A Comprehensive Survey of Privacy-Preserving in Smart Grid

Guanlin Si[1], Zhitao Guan[1(✉)], Jing Li[1], Peng Liu[2], and Hong Yao[3]

[1] School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
guanzhitao@126.com
[2] School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China
[3] School of Computer Science, China University of Geosciences, Wuhan 430074, China

Abstract. As the next generation of power system, smart grid provides people with great convenience in efficiency and quality. It supports two-way communication and extremely improves the efficiency of utilization of energy resource. In order to dispatch accurately and support the dynamic price, a lot of smart meters are installed at users' house to collect the real-time data or power plan. Besides, many entities in smart grid need to share some necessary data. However, all these collected data are related to user privacy. In this paper, we make a comprehensive survey of privacy-preserving in smart grid, analyze the current privacy problems and list the corresponding solutions from a holistic angle. At last, we discuss the future work and make the conclusion.

Keywords: Privacy-preserving · Homomorphic encryption · Blind signature · Secret sharing scheme · Bilinear pairing · Smart grid

1 Introduction

As a new generation of energy network, smart grid is considered a useful way to solve the severe environment and resource problems. It is the product of the combination of energy network and information technology. Differing from the unidirectional centralized grid, the control mode of the smart grid is more flexible and reliable. It supports bidirectional power flow between the users and grid, that is to say, the user in smart grid is not only a consumer but also a generator. Smart grid can supply users with electricity, on the contrary, the users can also provide smart grid with their superfluous electricity which comes from their household energy. What's more, to realize the optimal scheduling, smart grid installs a smart meter at each house to collect the real-time electricity data, draw the real-time load curve, and create the plan for electricity generation. Not only that, smart grid also adopts many new service modes. For example, the dispatch center can make the dynamic price to encourage users to adjust their power consumption behaviors, collect the electricity consumption requirements and create the electricity generation plan in advance.

G. Wang et al. (Eds.): SpaCCS 2016, LNCS 10066, pp. 213–223, 2016.
DOI: 10.1007/978-3-319-49148-6_19

Although smart grid has many advantages in those aspects, there also exist several risks which may disclose user privacy in some degree. An adversary can infer user's family behaviors through his real-time electricity data. For example, when you get up, when you go to work, when you take a shower and so on. Thus, user privacy can't be guaranteed even he is at his own home, and thieves may gain entry to user's house when they notice that there is nobody home. Therefore, the privacy-preserving in smart grid becomes an extremely important problem which can hamper the implement of smart grid.

For the privacy-preserving in smart grid, many scholars proposed various solutions. In this paper, we analyzed some correlative problems in smart grid which are related to user privacy. Later, we summarized the solutions for the privacy-preserving from a holistic angle. Privacy-preserving and authentication are two closely related issues in smart grid. We also analyzed correlative problems about authentication. What's more, we analyzed the flaws of the various solutions and pointed out our future work. The service model of smart grid is showed in Fig. 1.

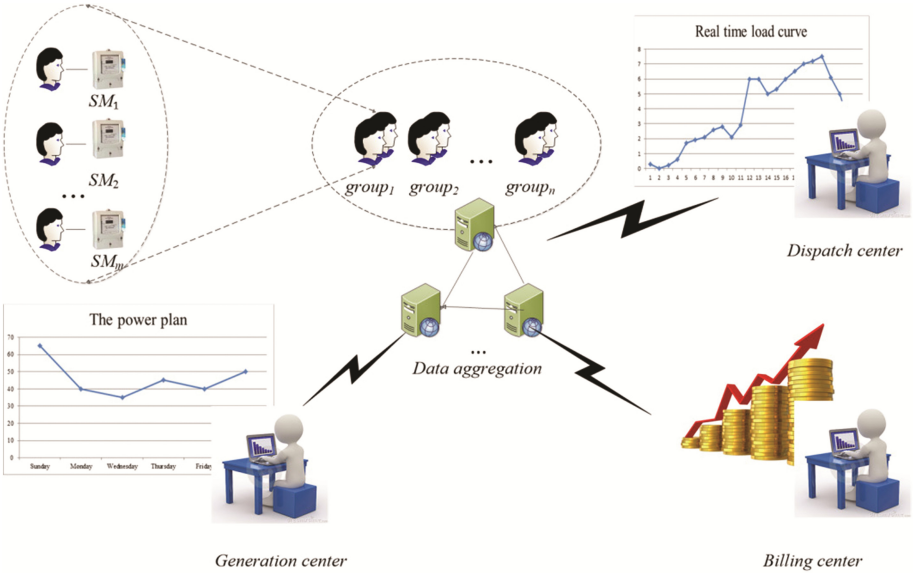

Fig. 1. The service model of smart grid

This paper is organized as follows. We showed the basic theory of the privacy-preserving in Sect. 2. Then, we described the privacy problems and protection strategies in smart grid in Sect. 3. Section 4 contained the main solutions to solve the privacy and authentication problems. Besides, we pointed out the future work in Sect. 5 and made a conclusion in Sect. 6.

2 Preliminaries

2.1 Homomorphic Encryption

Homomorphic encryption is a data aggregation encryption which can operate the cipher text to achieve ideal effect without knowing the plaintext. For example, there are three entities in our system: sender, intermediate, and receiver. All of the senders encrypt their own values and send them to the intermediate. The receiver wants to achieve the summary of the values from all the senders. In order to achieve this purpose, the intermediate can multiply these cipher texts and send the result to the receiver, thus, receiver can achieve the summary of the values coming from the senders while the intermediate provides the middle service without knowing the real values. We call this algorithm additive homomorphic encryption. Paillier encryption and Bone-Goh-Nission encryption are two common additive homomorphic algorithms.

2.2 Blind Signature and CL-Signature

A user times his data with a random number called blind factor and sends the result to the third party. Later, the third party authenticates the user's identity, signs the result with its private key and returns the signed data to the user. Thus, the user can obtain the right signature by multiplying the signed data with his inverse of the blind factor, while the third party doesn't know the content of the user's data. We call this algorithm blind signature.

The CL- Signature is similar to the blind signature, and it can realize that a user could obtain a signature while the signer has no information about the value. The detailed process is as follows:

Chooses two big primes p, q and calculates $n = pq$. $h, a, g_1 \ldots g_n$ are random numbers. Publish $(n, h, a, g_1 \ldots g_n)$ and keep p as the private key.

Users set their data $m_1 \ldots m_n$ as $g_1^{m_1} g_2^{m_2} \ldots g_n^{m_n}$ and send them to the signer. The signer calculates $l_r = l_m + l_n + l_s$ where l_m denotes the length of m_i and l_n denotes the length of n. l_s is a random number selected by signer. Then, the signer chooses a parameter e which satisfies $l_e > l_m + 2$. At last, calculates a parameter v through

$$v^e = a g_1^{m_1} g_2^{m_2} \ldots g_n^{m_n} h_r \bmod n \tag{1}$$

(e, v, r) is the CL- Signature and verifier can verify the signature by checking $v^e = a g_1^{m_1} g_2^{m_2} \ldots g_n^{m_n} h_r \bmod n$.

2.3 Secret Sharing Scheme

The secret sharing scheme is a scheme which splits a secret into n pieces and distributes these pieces with different valid members. If an adversary captures a member in the system, he can only get a piece of the secret. Only if the adversary gets at least k pieces of the secret, can he get the whole secret. We usually adopt the Shamir technique to realize this result.

The trusted party chooses a polynomial to split a secret denoted by s

$$f(x) = s + r_1 x + r_2 x^2 + \ldots r_{k-1} x^{k-1} \bmod p \tag{2}$$

(x_i, y_i) is the corresponding share. Remarkably, the Shamir secret sharing scheme is the fully homomorphic and can be designed as a better scheme to realize the data aggregation.

2.4 Bilinear Pairing

Let G be an additive group and P, Q are the generators of G. e:$G \times G = G_t$, where G_t is a multiply group. The bilinear pairing satisfies several properties:

$$e(P, Q)^x = e(P, Qx) = e(Px, Q) \tag{3}$$

$$e(P, Q)^{x_1} e(P, Q)^{x_2} = e(P, Q)^{x_1 + x_2} \tag{4}$$

$$e(P, P) \neq 1 \tag{5}$$

2.5 Commitment and BBS+ Signature

Commitment is a scheme which can commit verifier a value without revealing it. After a period of time, the sender reveals the value and the verifier can judge whether the value is the same as the previous one. Commitment is often used for verifying the message authenticity. A famous commitment scheme called Pedersen Commitment is created as follows: let g and h be the generator of a group. To commit a value m, the verifier chooses a random number r and computes $C = g^m h^r$ as the commitment.

BBS+ Signature is a partly blind signature based on the Pedersen Commitment. For a bilinear pairing: e:$G \times G = G_t$, let g, g_0, g_1 be generators of G. A sender chooses a random number r and computes $w = g^r$ as public key. To sign a message m, the signer computes $A = (g^c g_0^z g_1^m)^{1/(c+r)}$. While, c and z are random numbers chose by the signer. Then, (A, c, z) is the BBS+ signature and one can verify it by checking if

$$e(A, wg^c) = e(gg_0^z g_1^m, g) \tag{6}$$

3 Privacy Problems and Protection Strategies in Smart Grid

3.1 Privacy Problems

Smart grid can provide people with great efficiency and economy, while, it may disclose user privacy in some degree. We show the privacy-risks as follows:

- For dealing with the electrical fault which may happen at any time, smart grid needs to collect the real-time electricity data from all of the users to watch the status of the system. However, this may disclose user's real-time privacy. By analyzing the real-time electricity curve, user's family behavior would be disclosed easily.

- For saving power as much as possible, smart grid collects the power plan from all of the users in advance. Obviously, the power plan from users will disclose their behaviors for a period of time in the future. For example, smart grid can infer that your family will go on a trip if your power plan is closed to zero.
- There are many entities in smart grid such as power plant, transformer substation, control center and billing center. In order to achieve a better control, all the entities need to share some necessary data. However, the relationship of these entities is competitive and cooperative. Some important data are related to the core interests of an entity and should not be disclosed to any other entities. Therefore, all the entities in smart grid need to establish a security data sharing program to protect each entity's privacy.

As the first situation, the privacy-preserving scheme is related to real-time electricity data, so it may affect the dynamic price in smart grid. Therefore, when we need to establish a privacy-preserving scheme, we must ensure the scheme doesn't disturb the normal billing. For the second situation, there also exists a problem. A malicious user may send smart grid a wrong power plan which is much larger than its real consumption; thus, smart grid will produce much unnecessary electricity and lead to power waste. So, we need to authenticate all the users and ensure the validity of the power plan when we establish our privacy-preserving scheme.

3.2 Protection Strategies

To protect the privacy of users in smart grid, there exist three basic strategies currently.

We can protect the identity of each user. Even an adversary or dispatcher can obtain user's accurate data, but not knowing the identity of the user can still protect the user privacy.

We can protect the data of each user. Through some schemes such as data aggregation or obfuscation, we can ensure that the control center runs in a right way while has no information about any user's data.

We can protect the route between the sender and receiver. Thus, if an adversary captures a message through the network, he can't ensure the sender and receiver through the message.

Authentication is closely related to privacy in smart grid and it contains the following aspects. *a.* Message integrity

When the control center receives a message sent by smart meter, we must ensure the integrity of the message. Given the privacy-preserving, the best strategy is that the control center can guarantee the message integrity while has no information about the content.

b. Validity of identity

When the control center receives a message from a user, we must ensure the identity of the user is validity and prevent the adversary from impersonating a legal user. Given

the privacy-preserving, the best strategy is that the control center can guarantee the validity of user's identity while has no information about user's real identity.

c. Message authenticity

Apart from the message integrity and the validity of user's identity, the message authenticity can't be ignored. For example, a malicious user may send a false power plan which is much larger than his real power consumption. Thus, smart grid will generate much unnecessary power, which causes a serious waste of energy. So, the best strategy is that smart grid can guarantee the authenticity of a message but has no information of the content or the sender's identity.

4 Countermeasures

Based on the aforementioned basic strategies, we can list the related solutions. We summarized some basic and classical solutions according to basic strategies at Table 1.

Table 1. Privacy strategies and major solutions

Privacy-strategies	Solutions
Mask identity	Virtual ring
	CL signature
	Blind signature
Mask data	Home battery
	BGN encryption
	Paillier encryption
	Data obfuscation
	Bilinear mappings
	Shamir secret sharing scheme
Mask route	Random re-transmission
Authentication	CL signature
	Blind signature
	BBS+ Signature
	Commitment

4.1 Privacy-Preserving by Masking the User's Identity

For masking the user's identity, a simple solution adopting a trusted- party to manage the identity list is proposed in [1]. However, finding a trusted-party is not easy. We have many better choices to select.

Some scholars proposed a scheme based on blind signature to solve the privacy-preserving and validity-authentication in [2]. The main idea of this scheme is as follows: a user times his data with a random number called blind factor and sends the result to the third party. Later, the third party authenticates the user's identity, signs the result with its private key and returns the signed data to the user. Thus, the user can obtain the right signature by multiplying the signed data with his inverse of the blind factor, while the third party doesn't know the content of the user's data. The downside of this scheme

is that users should send their electricity data to the third party for authentication before communicating with the control center, which is against the real-time property of the power grid. Camenisch and Lysyanskaya proposed a scheme named CL -Signature scheme which is similar to the blind signature in [3]. It can ensure that a user can obtain a signature while the signer has no information about the value.

An effective scheme based on virtual ring is presented in [4]. It groups the users by their geographical positions and distributes each member in the same group with the same key. In this way, control center can obtain all of the users' data without knowing the senders' ID. Obviously, it's a good way to protect user privacy, but the validity-authentication can't be guaranteed because of the anonymity.

4.2 Privacy-Preserving by Masking the Real-Time Data

A solution using a battery to hide the real-time data is proposed in [5–7]. In these schemes, smart grid and the household battery provide users with electricity at the same time. When the household consumption curve goes high, the battery discharges. Otherwise, it charges. In this way, we can hide the user's real-time data to protect user privacy. The downside is that the effect depends on the battery capacity, besides, charging and discharging the battery frequently is detrimental to the battery life and may collide with dynamic electricity price.

A solution using the data aggregation is also popular in smart grid for privacy-preserving. It often uses homomorphic encryption and there are many algorithms which have the property of homomorphism, such as the Paillier and Bone-Goh-Nission encryption, secret sharing scheme, bilinear mapping and so on. Next, we will show them in details.

The Paillier encryption and Bone-Goh-Nission encryption are classical algorithms as the homomorphic encryption and they are used in many schemes such as [8–12]. However, the computational complexity of them can't be ignored. In [13], scholars try to group the members to realize a distributed authentication in order to reduce the complexity, but the effect is not very ideal. Thus, many scholars try to use other simple algorithms to substitute them.

Secret sharing scheme is proposed to realize the data aggregation. It adopts the Shamir technique to encrypt the electricity data [14]. As we mentioned in Sect. 2, users can choose the same argument x to create their shares like (x, y_i) and the shares have the homomorphic property. Of course, secret sharing scheme can be also applied into other aspects, such as the key management, but this property is not taken into our consideration.

The bilinear mapping is also a common solution for data aggregation. We usually use the formula (4) to create homomorphic encryption such as [16, 23, 25]. Remarkably, we usually use the formulas (3) and (4) to realize the key-exchange.

Besides those mentioned above, there are many algorithms to create the homomorphic encryption, such as [15]. It constructs an equation $C_i = M_i \bullet S + r$, where C_i and M_i stand for the cipher text and plaintext of the electricity data, and S is a common number which is shared by all the members. r is a random number. Smart grid can obtain the summary of M_i easily without disclosing user privacy. Obviously, this scheme does a

good job in complexity, but it also has several flaws. To share the same parameter, all the smart meters have to communicate with each other and reach agreement on the *S*, which would increase the traffic in the network.

While, no matter which algorithm we choose to create the homomorphic encryption, there are always two problems to consider: error-tolerance and differential privacy. These two problems are discussed at [10].

Besides the homomorphic encryption, the data –obfuscation solution is also popular to realize the data aggregation for privacy-preserving. In [17], the author shows a scheme based on data obfuscation, which adds a random number to each electricity data to protect the real-time data from being disclosed by the adversary and control centre. But it will cause some large errors if the random numbers are not reasonable. If we consider the summary of random numbers is zero, then we have to face the problem about error-tolerance.

For the data sharing among different entities, a security data sharing program based on attribute has been presented in [18]. The main idea of this kind of scheme is as follows: the owner of a file sets some attributes in his file, and saves them in the data access centre. If someone wants to access a file, he must satisfy the attributes which are set by its owner. In [18], it not only encrypts a file based on attributes, but also encrypts the attributes themselves. In this way, each entity can set property attributes to protect their privacy during the common data sharing. Because the relationship of these entities is competitive and cooperative, some scholars devote into searching an approach to protect the sensitive information and solving the multi-party cooperation problem such as the optimal power flow in a shared computing platform [20].

4.3 Privacy-Preserving by Masking the Route Between Sender and Receiver

To mask the route between sender and receiver, we usually use the random re-transmission to mask the real route such as [21, 22]. Generally, we often use the topological matrix to achieve a better effect. There is no doubt that random re-transmission would increase the network traffic. Therefore, we need find some better solutions to mask the message route.

4.4 Privacy-Preserving and Authentication

Privacy-preserving and authentication are two closely related issues. As we analysed in Sect. 3, authentication in smart grid contains three aspects.

For the message integrity, there are many solutions to achieve a good effect, such as the message digest. If we want to realize the message integrity, while the verifier has no information about the content of the message, we can adopt the blind signature [2] or CL-signature [3] to achieve this effect.

For the validity of identity, we usually adopt asymmetric encryption to create digital signature. If we want to authenticate the validity of a user's identity, while the verifier has no information about the user's real identity, we can adopt commitment to create BBS+ Signature [23] and create a pseudonym to communicate with the control centre [24].

For the message authenticity, especially for the power plan, commitment is a good choice [19]. A sender firstly sends the control centre encrypted electricity data and a commitment. After a period of time, the sender has to send his real electricity data and the decryption key to the commitment. Thus, the verifier can easily verify whether the two electricity data are the same.

According to the above analysis, the zero-knowledge proof is widely adopted in the authentication related to privacy-preserving in smart grid.

5 Future Works

According to the analysis and aforementioned survey, there exist several disadvantages in current solutions to be solved in future.

As the solution which uses a battery to hide the real-time data, if we don't consider the capacity of the battery, that would be a good choice. But there also exists a severe problem which needs us to consider. The charging and discharging of the battery may conflict with the dynamic price, which would cause huge losses to consumer. That is to say, there is a tradeoff between the privacy-preserving and the dynamic price. Some papers have considered this problem and given some solutions like [11], but people maybe not satisfied with this tradeoff. So, finding a novel solution which can protect user privacy and doesn't damage his economy would be our future work.

As the homomorphic encryption and data-obfuscation, the main concerns are differential privacy and error-tolerance. For the differential privacy, many scholars proposed lots of novel solutions such as adding a random number in each electricity data, where the random number must obey some distribution (For example, it obeys the Laplace distribution) so that the whole aggregation can resist differential privacy [9]. For the error-tolerance, there are still some solutions to solve this problem. A representative solution is proposed by Zhiguo at [10], and it is a solution based on grouping to realize the error- tolerance. Although this solution has advantage in computational complexity to some extent, its accuracy isn't satisfied with our ideal requirement. Therefore, we should find a novel solution to realize the error-tolerance, which has an obvious advantage both in complexity and accuracy.

As we discussed before, we usually adopt Paillier encryption as homomorphic encryption to realize the data aggregation, but the complexity of the Paillier encryption cannot be ignored, which collides with the character of real-time in smart grid. Shamir secret scheme and bilinear mapping would be better choices, but there is not a perfect scheme to make the best of them. So, we had better find a solution to institute the Paillier encryption or create a novel solution which makes the best of the existing algorithms.

6 Conclusion

The current problem about privacy-preserving in smart grid has not been solved perfectly. Many solutions have been proposed, but there still exist some flaws such as complexity and the lack of feasibility. In this paper, we summarized the current privacy problems in smart grid from a holistic perspective and listed corresponding solutions

currently. At last, we discussed the main flaws in the recent solutions and gave the advice for our future work. It's expected that the problem of privacy-preserving would be solved efficiently in the future.

Acknowledgments. This work is partially supported by Natural Science Foundation of China under grant 61402171, Central Government University Foundation under grant JB2016045.

References

1. Efthymiou, C., Kalogridis, G.: Smart grid privacy via anonymization of smart metering data. In: 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm). IEEE (2010)
2. Cheung, J.C.L., et al.: Credential-based privacy-preserving power request scheme for smart grid network. In: 2011 IEEE Global Telecommunications Conference (GLOBECOM 2011). IEEE (2011)
3. Diao, F., Zhang, F., Cheng, X.: A privacy-preserving smart metering scheme using linkable anonymous credential. IEEE Trans. Smart Grid **6**(1), 461–467 (2015)
4. Badra, M., Zeadally, S.: Design and performance analysis of a virtual ring architecture for smart grid privacy. IEEE Trans. Inf. Forensics Secur. **9**(2), 321–329 (2014)
5. McLaughlin, S., McDaniel, P., Aiello, W.: Protecting consumer privacy from electric load monitoring. In: Proceedings of the 18th ACM conference on Computer and communications security. ACM (2011)
6. Yao, J., Venkitasubramaniam, P.: The privacy analysis of battery control mechanisms in demand response: revealing state approach and rate distortion bounds. IEEE Trans. Smart Grid **6**(5), 2417–2425 (2015)
7. Yang, L., et al.: Cost-effective and privacy-preserving energy management for smart meters. IEEE Trans. Smart Grid **6**(1), 486–495 (2015)
8. Marmol, F.G., et al.: Do not snoop my habits: preserving privacy in the smart grid. IEEE Commun. Mag. **50**(5), 166–172 (2012)
9. Bao, H., Rongxing, L.: A new differentially private data aggregation with fault tolerance for smart grid communications. Internet Things J. IEEE **2**(3), 248–258 (2015)
10. Shi, Z., et al.: Diverse grouping-based aggregation protocol with error detection for smart grid communications. IEEE Trans. Smart Grid **6**(6), 2856–2868 (2015)
11. Liang, X., et al.: UDP: usage-based dynamic pricing with privacy preservation for smart grid. IEEE Trans. Smart Grid **4**(1), 141–150 (2013)
12. Chen, L., et al.: MuDA: multifunctional data aggregation in privacy-preserving smart grid communications. Peer-to-peer Netw. Appl. **8**(5), 777–792 (2015)
13. Jo, H.J., Kim, I.S., Lee, D.H.: Efficient and privacy-preserving metering protocols for smart grid systems. IEEE Trans. Smart Grid **1**(1), 65–75 (2015)
14. Barletta, A., et al.: Privacy preserving smart grid Communications by verifiable secret key sharing. In: 2015 International Conference on Computing and Network Communications (CoCoNet). IEEE (2015)
15. Dong, X., Zhou, J., Cao, Z.: Efficient privacy-preserving temporal and spacial data aggregation for smart grid communications. Concurrency Comput. Pract. Experience **50**(9), 98–114 (2015)
16. Akula, P., et al.: Privacy-preserving and secure communication scheme for power injection in smart grid. In 2015 IEEE International Conference on Smart Grid Communications (SmartGridComm). IEEE (2015)

17. Beussink, A., et al.: Preserving consumer privacy on IEEE 802.11 s-based smart grid ami networks using data obfuscation. In: 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE (2014)
18. Hur, J.: Attribute-based secure data sharing with hidden policies in smart grid. IEEE Trans. Parallel Distrib. Syst. **24**(11), 2171–2180 (2013)
19. Chim, T.W., et al.: PRGA: privacy-preserving recording & gateway-assisted authentication of power usage information for smart grid. IEEE Trans. Dependable Secur. Comput. **12**(1), 85–97 (2015)
20. Wu, D., et al.: Preserving privacy of AC optimal power flow models in multi-party electric grids. IEEE Trans. Smart Grid **7**(4), 2050–2060 (2016)
21. Nicanfar, H., et al.: Enhanced network coding to maintain privacy in smart grid communication. IEEE Trans. Emerg. Top. Comput. **1**(2), 286–296 (2013)
22. Rottondi, C., Verticale, G.: Privacy-friendly load scheduling of deferrable and interruptible domestic appliances in smart grids ☆. Comput. Commun. **58**(1), 29–39 (2014)
23. Gong, Y., et al.: A privacy-preserving scheme for incentive-based demand response in the smart grid. IEEE Trans. Smart Grid **7**(3), 1304–1313 (2016)
24. Tan, X., et al.: Pseudonym-based privacy-preserving scheme for data collection in smart grid. Int. J. Ad Hoc Ubiquit. Comput. **22**(2), 120–127 (2016)
25. Chen, J., Shi, J., Zhang, Y.: EPPDC: an efficient privacy-preserving scheme for data collection in smart grid. Int. J. Distrib. Sens. Netw. **11**, 1–12 (2015)