes computationally and communicationally more lightweighted than the scheme in the literature, especially when the number of SMs grows in the network. Therefore, the proposed scheme can be used in large-scale smart grids in order to aggregate the data consumption of SMs.

References

[1] IoT Analytics, “Why the internet of things is called internet of things: definition, history, disambiguation,” https://iot-analytics.com/internetof-things-definition, 2014.

[2] M. A. Ferrag, L. A. Maglaras, H. Janicke, J. Jiang, L. Shu, “Authentication protocols for internet of things: A Comprehensive Survey”, Hindawi, Security and Communication Networks, vol. 2017, Article ID 6562953.

[3] G. Gharepatyan, M. Shahidehpour and B. Zaker, Smart Grids and Microgrids, Tehran: Amirkabir University of Technology, 2019.

[4] N. Saxena, B. J. Choi, and R. Lu, “Authentication and authorization scheme for various user-roles and devices in smart grid,” IEEE Trans. on Information Forensics and Security, vol. 11, pp. 907-921, May 2016

[5] E. Kabalci, Y. Kabalci, Smart Grids and Their Communication Systems, Singapure: Springer, 2019

[6] J. Zhang, Y. Zhao, J. Wu, and B. Chen, “A lightweight privacy-preserving data aggregation scheme for edge computing,” in 15th International Conference on Mobile Ad-hoc and Sensor Systems, October 2018.

[7] H. Shen, M. Zhang and J. Shen, "Efficient Privacy-Preserving Cube-Data Aggregation Scheme for Smart Grids," IEEE Trans. on Information Forensics and Security, vol. 12, pp. 1369-1381, 2017.

[8] Z. He, Sh. Pan, D. Lin, “PMDA: privacy-preserving multi-functional data aggregation without TTP in smart grid,” in 17th IEEE International Conference on Trust, Security and Privacy In Computing and Communications, 2018

[9] R. Lu, X. Liang, X. Li and X. Shen, “EPPA: an efficient privacy-preserving aggregation scheme for secure smart grid communications,” IEEE Trans. on Parallel Distribution Systems, vol. 23, pp. 1621-1631, September 2012.

[10] S. Li, K. Xue, Q. Yang and H. Peilin, "PPMA: Privacy-Preserving Multisubset Data Aggregation in Smart Grid," IEEE Trans. on Industrial Informatics , vol. 14, pp. 462-471, 2018.

[11] Y. Chen, J. Ortega, P. Castillejo and L. Lopez, "A Homomorphic-Based Multiple Data Aggregation Scheme for Smart Grid," IEEE Sensors Journal, vol. 19, pp. 3921-3929, 2019.

[12] J. H. Jo, S. I. Kim and H. D. Lee, "Efficient and Privacy-Preserving Metering Protocols for Smart Grid Systems," IEEE Trans. on Smart Grids, vol. 7, pp. 1732-1742, 2016.

[13] A. Abdallah and X. Shen, “A Lightweight Lattice-Based Homomorphic Privacy-Preserving Data Aggregation Scheme for Smart grid,” IEEE Trans. on Smart Grid, vol. 9, pp. 396-405, January 2018.

[14] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu and X. Shen, “A lightweight message authentication scheme for smart grid communication,” IEEE Trans. on Smart Grid, vol. 2, pp. 675-685, December 2011.

[15] H. J. Jo, I. S. Kim and D. H. Lee, “Efficient and privacy-preserving metering protocols for smart grid,” IEEE Trans. on Smart Grid, vol. 7, pp. 1732-174, May 2016.

[16] E. Vahedi, M. Bayat, M. Pakravan and M. Aref, “Secure ECC-based privacy preserving data aggregation scheme for smart grids,” in Computer Networks, vol. 129, no. 1, pp. 28-36, 2017

[17] Y. Liu, W. Guo, C. Fan, L. Chang, and C. Cheng. “A practical privacy-preserving data aggregation (3PDA) scheme for smart grid,” IEEE Trans. on Industrial Informatics, vol. 15, pp. 1767-1774, March 2019.

[18] D. He, N. Kumar, Sh. Zeadally, A. Vinel, L. T. Yang, “Efficient and privacy-preserving data aggregation scheme for smart grid against internal adversaries”, IEEE Trans. on Smart Grids, vol. 8, pp. 2411-2419, 2017

[19] A. Abdallah and X. Shen, “Lightweight security and privacy preserving scheme for smart grid customer-side networks,” IEEE Trans. on Smart Grids, vol. 8, pp. 1064-1074, May 2017

[20] M. A. Mustafa, S. Cleemput, A. Aly and A. Abidin, "A Secure and Privacy-Preserving Protocol for Smart Metering Operational Data Collection," IEEE Trans. on Smart Grids, vol. 10, pp. 6481-6490, 2019.

[21] F. Knirsch, G. Eibl and D. Engel, "Error-Resilient Masking Approaches for Privacy Preserving Data Aggregation," IEEE Trans. on Smart Grid, vol. 9, pp. 3351-3361, 2018.

[22] P. Gope and B. Sikdar, “Lightweight and privacy-friendly spatial data aggregation for secure power supply and demand management in smart grids,” IEEE Trans. on Infornation Forensics and Security, vol. 14, pp. 1554-1566, June 2019.

[23] F. Maqsood, M. Ahmed, M. M. Ali and M. A. Shah, “Cryptography: a comparative analysis for modern techniques”, International Journal of Advanced Computer Science and Applications, vol. 8, Issue 6, 2017

[24] H. Gilbert and H. Handschus, “Security analysis of SHA-256 and sisters”, Internationatl Workshop on Selected Areas in Cryptography, Springer, SAC 2003, pp. 175-193.

[25] J. So, C. He, C. Yang, S. Li, Q. Yu, R. E. Ali, B. Guler and S. Avestimehr, “LightSecAgg: a lightweight and versatile design for secure aggregation in federated learning”, in Proceedings of Machine Learning and Systems, vol. 4, pp. 694-720, 2022.

[26] M. Zhang, Y. Li, Y. Ding, and B. Yang, “A Lightweight and Robust Multi-Dimensional Data Aggregation Scheme for IoT”, IEEE Internet of Things Journal, vol. 1, no. 1, pp. 1-1, 2023

[27] J. Qian, Z. Cao, X. Dong, J. Shen, Z. Liu, and Y. Ye, "Two Secure and Efficient Lightweight Data Aggregation Schemes for Smart Grid," IEEE Trans. on Smart Grid, vol. 12, no. 3, pp. 2625-2637, May 2021

[28] Junhua Wu, Zhuqing Xu, Guangshun Li, Cang Fan, Zhenyu Jin, Yuanwang Zheng, "E-LPDAE: An Edge-Assisted Lightweight Power Data Aggregation and Encryption Scheme", Security and Communication Networks, vol. 2022, Article ID 6218094, 12 pages, 2022