PRIVACY-PRESERVING OF ELECTRICITY DATA BASED ON GROUP SIGNATURE AND HOMOMORPHIC ENCRYPTION

MING AI and HAIQING LIU

ABSTRACT

Smart grid is a highly integrated power system, which combines modern advanced information communication, sensor measurement and automatic control technology, including a large number of smart devices, with the advantages of reliability, reliability, safety and self-healing, providing a great service. However, when collecting, processing, and transmitting large amounts of data, devices such as smart dashboards in the network face the problem of discovering user-specific issues. In order to solve this problem, a scheme based on improved group signature and homomorphic encryption is proposed here. The improved group signature is used to determine the public and private keys of the smart meter and the signature of the electricity consumption data, and the homomorphic encryption is used for the electricity consumption. Data is encrypted and aggregated. In the end, the electricity consumption data is always in cipher text during the data transmission process. The forward-secure group signature scheme can avoid the harm caused by the interference of the private key, thus realizing the protection effect.

Index Terms: *Group Signature; Smart Grid; Homomorphic Encryption; Data Aggregation; Privacy Protection.*

Reference to this paper should be made as follows:

MING AI and HAIQING LIU, (2021), "PRIVACY-PRESERVING OF ELECTRICITY DATA BASED ON GROUP SIGNATURE AND HOMOMORPHIC ENCRYPTION" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 11-20, DOI 10.30696/IJEEA.IX.I.2021.11-20.

Biographical notes:

MING AI*, female, born in Xinyang, Henan in 1996. Now she is a master's student in the School of Control and Computer, North China Electric Power University. The main research direction is data privacy protection.*

HAIQING LIU*, currently associate professor at the School of Control and Computer Engineering, North China Electric Power University.*

1. INTRODUCTION

At present, the society's demand for electricity is gradually increasing, and the application of electricity is also increasing exponentially. And with the construction of the power grid, the reform of the power system is also deepening, and the smart grid has become a new direction for the development of the power grid. The smart grid collects user power consumption data and grid monitoring data in real time, and grasps the real-time operation status of the grid in time, so as to perform data analysis, equipment detection and safety protection of the entire system. However, in the smart grid system, users are installed with smart meters to collect power data, if the data is not encrypted in the process of data transmission, it is likely to lead to the leakage of power data. Once data leakage occurs, it is easy for attackers to analyze the electrical appliances used by users through their daily power consumption curve, and then analyze the number and behavior of people in the user's home, so as to master the user's life rules. If this information is used illegally, it will bring huge security risks to the power grid and users.

2. BACKGROUND

In order to deal with the leakage of user power data and identity privacy and the aggregation of encrypted data, researchers have proposed many related solutions. Castelluccia et al. [1] proposed in 2005 that a homomorphic encryption technology was used to complete the effective aggregation of encrypted data when the intermediate data processing node was not decrypted. Although the data processing center may encounter attacks, the attacker No privacy-sensitive information related to regional users will be obtained. In 2006, Westhoff et al. [3] pointed out that [1] Castelluccia uses different keys at each aggregation point and forcibly transmits a list of the identity information of encrypted nodes, which causes a large amount of communication overhead for each node. Therefore, Westhoff et al. [3] designed an early key distribution scheme to adapt to the end-to-end encryption in the reverse multicast flow sensor network. In 2009, Castelluccia et al. [2] extended the aggregation scheme [1], realized a provably safe and effective encryption aggregation scheme, and improved calculation and communication overhead. In 2011, in the scheme designed by Shi [4] and others, they adopted the method of homomorphic encryption in Castelluccia [1], and at the same time required users to divide their own information into blocks, and coordinated aggregation with other people's block information for protection User privacy. In 2016, Liu Xueyan and others proposed a model of multi-dimensional data aggregation and feedback mechanism for smart grid [6], which can carry out one-to-many communication. At the same time, they also designed a signature batch verification algorithm to reduce the computational overhead of smart grid signature verification. However, this model sets the user's access authority, and only users with corresponding attributes can obtain the corresponding control policies. In 2017, Zhou Hua and others proposed a power multi-dimensional data aggregation scheme under the smart grid multi-level gateway mode [7], which optimized the Paillier algorithm [8], and some studies also designed a fault-tolerant mechanism [9-11], but it also adds additional overhead. In 2017, Shen et al. proposed a privacy protection multi-dimensional data aggregation scheme that considers the residential area of multiple users [12]. According to Horner's law [13], the Paillier cryptosystem is used to hide polynomials. By aggregating data from multiple regions, the region-level polynomial is hidden in the final output.

This paper adopts the method of combining group signature and homomorphic encryption. First, the method of group signature is used to obtain the public and private key of the confirmed group member SM_i and the signature of the power data, and in order to avoid the harm caused by the leakage of the private key, proposed a key update scheme in which the private key of the group member

smart meter SM_i is naturally updated over time; then, through homomorphic encryption method, the power consumption data is always in the form of ciphertext in the transmission process from the smart meter to the regional center, and then from the regional center to the control center, which plays the role of privacy protection.

3. MODEL DESIGN

[3.1] SMART GRID SYSTEM MODEL

The smart grid is a complex system; the system model diagram presented in this paper is shown in Figure1, including the Control Center, Regional Center and the Residential Area of three parts. Control Center is responsible for the initialization of the system and the allocation of relevant system keys to the Regional Center and Residential Area, which have strong computing and storage capabilities and are assumed reliable at all time. In the system, the Residential Area contains w users, respectively, representing the $\{U_1,\ U_2,\ \ldots,\ U_w\}$, each user is equipped with a smart meter SM_i, which can safely transmit the user's electricity data to Regional Center at specific time intervals (e.g. 30min). SM_i encrypts electricity data and performs group signature operations of users. Regional Center has strong computing and storage capabilities, and it can determine whether SM_i is a legitimate member of the group, aggregate the user's electricity data, and send the aggregated data signature to Control Center, and if SM_i conflicts with the signature of power consumption data, Regional Center can track. the signature of the signer SM_i identity by querying the T_i and V_i in the tracking list. Control Center verifies the signature of Regional Center and then decrypts and analyzes the aggregated data. After understanding the power consumption characteristics of regional users, Control Center will dynamically adjust the ladder power price to achieve the grid's load balance, improve the reliability of grid operation, and maximize the interests of both parties.

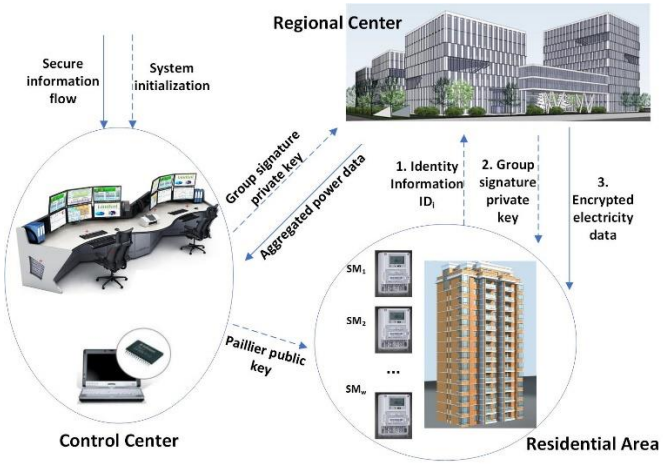

Figure:1. Smart grid system model

First, the control center generates various initial parameters of the system and assigns them to SM_i and the regional center. The control center selects a security parameter k, and the Paillier cryptographic system generates system initial parameters$(g,\ \mu,\ \lambda,\ p,\ q,\ n = p \cdot q)$, where $pk = (n,\ g),\ sk = (\lambda,\ \mu)$. Then the control center selects a parameter $r \in Z_n^*$ to generate the f-order cyclic group G, and selects a safe hash function $h(\cdot)$, where $h : \{0,\ 1\}^* \to Z_n^*$. Next, the control center selects a positive integer e, which satisfies $1 \le e \le \Phi(n)$ and $gcd\big(e,\ \Phi(n)\big) = 1$, and then calculates d, which satisfies $d \cdot e \equiv 1\ mod\ \Phi(n)$. Finally, the control center chooses a

random number$x_{CC} \in Z_n^*$ as the private key and calculates the public key $y_{CC} = r^{x_{CC}}$; the regional center chooses a random number $x_{RC} \in Z_n^*$ as its private key and calculates its public key $y_{RC} = r^{x_{RC}}$.

[3.2] PRIVACY-PRESERVING OF ELECTRICITY DATA BASED ON GROUP SIGNATURE AND HOMOMORPHIC ENCRYPYION

First, we must collect the user's electricity consumption data. In order to realize the collection and analysis of user electricity data every 30 minutes, each user's SM_i will obtain the corresponding elevator power $(T_1, T_2 \dots, T_K)$ types of data $(d_{i1}, d_{i2} \dots, d_{iK})$, and then After homomorphic encryption, it is sent to the regional center, and $d_{ij} < D$ (D is a constant). Suppose the power consumption collected by SM_i is p_i, located at [t, t+1], and then calculate the user's power consumption:

$$d_{ij} = dt_{i(t+1)} = p_i - T_t + Z \quad (2 \le j \le t);\ dt_{i(t+1)} = p_i - T_t,\ \ dt_{ij} = 0 \quad (t+1 \le j \le K) \qquad (1)$$

Finally, SM_i obtains $(d_{i1}, d_{i2} \dots, d_{iK})$, $d_{im} = dt_{im}(m = 1, \dots, t, t+2, \dots, K)$, $d_{i(t+1)} = dt_{i(t+1)} + Z$. Smart meter SM_i selects a random number $r_i \in Z_n^*$, and uses the following formula to encrypt power data:

$$c_i = \prod_{j=1}^{K} g_j^{d_{ij}} \cdot r_i^n \ mod\ n^2 \qquad (2)$$

Next, as shown in Figure 2, a group signature method is used to sign power data and so on. First select $G_1 = \langle P \rangle$, $G_2 = \langle Q \rangle$, $|G_1| = |G_2| = pp$, where $P \in E(F_q)$, $Q \in E(F_{q^\alpha})$. There is a corresponding non-degenerate bilinear mapping e: $G_1 \times G_2 \to G_T$, $H: \{0,\ 1\}^* \to G_1$, $H_1: \{0,\ 1\}^* \to Z_p$, isomorphic mapping $\psi: G_2 \to G_1$. For the regional center, take $gsk \in Z_P$ and calculate $gpk = gsk \cdot Q \in G_2$, where gsk is the private key of the group administrator and gpk is the public key, so the group public parameters are $\{pp,\ P,\ Q,\ G_1,\ G_2,\ e,\ H,\ \psi\}$, the group public key is gpk.

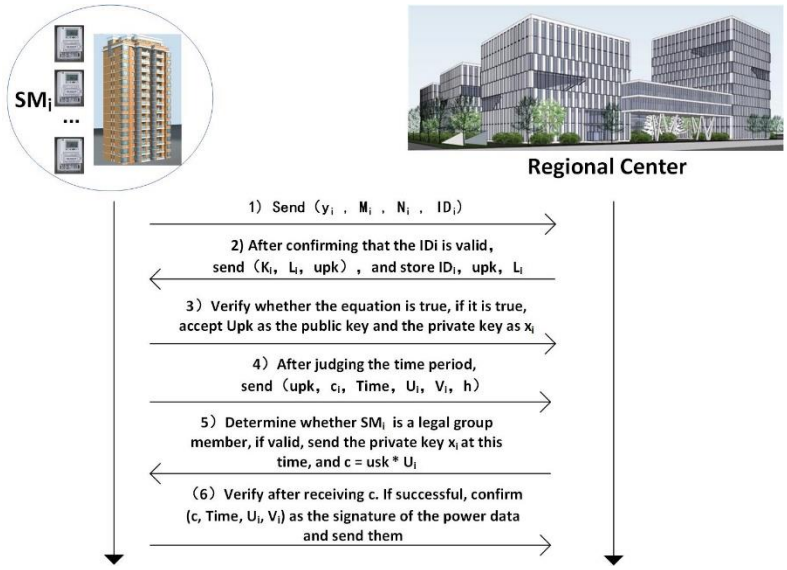

Figure:2. the process of group signature

For group member SM_i, SM_i takes x_i, $r_i \in Z_p$, calculates $y_i = x_i Q \in G_2$, $M_i = r_i H(ID_i)$, $N_i = r_i Q$, and then sets $(ID_i,\ y_i,\ M_i,\ N_i)$ are sent to the group manager regional center. After receiving $(ID_i,\ y_i,\ M_i,\ N_i)$, the regional center first verifies whether the equation $e(M_i,\ Q) = e(H(ID_i),\ N_i)$ is valid, and confirms the validity of ID_i Validity, namely: $e(M_i,\ Q) = e(r_i H(ID_i),\ Q) = e(H(ID_i),\ r_i Q) = e(H(ID_i),\ N_i)$, if it is true, calculate $K_i = e(gsk,\ y_i)$, $L_i = aQ$, $upk = y_i + L_i$, and then send $(K_i,\ L_i,\ upk)$ to SM_i, and store ID_i、upk、L_i, where ID_i、upk and timestamp Time constitute the public key list PKSL. After group member SM_i receives $(K_i,\ L_i,\ upk)$, verify the equation $K_i = e(x_i,\ gpk)$, $L_i = upk - y_i$ is all established, if so, accept upk as public key and private key is x_i.

The leakage of member keys in signatures often causes serious consequences, thus in this paper, the key is updated by using a one-way function. That is, (1) the entire effective time is divided into several periods 1,2 ,...,L; (2) in the first j phase, the key of the SM_i is $x_{i,j}$, the key of the group administrator is x_j , and the key in the first j+1 stage is $x_{i,j+1} = x_{i,j} + H_1(r_1 P \parallel j+1) - H_1(r_1 P \parallel j)$, $x_{j+1} = x_j + pH_1(aP \parallel j+1)$, the r_i is the number of randoms selected by SM_i when applying for a group, and $a \in Z_P$ is a fixed constant. After calculating the key of stage j+1, delete the key of stage j immediately.

Next, SM_i needs to sign the power data. First, SM_i judges the time period according to Time, set it as the jth time period, calculate $h_1 = H_1(r_1 P \parallel j)$, $h_2 = H_1(r_1 P \parallel 0)$, $h = h_1 - h_2$ h_2, $x_{i,j} = x_{i,0} + h$; secondly, take $b_i \in Z_P$, calculate $U_i = H(M \parallel Time \parallel Q \parallel b_i Q)$, $V_i = b_i - U_i x_{i,j}$, and then$(upk,\ M,\ Time,\ U_i,\ V_i,\ h)$ is sent to the regional center. The regional center first searches the PKSL table by the upk value to see if it is a valid value; if it is valid, according to the corresponding L_i value, verify the equation $U_i = H\left(M \parallel Time \parallel Q \parallel \left(V_i Q + U_i(upk - L_i + hQ)\right)\right)$ is established, so as to determine whether it is a legal member of the group. If the equation holds, the regional center stores U_i, and V_i constitutes a tracking list. At the same time, it judges the corresponding time period according to Time, and calculates the private key usk and $c = usk \cdot U_i$ at this time, and then sends c to the member SM_i, If it is not established, the signature request is rejected. After SM_i receives c, verify whether $e(c,\ Q) = e(U_i,\ gpk)$ is true. If it is true, then $(c,\ Time,\ U_i,\ V_i)$ is the value of SM_i for power data c_i signature. If it is not true, that is, the signature has conflicts and differences, the group administrator can track the identity of the signer of the signature by querying the corresponding U_i and V_i in the tracking list.

After the group signature has signed the encrypted power consumption data, SM_i send c_i and $(c,\ Time,\ U_i,\ V_i)$ to Regional Center. After Regional Center receives the signature, verify that $e(c,\ Q) = e(U_i,\ gpk)$ is true, and if so, accept the signature and calculate the aggregated data C according to the following formula:

$$C = \prod_{i=1}^{w} c_i \ mod\ n^2 \tag{3}$$

Then the Regional Center selects a random number $\alpha \in Z_n^*$, α meets $gcd(\alpha,\ n-1) = 1$ and then calculates its signature $Sig_{RC} = (Time_{RC},\ R_{RC},\ T_{RC})$ with the private key SK_{RC}, $RR_{RC} = r^t \ mod\ n$; $S_{RC} = h(C,\ Time,\ R_{RC})$; $T_{RC} = (S_{RC} - SK_{RC} \cdot R_{RC})\alpha^{-1}\ mod\ (n-1)$,

where the $Time_{RC}$ is the timestamp of the time. Finally, Regional Center sends the aggregate power consumption data C and the signature Sig_{RC} to Control Center.

After receiving the $\langle C,\ \ Sig_{RC}\rangle$ from Regional Center, the Control Center checks the validity of the signature according to the following formula:

$$PK_{RC}^{R_{RC}} \cdot R_{RC}^{T_{RC}} = r^{h(C,\ \ Time,\ \ R_{RC})}\ mod\ n \tag{4}$$

After the signature verification is completed, Control Center decrypts the aggregated power consumption data C, the decryption process is as follows:

$$\mathrm{C} = \prod_{\mathrm{i}=1}^{\mathrm{w}} \mathrm{c_i}\ \mathrm{mod}\ \mathrm{n}^2$$

$$= \prod_{i=1}^{w} g_1^{d_{i1}} \cdot g_2^{d_{i2}} \cdot \ldots \cdot g_L^{d_{iK}} \cdot r_i^n\ mod\ n^2$$

$$= g_1^{\sum_{i=1}^{w} d_{i1}} \cdot g_2^{\sum_{i=1}^{w} d_{i2}} \cdot \ldots \cdot g_L^{\sum_{i=1}^{w} d_{iK}} \cdot \prod_{i=1}^{w} r_i^n\ mod\ n^2$$

$$= g^{a_1 \sum_{i=1}^{w} d_{i1}} \cdot g^{a_2 \sum_{i=1}^{w} d_{i2}} \cdot \ldots \cdot g^{a_K \sum_{i=1}^{w} d_{iK}} \cdot \prod_{i=1}^{w} r_i^n\ mod\ n^2$$

$$= g^{a_1 \sum_{i=1}^{w} d_{i1} + a_2 \sum_{i=1}^{w} d_{i2} + \cdots + a_K \sum_{i=1}^{w} d_{iK}} \cdot \prod_{i=1}^{w} r_i^n\ mod\ n^2$$

$$= g^M \cdot R^n\ mod\ n^2$$

$$\left(M = a_1 \sum_{i=1}^{w} d_{i1} + a_2 \sum_{i=1}^{w} d_{i2} + \cdots + a_K \sum_{i=1}^{w} d_{iK}\ \ mod\ n,\ \ R = \prod_{i=1}^{w} r_i\right) \tag{5}$$

Control Center then restores plaintext $M = K\left(C^{\lambda\ mod\ n^2}\right)\mu\ mod\ n$ by using the master key $(\lambda,\ \ \mu)$.

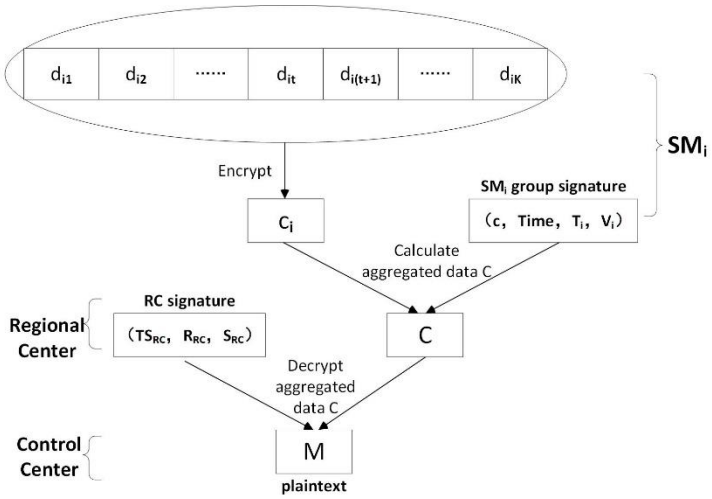

Figure:3. Encryption process combining homomorphic encryption and group signature

4 EXPERIMENTS

[4.1] PRIVACY ANALYSIS

First discuss the security of signatures and public and private keys.

(1) Correctness. In the process of joining the member, the identity of the member and the administrator is verified. SM_i received $(K_i,\ L_i,\ upk)$ sent by RC. First, verify $K_i = e(gsk,\ y_i) = e(x_i,\ gpk)$ through the properties of bilinear mapping, and then accept the sending $(K_i,\ L_i,\ upk)$ after it is proved that it is indeed the group manager. Secondly, confirm the validity of L_i through $L_i = upk - y_i$. Not only that, but group administrators and members must mutually verify their identities during the signing process. The signature is jointly generated by the group manager and the group member SM_i. On the one hand, after receiving the $(upk,\ M,\ Time,\ U_i,\ V_i,\ \mathrm{h})$ sent by SM_i, the regional center will first determine whether the upk is within the valid time according to the PKSL table, and then verify the membership. That is to verify whether $U_i = H\left(M \parallel Time \parallel Q \parallel \left(V_i \cdot Q + T_i(upk - L_i + \mathrm{hQ})\right)\right)$ is established, if it is established, it proves that the sender is indeed the group member SM_i and $(upk,\ M,\ Time,\ U_i,\ V_i,\ \mathrm{h})$. On the other hand, after receiving c from the regional center, the member of the group SM_i verifies whether $e(c,\ Q) = e(U_i,\ gpk)$ is true through the properties of bilinear mapping. If it is true, prove c It is indeed a valid value generated by the regional center.

(2) Anonymity. After receiving the signature $(c,\ Time,\ U_i,\ V_i)$, it is anonymous simply by verifying the validity of $e(c,\ Q) = e(U_i,\ gpk)$, and only the public key y_{RC} of Regional Center, does not involve information about SM_i .

(3) Anti-collusion. Because this paper adopts the form of verifying the identity of the group member SM_i, and then the signature form is produced by SM_i and Regional Center, the contact between the group member and the administrator in the signing process is strengthened, so the group members conspire to produce the signature that cannot be traced. Not only that, all members of the group, as well as the private keys of the administrator, are completely confidential and irrelevant to each other, so they are more resistant to collusion.

(4) The traceability of malicious users. Upon receipt of a signature $(c,\ Time,\ U_i,\ V_i)$, Regional Center can look up the tracking list directly according to U_i and V_i , thereby tracking the identity of the signer and successfully completing the tracking.

Then about the external attack, this paper uses the homomorphic encryption method so that the user's electricity data from SM_i to Regional Center, and then from Regional Center to Control Center transmission process, is always in ciphertext form, so can ensure the security of user data.

[4.2] TIME COMPLEXITY ANALYSIS

To test the efficiency of the proposed scheme, the experiment used the Irish Electric Power Company's home user smart meter data, which ranged from July 14, 2009, to December 31, 2010, at an interval of 30min. The experiment was implemented in MATLAB, and 2000 household users were randomly selected for the three-month smart meter data from January 15 to April 15, 2000, as experimental data.

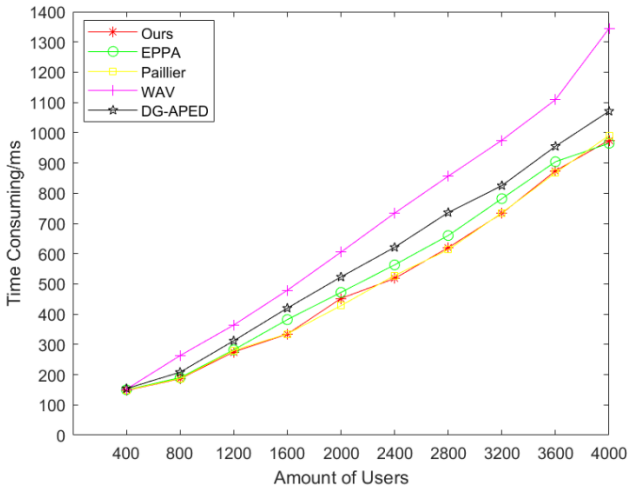

Figure:4. Efficiency curves of different methods

First, the efficiency of these methods is compared based on the number of different users. As shown in Figure 4, when the number of users is large, the scheme of this article does not have a clear advantage, but when the number of users is small, although the efficiency difference is not much, the scheme of this article is less time consumed.

In Figure 5, 1-r represents the privacy estimate of the scenario, and the lower the value indicates that the program's privacy-preserving is better. As shown in Figure 5, in most cases, the privacy estimate of this scenario is lower than the privacy estimate of the other scenarios, therefore, it can be considered that this paper has advantages in terms of privacy-preserving.

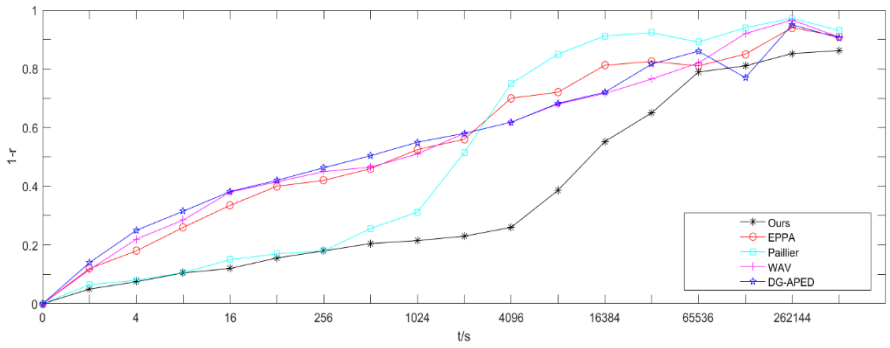

Figure:5. Privacy curve of different methods

[5] CONCLUSION

The wide application of the smart grid brings convenience and exposes users to the risk of identity and power data leakage. After studying the privacy-preserving of electricity data in the smart grid, this paper proposes a privacy-preserving scheme for electricity data based on improved group signature and homomorphic encryption, first using the method of improved group signature to obtain the user's smart meter signature and public and private key, and then using the method of homomorphic encryption to encrypt the data in the process of transmission of power data. In this scheme, the user hides the real electricity data, in the transmission process of the data, the data appears in the form of ciphertext, and the private key of the smart meter has changed over time, to

achieve the user's electricity data privacy-preserving. Finally, on the MATLAB platform, the simulation experiment of the intelligent meter data of the Irish electric power company is used to verify the validity of this paper in the protection of privacy.

REFERENCES

[1] Mykletun E . Efficient aggregation of encrypted data in wireless sensor networks[J]. IEEE, 2005.

[2] CASTELLUCCIA C, CHAN A C, MYLIETUN E, et.al. Efficient and Provably Secure Aggregation of Encrypted Data in Wireless Sensor Networks[J]. ACM Transactions on Sensor Networks (TOSN), 2009, 5(3):20.

[3] WESTHOFF D, GIRAO J, ACHARYA M. Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation[J]. Mobile Computing, IEEE Transactions on, 2006, 5(10):1417-1431.

[4] SHI J, ZHUANG R, LIU Y, et.al. Prisense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems [C]//INFOCOM, 2010 Proceedings IEEE. IEEE, 2010:1-9.

[5] LIN X, LURX, SHEN X M. MDPA: Multidimensional Privacy-Preserving Aggregation Scheme for Wireless Sensor Networks[J]. Wireless Comm. and Mobile Computing, 2010, 10(6):843-856.

[6] 刘雪艳，张强，李战明，等.面向智能电网通信系统的数据聚合和访问控制方法[J].电力系统自动化，2016, 40(14): 135-144.

[7] 周华，陈杰，张跃宇，等. 智能电网多级网络下多维数据聚合方案[J]. 密码学报，2017, 4(2): 114-132.

[8] Paillier P. Public-key cryptosystems based on composite degree residuosity classes[C]. International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg, 1999, 99:223-238.

[9] Dong M, Ota K, et al. Big data analysis-based secure cluster management for optimized control in software-defined networks [J]. IEEE Transactions on Network and Service Management, 2018, 15(1): 27-38.

[10] Zhang X, Tan Y, Xue Y et al. Cryptographic key protection against FROST for mobile devices[J]. Cluster Computing, 2017, 20(3): 2393-2402.

[11] Gao S, Ma X, Zhu J, et al. APRS:a privacy-preserving location-aware recommender system based on differentially private histogram [J]. Science China Information Sciences, 2017, 60(11):119103.

[12] Shen H, Zhang M, Shen J. Efficient privacy-preserving cube-data aggregation scheme for smart grids[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(6): 1369-1381.

[13] Anany L. Introduction to the design & analysis of algorithms[M].Boston: Addlson-Wesley .