SBPP: Statistical-Based Privacy-Preserving Approach for Data Gathering in Smart Grid

A. Ahadipour[1], M. Mohammadi[2], and A. Keshavarz-Haddad[3]
[1,2] *PhD Candidate of Electrical Engineering*
[3] *Faculty Member of Electrical Engineering*
School of Electrical and Computer Engineering
Shiraz University, Shiraz, Iran
ahadipour.alireza@shirazu.ac.ir, mojtaba.mohammadi@shirazu.ac.ir, keshavarz@shirazu.ac.ir

***Abstract* - As smart grids are getting popular and being employed widely, the privacy of users in such networks is getting more and more substantial. Decision making in smart grids depends on the information gathered from the users periodically. However, having access to the data relevant to the electricity consumption of users is inconsistent with their privacy. On the other hand, it is not sensible to entrust the responsibility of billing to consumers themselves. In this paper, we propose a statistical-based method for data gathering and billing in which the privacy of users is preserved, and at the same time, malicious consumers who try to send erroneous data would be detected.**

KEYWORDS - Data Aggregator, Correlation Coefficient, Privacy, Smart Grid, Supplier, Statistical Method

I. INTRODUCTION

Recently, traditional grids underwent an alteration to smart grids which leads to many benefits including enhanced reliability and resilience, higher intelligence and optimized control, decentralized operation, higher operational efficiency, more efficient demand management, better power quality, and fraud detection [1]. Indeed, consumers minimize their expenses while providers maximize their revenue so that, a win-win partnership can be achieved.

The smart grid is envisaged to be the next generation of traditional grid. In contrast to the traditional grids, there is a bidirectional information flow between suppliers and consumers in smart. To provide this two-way communication, consumers should be equipped with smart meters by which they can measure their usage and send and receive their messages over various communication technologies such as power line communication, cable communication, and wireless communication.

Bidirectional information flows the supplier to generate the electricity based on the demands at any given time period; and at the same time, the supplier can define dynamic billing tariff, and regard to these tariffs that are sent to user periodically (e.g. every 15 minutes). Then, each user can decide whether to decrease its power consumption or not. Thus, electricity is consumed in a more efficient way. On the other hand, in traditional grids, each user sends its electricity usage (by means of a third party) in fixed intervals (e.g. monthly) and its bill is calculated based on their whole usage; no matter their power consumption was in the pick hours or not. However, in smart grid, in the other direction of information flow, the users can declare their need for electricity; indeed, the users send their momentary electricity usage to the suppliers. As a result, unlike traditional grids, in smart grids suppliers provide electricity based on the need of consumers. Hence, ideally, no resource is wasted in the network [2].

In smart grids, one scenario for billing is that users send their electricity usage to local servers – which are responsible for gathering data – periodically by means of smart meters and then, local servers send the gathered data from users to local or central database. Then the server calculates the price of consumed electricity of each user based on the received data of that user. Criticism to this scenario is that the privacy would not be preserved in this method. As all consumers send their usage data to the server and these data are stored in a database, the pattern of each user's power consumption can be obtained by supplier; for instance, inhabitant’s personal schedules, habits, religion, and so on.

Another scenario is that the supplier sends the time-varying tariffs periodically to the consumers and consumers compute their electricity consumption price in the defined period (e.g. one month) based on the received tariffs. Eventually, at the end of each period, every user just sends its total billing amount to the supplier. In this case, the privacy of each consumer would be preserved. It is assumed that based on the existing information archived in databases regard to the power consumption of each user, the database can distinguish whether users are presenting correct billings or not. Consequently, one disadvantage of this scenario is that not only the supplier cannot find the malicious users, but also it would consider the honest ones guilty. For instance, if the power consumption pattern of a user alters over time, this user would be considered as a consumer who is declaring incorrect information; on the other hand, if there is a malicious user who ever sends artificial data, the database cannot notice this fact at all.

According to the afore mentioned scenarios, the main challenge in communications between consumers and suppliers is preserving the privacy of consumers and finding the malicious users simultaneously. To aim this goal, we propose a new statistical-based method for preserving privacy in data gathering of smart grids and at same time detecting the malicious users which manipulate their metering.

The remainder of this paper is organized as follows: In Related Works section, we briefly discusses related works. In System Model section, we introduce our system model. In Proposed Scheme section, we describe our proposed statistical-based scheme for data gathering in smart grid. In Simulation Results section, the simulation results of our scheme are presented. Finally, we conclude the paper in the last section.

II. RELATED WORK

Several algorithms for data gathering in smart grids have been studied in literature. In this section, we briefly review various privacy-preserving schemes for data gathering in smart grids.

In [3], an algorithm of data collection with self-awareness protection is proposed. They considered data collectors and respondents in their scheme and expressed that some of the respondents may not participate in contributing their personal data or submit erroneous data. To overcome this issue a self-awareness protocol was studied to enhance trust of the respondents when sending their personal data to the data collector. All respondents collaborate with each other to preserve their privacy. The authors hired an idea, which allows respondents to know protection level before the data submission process is initiated. The paper is motivated by [4] and [5]. In [4], co-privacy (co-operative privacy) is introduced. Co-privacy claims that best solution to achieve privacy is to help other parties to achieve their privacy. More of co-privacy can be found in [4].

Many researchers focused on self-oriented privacy protection. One of the most interesting ones is [6] which proposes self-enforcing privacy (SEP) for e-polling. In this scheme, pollster must allow the respondents to track their submitted data in order to protect their privacy. In this case, respondents can accuse the pollster based on data they gathered during the collection process. Following this idea, a fair approach for accusation is presented in [7]. In [8], a respondent-defined privacy protection (RDPP) is introduced. It means that respondents are allowed to determine their required privacy protection level before delivering data to data collector. The main difference of this method is that unlike other methods, which data collector decides about the privacy protection level, respondents can freely define the privacy protection level.

To obtain privacy of residential users, a scheme named APED is proposed in [9]. It employs a pairwise private stream aggregation. They have shown that their scheme achieves privacy preserving aggregation and also executes error detection when some nodes fail to function normally. DG-APED is an improved form of APED, suggested in [10]. DG-APED propounds diverse grouping-based protocol with error detection. This research added differential privacy technique to APED. Moreover, DG-APED has an advantage of being efficient in term of communication and computation overhead compared to APED.

Authors in [11] first presented a new kind of attack, which adversary extracts information about the presence or absence of a specific person to access the smart meter information. They named this type of attack, human-factor-aware differential aggregation (HDA) attack and claimed that other proposed protocols cannot handle it. To solve this issue, they introduced two privacy-preserving protocols, a basic one and an advanced one.

They corroborated that their research can stand out against HDA attack by transmitting encrypted measurements to an aggregator in a way that aggregator cannot steal any information of human activities. By some implementations, it is demonstrated that the proposed method in [11] can guarantee privacy.

PDA is a scheme presented in [12]. It is a privacy-preserving dual-functional aggregation technique for smart grids in which, every user disseminates only one data and then data and control centre computes two statistical averages (mean and variance) of all users. Their simulations show that PDA is efficient concerning computational and communication overheads. The authors of [12], continued their researches leading to a privacy-preserving data aggregation with fault-tolerance called PDAFT [13]. In this work, a strong adversary is not able to gain any information, even in the case of compromising a few servers at the control centre (CC). Like PDA, PDAFT has a good communication overhead and is tenacious against many security threats. In a condition, which some users or servers fail, PDAFT can still work and this is the reason why they claimed that their proposed method has the fault-tolerance feature. DPAFT [14] is another privacy-preserving data collection scheme which supports both differential privacy and fault tolerance at the same time. It is claimed that, DPAFT surpass other schemes in many aspects, such as storage cost, computation complexity, utility of differential privacy, robustness of fault tolerance, and the efficiency of user addition or removal [14]. A new malfunctioning data aggregation scheme, named MuDA, is introduced in [15]. It is resistant to differential attacks and keeps users' information secret with an acceptable noise rate. PDAFT [15], DPAFT [14], and MuDA [15], shows nearly same characteristics. Their difference is in the cryptographic methods they use [16]. PDAFT employs homomorphic Paillier cryptosystem [17], while DPAFT and MUDA use Boneh-Goh-Nissim cryptosystem [18].

The paper [19] presents a secure power usage data aggregation for smart grid. By this method, supplier understands usage of each neighbourhood and makes decision about energy distribution, while it has no idea of the individual electricity consumption of each user. This scheme is designed to barricade internal attacks and provide batch verification. Authors of [20] found out that [19] has the weakness of key leakage and the imposter can obtain the private key of user easily. It is proved that by using the protocol in [20], key leakage problem is solved and a better performance in term of computational cost is achieved. Neglecting energy cost is the disadvantage of this method.

Some other researches are also investigated in the field of privacy-preserving data collection. For example, in [21], authors designed a balanced anonymity and traceability for outsourcing small-scale linear data aggregation (called BAT-LA) in smart grid. They designed their protocol with the concern of providing both anonymity and traceability. Anonymity means that users' identity should be kept secret and traceability means that imposter users should be traced. Another challenge is that many devices are not capable of handling required complicated computations. Hence, they hired the idea of outsourcing computations with the help of public cloud. Authors of [21] utilized elliptic curve cryptography and proxy re-encryption to make BAT-LA secure. BAT-LA is evaluated by comparing it to two other schemes, RVK [22], and LMO [23] and it is shown that BAT-LA is more efficient in terms of confidentiality compared to the other two schemes [16].

The manuscript [24], a privacy-preserving protocol for smart grid is designed, which outsources computations to cloud servers completely. In this protocol, the data is encrypted before outsourcing and consequently cloud can perform any computations without decrypting data. It is claimed that their work became secure and efficient by using a multi-server framework. The paper [25] adopts perturbation techniques to preserve privacy and uses perturbation techniques and cryptosystems at the same time. This is designed in a way to be suitable for hardware-limited devices. Evaluations show that [25] is resilient to two types of attack, filtering attack, and true value attack. Authors of [26] divided their contribution to two parts. First it is described how an individual meter shares its readings to multiple users, and then the second part, where a user receives meter readings from multiple meters. Finally, they proposed a polynomial-based protocol for pricing. TPS3 [27] is security protocol, which is got its idea from Temporal Perturbation and Shamir's

Secret Sharing (SSS). Using both of these schemes simultaneously, makes it harder for adversary to obtain critical data of users. TPS3 guarantees privacy and reliability of users' data and begets a trade-off between communication cost and security. In [28], data collector tries to preserve privacy by adding some random noise to its computation result. To overcome the problem of computation accuracy reduction, an approximation method is proposed in [28] which leads to obtain a closed form of collector's decision problem.

In [29], a slightly different scenario is considered which data collector collects data from data providers and then spread it to data miner. The goal is to preserve providers' data privacy. Anonymization might be an answer, but it has its own challenges. To achieve a trade-off between privacy protection and data utility, interactions among three elements of scenario (data providers, data collector, and data miner) is modelled as a game and the Nash equilibria of the game is found. Simulations prove that the founded trade-off made an improvement to previous researches.

Some of the reviewed researches, such as [21] and [24] focused on outsourcing to clouds or distributed systems and prior to this, an encryption improves the security significantly. Based on which encryption method we use, it is important to use a secure key management scheme. The cryptographic technique ensures that no privacy sensitive information would be revealed. But, there is still the challenge of how to efficiently query encrypted multidimensional metering data stored in an untrusted heterogeneous distributed system environment [30]. The paper focused on this challenge and introduced a high performance and privacy-preserving query (P2Q) scheme and shows that it brings confidentiality and privacy in a semi-trusted environment.

III. SYSTEM MODEL

In this section, we present our system model. The essential elements of our SPBB approach include:

i. *Consumer*: those who consume energy in a grid.

ii. *Benign Consumer*: a consumer who reported its power consumption correctly.

iii. *Malicious Consumer*: a consumer who reported its power consumption wrongly due to some purposes such as fraud or subversive goals.

iv. *Supplier*: an entity whose responsibility is to provide energy for power consumers in a region.

v. *Data Aggregator*: a local server whose liability is gathering the amount of power consumption information from consumers periodically and dispatching the gathered data to a supplier.

vi. *Electricity Leakage*: the difference between the actual amount of consumed energy and the sum of quantity expressed by consumers as their power consumption.

Consider a grid consisting of M regions, each comprises one data aggregator and n_j consumers where j denotes the index of the region, that is $j \in \{1, \dots, M\}$. Consumers send their power consumption information measured by smart meters to the local aggregators. Data aggregators are responsible of gathering local data and sending it to the power supplier with a specific mechanism which will be presented in the subsequent section.

It is assumed that data aggregators are trusted. Indeed, no information leakage occurs at data aggregators, supposedly because after aggregation takes place, no raw information concerning power consumption of consumers would be at hand.

Besides, we assume that connections among above entities are secured by means of some cryptographic shared or public keys. Since smart meters on consumers' side cannot perform high computationally complex calculations, utilization of public key cryptography may not be sensible. Thus, employment of secret key cryptography would be a better option.

IV. PROPOSED SCHEME

In this paper we propose a method for data gathering with the purpose of informing the supplier of the instant power consumption. This algorithm provides the supplier with enough information about the demand for the power in the grid. Consequently, the power energy is produced based on the instantaneous requirement and this would prevent waste of energy and supplies.

A. Data Gathering

Although the accuracy of smart grids' performance is engaged with the correctness of data gathered from consumers, this data gathering should not be in contrast with the privacy of consumers.

In this section we present a method for data gathering in smart grids which provide suppliers with data while keeping the users' power consumption information private and more importantly, find malicious consumers who try to send erroneous data to suppliers. We refer to this method as SBPP approach.

The proposed SBPP scheme for data gathering works as the following:

i. Consumers send their power consumptions periodically to a local centre called data aggregator.
ii. Each data aggregator selects one consumer randomly in each period.
iii. It aggregates the power consumption of all consumers in that period except the randomly selected one.
iv. Each data aggregator sends the aggregated amount of the previous step in accompany with the power consumption of the randomly selected consumer to the supplier.
v. The supplier provides energy based on the received power consumptions from data aggregators.

Figure 1 depicts how data gathering takes place. It is assumed that data aggregators are trusted, then power consumption information are not at hand any more after being aggregated by the data aggregators and being sent to the supplier. By this assumption, instead of having access to power consumption information of everyone at any period, a little portion of information is available about power consumption of each consumer. Suppose, for instance, there exist 100 consumers in a region with one data aggregator and let the period of data gathering be every 15 minutes. Without any data gathering algorithm, consumers would send their power consumption information to the supplier 2880 times (30*24*60/15) in a month, instead, by utilization of the above algorithm for data gathering, we have access to 0.01 of information corresponding to power consumption of users, that is, at most 29 times (0.01*2880) in a month.

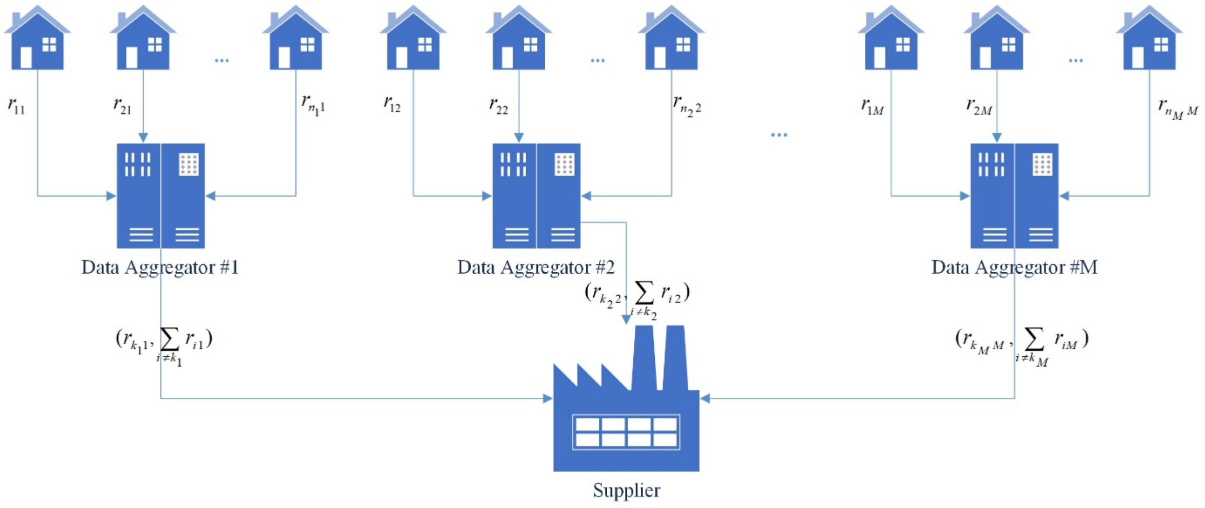


Figure 1: How power consumption information is sent to the supplier by data aggregators. Let P_{ij} be the power consumed by consumer i in region j and let k_j denotes the index of randomly chosen consumer in region j .

On the other hand, by utilization of the SBPP algorithm for data gathering, only 29 information regarding the power consumption of each consumer is available at the supplier in an analogous period. Although it may seem that having access to power consumption information of consumers is in contradiction with their privacy, availability of these information 29 times a month would not reveal any data concerning their life style compared with approachability of these information 2880 times within a month.

B. Finding Malicious Consumers

Malicious consumers pursue two distinct aims by sending erroneous data to suppliers. Either they declare their amount of power consumption lesser than their real consumed power to pay lower fee; or, they express their power consumption quantity much more so as to impose more expenditure to the supplier.

In this paper, we get use of correlation coefficient of power consumption of consumers to find malicious consumers in each region who try to send erroneous data to the supplier.

Correlation coefficient illustrates the statistical relationship between two variables and it is defined as follows:

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{cov(X,X)cov(Y,Y)}} \qquad (1)$$

where $corr$ is a widely used alternative notation for the correlation coefficient and cov means covariance. Correlation coefficient possesses values in the range of -1 to +1, where -1 and +1 indicate the strongest possible agreement and disagreement respectively.

In order to find malicious consumers, it is assumed that data aggregators are aware of the total amount of power consumed in each region. By comparing this amount with the aggregated quantity declared by consumers, the shortage amount can be determined.

Having access to merely one quantity of power consumption information corresponding to a consumer does not suffice to distinguish if that consumer is benign or malicious. In other words, the more information we have regarding power consumption of each consumer, the better decision we can make about the sabotage of consumers. Thus, the algorithm for finding malicious consumers takes place at the end of each month.

So as to detect malicious consumers, each data aggregator stores the identity (ID) of the randomly selected consumer, its declared power consumption, and the leakage amount of power consumed in that region at every period. At the end of each month, for each consumer, the data aggregator computes the correlation coefficient of its reported consumed energy and the leakage amounts of power consumption. Henceforth, we define the leakage quantity as:

$$leakage = actual\ amount - reported\ amount \qquad (2)$$

If the correlation coefficient turns to +1 for a consumer (according to (2), it means that consumer had expressed its power consumption less than its actual used power. On the other hand, if the correlation coefficient for a user turns to -1, it means that consumer is declaring its power consumption more than its usage due to some subversive goals. Thus, the proposed scheme is capable of not only detecting malicious users, but also comprehending if that user is declaring its amount of power consumption less or more than its actual quantity.

Furthermore, it is possible that there exists more than one malicious user in a region. In this case, although the correlation coefficient corresponding to these users would not be equal to ± 1, their correlation coefficient quantity will be maximum (or minimum) amongst other consumers. As a result, it is needed that a threshold (th) be defined where the absolute value of correlation coefficients fewer or more than the threshold indicate benign or malicious users respectively, as:

$$\begin{cases} malicious\ user, & -1 \leq corr \leq -th \\ benign\ user, & -th \leq corr \leq -th \\ malicious\ user, & th \leq corr \leq 1 \end{cases} \qquad (3)$$

It is apparent that the more the threshold is, the less malicious consumers are detected and on the other hand, the less the threshold is, the more benign users are considered malicious. Thus, a question that arises here is that *how should this threshold be determined?* The analysis concerning the detection of several malicious users in a region is left for future works, however, we briefly discuss the problem in the following. In this paper, according to the setting of the problem, we set the threshold to a fixed value namely 0.5.

As the proposed scheme is a statistical one, it is probable that the correlation coefficient of a benign user lies out of its defined region depicted in (3), or vice versa, that is, the correlation coefficient corresponding to a malicious consumer lies in the region belonging to benign ones.

C. Billing

In this section, we propose an algorithm for billing. As discussed in the preceding section, malicious consumers can be distinguished by computing correlation coefficient of all consumers in a region. Malicious consumers' being determined, sent data corresponding to other consumers are considered trustworthy and error free. By this assumption, the liability for billing can be assigned to data aggregators. In every period, consumers send their amount of consumed energy to data aggregators. Based on the received data from consumers and the received tariffs from the supplier, data aggregators compute the cost of consumed power for each consumer before data aggregation takes place. In each period, data aggregators calculate the cost of consumed power for each consumer and add the cost to the previously calculated cost for that consumer and by the end of month, a bill will be issued and sent to each consumer.

Not only this algorithm decreases the signalling overhead, but also the privacy of consumers would be protected. It is merely required that suppliers send tariffs periodically to data aggregators and consumers simultaneously. Data aggregators compute the cost of consuming energy for every consumer and smart meters on the consumers' side adjust the power consumption based on the received tariffs, i.e., if tariff increases, smart meters force dispensable devices to be turned off. In this case, no information leakage and thus no privacy invasion would occur.

Besides, by finding malicious consumers in each region and by comparing the amount of power consumed by other consumers and the total amount of produced energy, the power consumption quantity of malicious consumers would be determined. However, that how the bill of these malicious consumers should be calculated and what penalties should be intended for these consumers are not considered in this paper.

V. SIMULATION RESULTS

In this section, we present the results of simulations for the proposed SBPP approach. We would show that our proposed scheme can detect malicious users who send bogus information concerning their power consumption quantity in a smart grid.

Consider a region consisting of 100 consumers and one data aggregator where data aggregation takes place every 15 minutes and assume that consumer # 25 is a malicious user. Two cases are studied; user # 25 in case (a) expresses one tenth of its power consumption and in case (b) it reports its power usage 10 times more than its actual consumption. Figure 2 (a) illustrates case (a) where the correlation coefficient of expressed consumed energy and the leakage amounts of power consumption turns to +1 and Figure 2 (b) depicts case (b) where the correlation coefficient turns to -1.

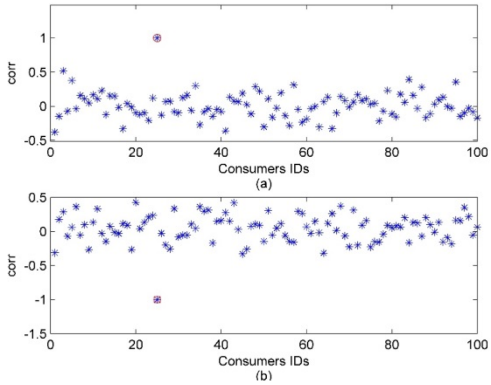
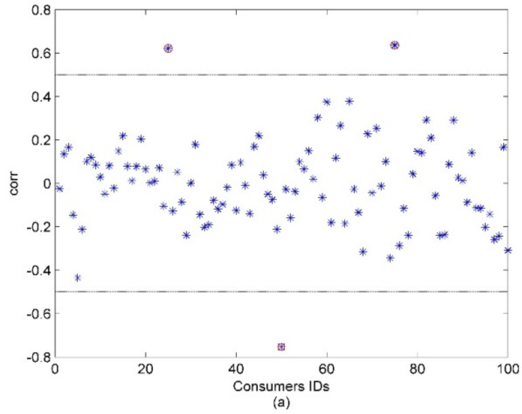


Figure 2: Correlation coefficient of reported energy consumption and the leakage amounts of power consumption for all users in the grid. (a) One malicious user declares its power consumption less than the actual quantity and (b) One malicious user declares its power consumption more than the actual quantity

Consider the previous assumptions except that there are three malicious consumers instead of one in that region with IDs 25, 50, and 75. Consumers with IDs 25 and 75 declare their power consumption less than their actual consumption and consumer # 50 expresses its power consumption more than its actual consumed energy. By setting the threshold to 0.5, consumers with absolute value of correlation coefficient greater than 0.5, that is, $|corr| \leq 0.5$, would be considered malicious, as depicted in Figure 3.



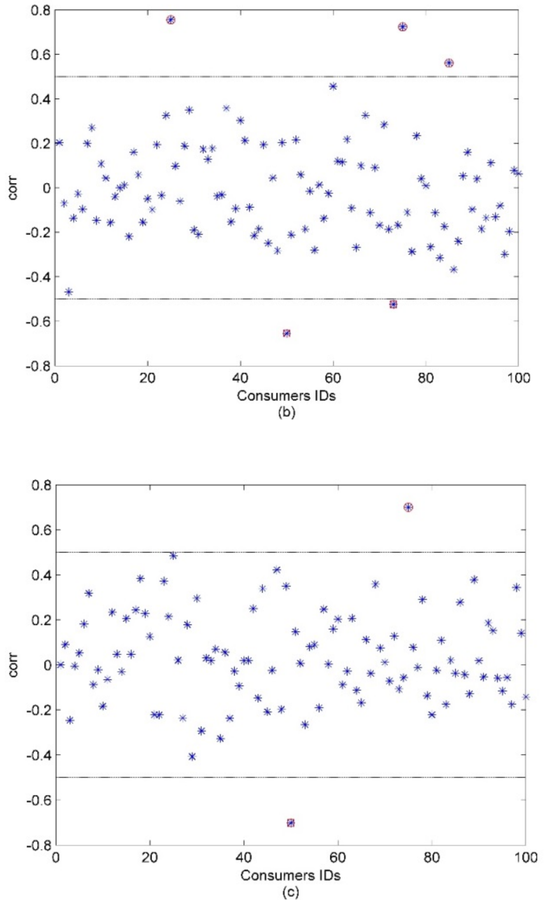


Figure 3: Detection of several malicious users (a) all malicious user are detected correctly, (b) in addition to malicious users, a number of benign users are found malicious, and (c) not all malicious users are detected.

As it can be seen from Figure 3, fixed threshold will result in 3 cases: 1) only malicious users been detected (Figure 3 (a)), 2) in addition to malicious users, some benign users found malicious (Figure 3 (b)), and 3) a subset of malicious users been detected (Figure 3 (c)).

VI. CONCLUSION

We presented a statistical-based approach for data gathering in smart grids which preserves the privacy of consumers. We investigated the capability of the proposed scheme in detecting malicious consumers who dispatch bogus data to service providers for a specific purpose such as abating their cost or imposing expenditure on suppliers (subversive goals). Furthermore, we showed that if there exists only one malicious user, it can definitely be detected if enough number of samples are gathered. When there are more malicious users, depending on the number of gathered samples, it is probable that all malicious consumers being detected, some benign consumers found malicious, or a subset of malicious users being detected. We also presented a scheme for billing which concede the liability of billing to data aggregators in each region. By employing this scheme, not only the signalling overhead decreases significantly, but also billing occurs at a trusted entity where malicious consumers are distinguished from benign ones. Our simulation results verified these terms.

VII. REFERENCES

[1] E. Fadel, V. C. Gungor, L. Nassef, N. Akkari, M. A. Malik, S. Almasri, and I. F. Akyildiz, “A survey on wireless sensor networks for smart grid,” *Computer Communications*, vol. 71, pp. 22–33, 2015.

[2] A. Alabdulatif, H. Kumarage, I. Khalil, M. Atiquzzaman, and X. Yi, “Privacy-preserving cloud-based billing with lightweight homomorphic encryption for sensor-enabled smart grid infrastructure,” *IET Wireless Sensor Systems*, vol. 7, no. 6, pp 182-190, 2017

[3] K.-S. Wong Wong and M. H. Kim, “Privacy-preserving data collection with self-awareness protection,” in Frontier and Innovation in Future Computing and Communications. Springer, 2014, pp. 365–371.

[4] J. Domingo-Ferrer, “Coprivacy: towards a theory of sustainable privacy,” in *International Conference on Privacy in Statistical Databases*. Springer, 2010, pp. 258–268.

[5] J. D. Ferrer, “Coprivacy: an introduction to the theory and applications of co-operative privacy,” *SORT: statistics and operations research transactions*, pp. 0025–40, 2011.

[6] P. Golle, F. McSherry, and I. Mironov, “Data collection with self-enforcing privacy,” *ACM Transactions on Information and System Security (TISSEC)*, vol. 12, no. 2, p. 9, 2008.

[7] M. Stegelmann, “Towards fair indictment for data collection with self-enforcing privacy,” in *IFIP International Information Security Conference*. Springer, 2010, pp. 265–276.

[8] R. Kumar, R. Gopal, and R. Garfinkel, “Freedom of privacy: anonymous data collection with respondent-defined privacy protection,” *INFORMS Journal on Computing*, vol. 22, no. 3, pp. 471–481, 2010.

[9] R. Sun, Z. Shi, R. Lu, M. Lu, and X. Shen, “Aped: An efficient aggregation protocol with error detection for smart grid communications,” in *Global Communications Conference (GLOBECOM), 2013 IEEE*. IEEE, 2013, pp. 432–437.

[10] Z. Shi, R. Sun, R. Lu, L. Chen, J. Chen, and X. S. Shen, “Diverse grouping-based aggregation protocol with error detection for

smart gri grouping-based aggregation protocol with error detection for smart grid communications," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2856–2868, 2015.

[11] W. Jia, H. Zhu, Z. Cao, X. Dong, and C. Xiao, "Human-factor-aware privacy-preserving aggregation in smart grid," *IEEE Systems Journal*, vol. 8, no. 2, pp. 598–607, 2014.

[12] C. Li, R. Lu, H. Li, L. Chen, and J. Chen, "Pda: a privacy-preserving dual-functional aggregation scheme for smart grid communications," *Security and Communication Networks*, vol. 8, no. 15, pp. 2494–2506, 2015.

[13] L. Chen, R. Lu, and Z. Cao, "Pdaft: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-to-peer networking and applications*, vol. 8, no. 6, pp. 1122–1132, 2015.

[14] H. Bao and R. Lu, "A new differentially private data aggregation with fault tolerance for smart grid communications," *IEEE Internet of Things Journal*, vol. 2, no. 3, pp. 248–258, 2015.

[15] L. Chen, R. Lu, Z. Cao, K. AlHarbi, and X. Lin, "Muda: Multifunctional data aggregation in privacy-preserving smart grid communications," Peer-to-peer networking and applications, vol. 8, no. 5, pp. 777–792, 2015.

[16] M. A. Ferrag, L. A. Maglaras, H. Janicke, and J. Jiang, "A survey on privacy-preserving schemes for smart grid communications (2016)," *arXiv preprint arXiv:1611.07722*, 2016.

[17] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 1999, pp. 223–238.

[18] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-dnf formulas on ciphertexts," in *Theory of Cryptography Conference*. Springer, 2005, pp. 325–341.

[19] C.-I. Fan, S.-Y. Huang, and Y.-L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial informatics*, vol. 10, no. 1, pp. 666–675, 2014.

[20] D. He, N. Kumar, and J.-H. Lee, "Privacy-preserving data aggregation scheme against internal attackers in smart grids," *Wireless Networks*, vol. 22, no. 2, pp. 491–502, 2016.

[21] H. Wang, D. He, and S. Zhang, "Balanced anonymity and traceability for outsourcing small-scale data linear aggregation in the smart grid," *IET Information Security*, vol. 11, no. 3, pp. 131–138, 2016.

[22] H. Wang, B. Qin, Q. Wu, L. Xu, and J. Domingo-Ferrer, "Tpp: Traceable privacy-preserving communication and precise reward for vehicle-to-grid networks in smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2340–2351, 2015.

[23] C. Rottondi, G. Verticale, and C. Krauss, "Distributed privacy-preserving aggregation of metering data in smart grids," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1342–1354, 2013.

[24] H. Chun, K. Ren, and W. Jiang, "Privacy-preserving power usage and supply control in smart grid," *Computers & Security*, 2018.

[25] U. B. BALOGLU and Y. DEMIR, "Lightweight privacy-preserving data aggregation scheme for smart grid metering infrastructure protection," *International Journal of Critical Infrastructure Protection*, 2018.

[26] A. Rial, G. Danezis, and M. Kohlweiss, "Privacy-preserving smart metering revisited," *International Journal of Information Security*, vol. 17, no. 1, pp. 1–31, 2018.

[27] M. U. Simsek, F. Yildirim Okay, D. Mert, and S. Ozdemir, "Tps3: A privacy preserving data collection protocol for smart grids," *Information Security Journal: A Global Perspective*, vol. 27, no. 2, pp. 102–118, 2018.

[28] G. Liao, X. Chen, and J. Huang, "Optimal privacy-preserving data collection: A prospect theory perspective," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[29] L. Xu, C. Jiang, Y. Qian, Y. Ren, L. Xu, C. Jiang, Y. Qian, and Y. Ren, "Privacy-preserving data collecting: A simple game theoretic approach," *Data Privacy Games*, pp. 45–57, 2018.

[30] R. Jiang, R. Lu, and K.-K. R. Choo, "Achieving high performance and privacy-preserving query over encrypted multidimensional big metering data," *Future Generation Computer Systems*, vol. 78, pp. 392–401, 2018.