Advances in Computer, Signals and Systems (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2023.070214
ISSN 2371-8838 Vol. 7 Num. 2

An aggregation scheme in a service-outsourced smart grid that protects against malicious data mining attacks

Baoyi Wang[a,*], Xian Du[b], Shaomin Zhang[c]

School of Control and Computer Engineering, North China Electric Power University, Baoding, Hebei, 071003, China
[a]wangbaoyi@126.com, [b]duxian_life_97@163.com, [c]zhangshaomin@126.com
[*]Corresponding author

Keywords: Service outsourcing, data aggregation, privacy protection, homomorphic encryption

Abstract: In price-based demand response, some sensitive computing is outsourced to service providers to reduce computational overhead for utilities, which poses some threats of privacy breaches to customers. Existing schemes protect user private data by data aggregation, but cannot resist the threat of data mining to user privacy leakage. An improved data aggregation homomorphic encryption scheme is proposed in this paper, which can effectively resist the threats to user privacy brought by data mining, and at the same time, the efficiency can be guaranteed by batch verification of encrypted messages. By analyzing the scheme's security and performance, demonstrated resistance to the threat of data mining attacks.

1. Introduction

Smart grids are considered as promising candidates for solving power imbalances and grid instability [1]. Demand response (DR) is one of the key features of smart grids and is possible to transfer users' electricity demand from peak to off peak periods [2]. Demand response plays an important role in the stabilization of power system supply and demand, and the calculation of electricity charges based on real-time prices [3] can respond to the cost of electricity and electricity consumption information in the short term, while better guiding customers' electricity consumption.

Real-time power data are characterized by volume, speed and variability [4], and the collection as well as the analysis of real-time power will be more complex. At the same time, some companies specializing in big data analysis can provide more efficient and accurate forecasting methods [5]. Service providers have introduced smart grid architectures that can remotely monitor and manage electricity consumption according to customer preferences [6, 7], outsourcing some complex processing to third parties and thus reducing the burden on the electric utility. However, the fine-grained power consumption data of smart meters reveal private information [8], and outsourcing of services will pose a threat to the privacy of the user's real-time power, and protecting the user's data privacy is the main research objective of this paper. In the scenario of service outsourcing, the most effective method for user data privacy preservation is secure data aggregation [5, 9], where user data is aggregated at the gateway and then sent to a outsourcer, where the data of multiple users is aggregated and the outsourcer can only obtain the aggregation results to protect user privacy. However, considering a more serious data mining attack, the attacker launches the attack intermittently, the attack disrupts communication between the user and the gateway [10], an attacker gets aggregated data at one moment by hacking into the database of the service provider, and then gets the aggregated data without this user at the next moment, then gets user’s sensitive information by calculation and analysis. However, there are still some limitations to protect the data privacy of this attack model. This paper aims to address the problem of resisting data mining attacks and obtaining accurate aggregation results efficiently in the smart grid scenario of service outsourcing.

2. Related technology

So far, although there are many responses to real-time pricing demands, there are very few studies that combine service outsourcing scenarios, as this scenario is completely different from previous studies. A scenario in this paper adds the role of a service provider who is able to use the customer's energy consumption for dynamic pricing forecasting without revealing any private information about the customer. Protecting the security and privacy of data transmission and privacy at this step as the data is transmitted from smart meters to gateways before it is sent to the service provider. In the literature [5], a homomorphic aggregation-based privacy-preserving approach is applied to a smart grid with real-time pricing demand response. The scheme also considers the scenario of service outsourcing, which not only protects user privacy but also allows outsourcing the prediction of dynamic electricity prices to service providers., but authentication is not completed between entities in this scheme and there may be malicious attacks by external attackers, while only aggregated encryption of data is considered in this paper's scheme, and the source legitimacy and reliability of data is not considered, while there may be internal attackers in this scheme, and in terms of data transmission, malicious The data mining attack may pose a threat to users' privacy and may leak users' privacy as well as cause damage to the power system.

3. Scheme design

The systems in this article include users, gateways (GW), outsourced service providers (SP), and utilities (PU). In this paper we assume a data mining attack model as in Figure 1, where the attacker blocks the communication between GW and the user at the moment T_t, and lets its communication return to normal at the moment T_{t+1}, and then launches the same attack again at the moment T_{t+2} as at the moment T_t. Suppose three users U_1, U_2, and U_3,d_1, d_2, and d_3 are the measurements of the three users at the moment T_{t-1}, and the attacker's goal is to infer the value of d_1. Suppose $d_1^0, d_2^0, d_3^0, d_1^1$、$d_2^1$、$d_3^1, d_1^2$、$d_2^2$、$d_3^2$ is the data from T_{t-1} to T_t, T_t to T_{t+1} and T_{t+1} to T_{t+2} time intervals respectively. At each moment the attacker compromises the SP database to obtain the aggregated data at each moment, and at moments T_t and T_{t+2} the attacker blocks the communication between GW and the user.D_0,D_1,D_2 and D_3 are the data of T_{t-1}, T_t, T_{t+1}, and T_{t+2} moments, respectively. The attacker can get $D_1 - D_0, D_2 - D_1, D_3 - D_2$ by using Eqs. (1) (2) (3) (4). Because the attacker's intrusion was very brief, Δd_2^0、Δd_2^1、Δd_2^2 and Δd_3^0、Δd_3^1、Δd_3^2 data values will be very close, The attacker deduces the value of $\Delta d_1^0 + \Delta d_1^1$, An attacker can infer the approximate value of d_1 from X in Eq. (5) to obtain sensitive information about the user.

$$D_0 = d_1 + d_2 + d_3 \tag{1}$$

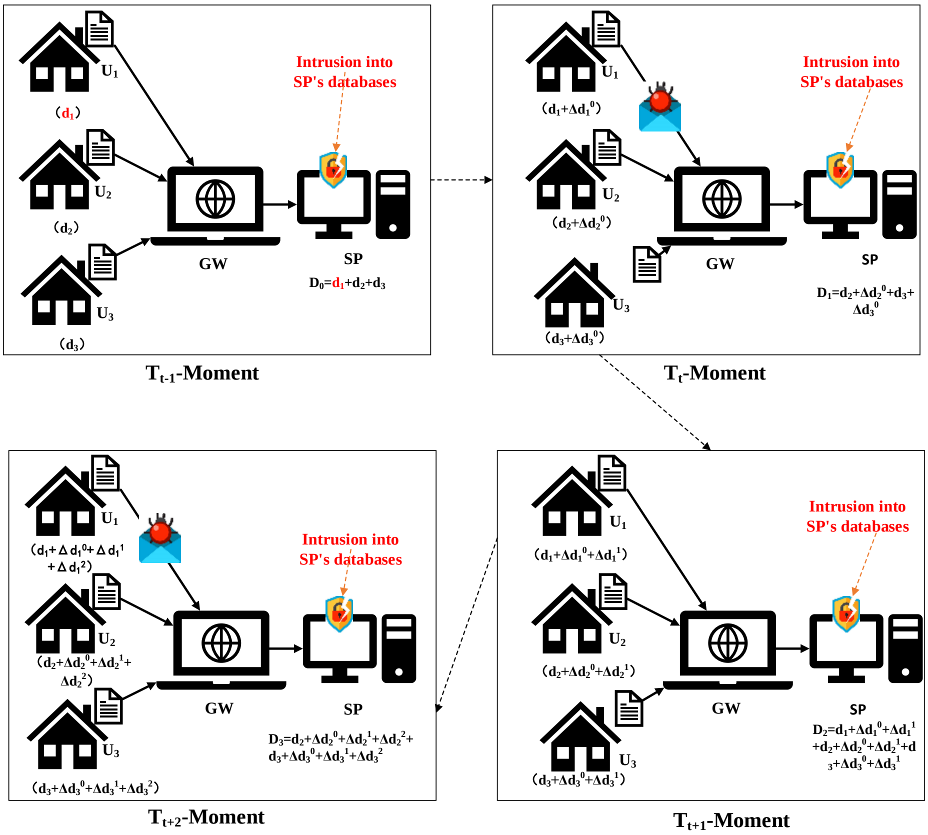
$$D_1 = d_2 + \Delta d_2^0 + d_3 + \Delta d_3^0 \tag{2}$$

$$D_2 = d_1 + \Delta d_1^0 + \Delta d_1^1 + d_2 + \Delta d_2^0 + \Delta d_2^1 + d_3 + \Delta d_3^0 + \Delta d_3^1 \tag{3}$$

$$D_3 = d_2 + \varDelta d_2^0 + \varDelta d_2^1 + \varDelta d_2^2 + d_3 + \varDelta d_3^0 + \varDelta d_3^1 + \varDelta d_3^2 \tag{4}$$

$$X = \varDelta d_1^0 + \varDelta d_1^1 \approx D_1 + D_2 - D_0 + D_3 \tag{5}$$

$$d_1 \approx \frac{(D_0+D_2-2D_1-X)-(2D_2-D_1-D_3-2X)}{4} \tag{6}$$



U_1
(d_1)
U_2
(d_2)
U_3
(d_3)
GW
SP
$D_0=d_1+d_2+d_3$
Intrusion into SP's databases
T_{t-1}-Moment
U_1
$(d_1+\Delta d_1^0)$
U_2
$(d_2+\Delta d_2^0)$
U_3
$(d_3+\Delta d_3^0)$
GW
SP
$D_1=d_2+\Delta d_2^0+d_3+\Delta d_3^0$
Intrusion into SP's databases
T_t-Moment
U_1
$(d_1+\Delta d_1^0+\Delta d_1^1+\Delta d_1^2)$
U_2
$(d_2+\Delta d_2^0+\Delta d_2^1+\Delta d_2^2)$
U_3
$(d_3+\Delta d_3^0+\Delta d_3^1+\Delta d_3^2)$
GW
SP
$D_3=d_2+\Delta d_2^0+\Delta d_2^1+\Delta d_2^2+d_3+\Delta d_3^0+\Delta d_3^1+\Delta d_3^2$
Intrusion into SP's databases
T_{t+2}-Moment
U_1
$(d_1+\Delta d_1^0+\Delta d_1^1)$
U_2
$(d_2+\Delta d_2^0+\Delta d_2^1)$
U_3
$(d_3+\Delta d_3^0+\Delta d_3^1)$
GW
SP
$D_2=d_1+\Delta d_1^0+\Delta d_1^1+d_2+\Delta d_2^0+\Delta d_2^1+d_3+\Delta d_3^0+\Delta d_3^1$
Intrusion into SP's databases
T_{t+1}-Moment

Figure 1: Example of a malicious data mining attack model

4. Scheme implementation

4.1. System initialization

The power company generates a private key, a public key and system parameters. The security parameter k is generated, SP generates the public key by the Paillier key $(N = pq, g)$and generate the corresponding private key(λ, β). The power company selects a large prime number p', while there exists a large prime number q' satisfying $q'|p' - 1$, and then selects two randomly generated numbers $g_1, g_2 \in \mathbb{Z}_{p`}^*$,It is also difficult to calculate the value of $\log_{g_1}^{g_2}$. Power company randomly selects n+1 keys $SK_i \in Z_{p`}^*$, and $\sum_{i=0}^{n} SK_i = 0 \ mod \ p$`(i=1, 2,……,n).Power company published parameters $\{N, g;\ g_1, g_2;\ q', p'\}$.

At this stage the three entities, user, GW and SP, complete the authentication and registration and the power company responds to the entity that registered the request. User U_i first selects $a_i, b_i \in Z_{q'}^*$ from a set of positive integers of prime numbers and calculates and sends the message$\{ID_i, X_i, Y_i, \alpha_i\}$ to GW. After receiving a message, verifies if $Y_i = g_2^{\alpha_i} \cdot X_i^{h_0(ID_i, X_i, Y_i)}$ holds at GW, if not GW terminates the registration process, otherwise GW announces the parameter$\{ID_i,\ X_i,\ Y_i,\ \alpha_i\}$, thus completing the authentication process for the registration of the meter. GW and SP similarly complete the authentication process based on the above process by selecting$a, b \in Z_{q`}^{*}$, from a set of positive integers of prime numbers and obtaining the announcement parameter $\{ID_G, X, Y, \alpha\}$.GW、SP、U$_i$($i \in \{1,2,\ldots\ldots,n\}$)request registration from SP and send registration message, which returns the private key(λ, β) to SP. PU sends authentication key SK$_0$ to GW. PU sends authentication key SK$_i$ to U$_i$.

$$X_i = {g_2}^{a_i} \ mod \ \ p` \tag{7}$$

$$X_i = {g_2}^{b_i} \ mod \ \ p` \tag{8}$$

$$\alpha_i = \{ID_i, X_i, Y_i\} \ mod \ \ q` \tag{9}$$

4.2. User Report Generation

U$_i$（i＝1,2,……, n） encrypts the meter data and generates the corresponding report, U$_i$ collects the meter data d$_i$,U$_i$ generates the timestamp t, Generate a random number r$_i$ in Z$_n$* and calculate:

$$\Gamma_i = g^{d_i} \cdot {r_i}^N \cdot {g_1}^{SK_i} \ mod \ \ N^2 \tag{10}$$

$$R_i = {g_2}^{SK_i} \ mod \ \ p \tag{11}$$

$$\Upsilon_i = h_1(ID_i, X_i, \Gamma_i, R_i, t) \tag{12}$$

$$E_i = SK_i + \Upsilon_i \cdot a_i \ mod \ \ q` \tag{13}$$

The user Ui sends$\{\Gamma_i, R_i, E_i, t\}$ to the GW.

4.3. Data aggregation and decryption phase

When GW receives $\{\Gamma_i, R_i, E_i, t_i\}$, verify the timestamp t_i and calculate $\Upsilon_i = h_1(ID_i, X_i, \Gamma_i, R_i, t_i)$ to verify that Eq. (14) holds. To reduce the cost, GW can use the small exponential test technique for batch verification, GW randomly selects a small integer s and a set of small numbers $\theta_1, \theta_2, \ldots\ldots, \theta_n \in [1, 2^s]$, and calculates whether Eq. (15) holds. According to Eqs. (7), (11), (13) can be derived to calculate Eqs. (16), (17).

$${g_2}^{E_i} = R_i \cdot {X_i}^{\Upsilon^i} \ (i \in \{1,2,\ldots\ldots,n\}) \tag{14}$$

$${g_2}^{\sum_{i=1}^{n} \theta_i \cdot E_i} = \prod_{i=1}^{n} ({R_i}^{\theta_i} \cdot {X_i}^{\theta_i \cdot \Upsilon_i}) \tag{15}$$

$$R_i \cdot {X_i}^{\theta_i} = {g_2}^{SK_i} \cdot ({g_2}^{a_i})^{\Upsilon_i} = {g_2}^{SK_i + a_i \cdot \Upsilon_i} = {g_2}^{E_i} \tag{16}$$

$$\prod_{i=1}^{n} ({R_i}^{\theta_i} \cdot {X_i}^{\Upsilon_i \cdot \theta_i}) = \prod_{i=1}^{n} ({g_2}^{SK_i \cdot \theta_i} \cdot ({g_2}^{a_i})^{\theta_i \cdot \Upsilon_i}) = {g_2}^{E_i \cdot \theta_i} \tag{17}$$

GW completes the aggregation operation of user-encrypted data according to the ciphertext calculation formula expressed in Equation (10),$\Gamma = \prod_{i=1}^{n} \Gamma_i$, similarly, GW generates informative reports on aggregated data that$R = \prod_{i=1}^{n} {g_2}^{SK_i}$, $R_G = {g_2}^R$,$\Upsilon_G = h_2(ID_G, X, \Gamma, R_G, T)$,$E_G = R + \Upsilon_G \cdot a$, GW sends $\{\Gamma, R_G, E_G, T\}$ to SP.

When SP receives$\{\Gamma, R_G, E_G, T\}$, Verify T and equation (18), if Eq. (18) does not hold, the system immediately terminates the operation, wait until the next time interval to receive data, calculate $\Upsilon_G = h_2(ID_G, X, \Gamma, R_G, T)$ and verify Eq. (19), If Eq. (19) does not hold, the system immediately terminates the operation, wait until the next time interval to receive data, otherwise, the received ciphertext is decrypted. SP decrypts the ciphertext based on the ciphertext derived from Eq. (20) and the Paillier decryption algorithm and the private key（λ, β）to obtain the aggregated plaintext $m = \sum_{i=1}^{n} d_i$.

$$R_G \cdot {g_2}^{SK_0} = 1 \tag{18}$$

$$ {g_2}^{E_G} = R_G \cdot X^{\Upsilon_G} \tag{19}$$

$$C = \Gamma \cdot {g_1}^{SK_0} \tag{20}$$

5. Security analysis

In the scheme of this paper, each entity completes authentication in the registration phase within the scope of mutual communication, which ensures that the messages are received from inside the system and can prevent internal attacks.SK_i is the blind factor of user U_i, SK_0 is stored in SP and, and $\sum_{i=1}^{n} SK_i = 0$,SK_i verifies that all users in the system are sending messages properly, and if some users fail to send messages, As in Figure 1, blocking T_t and T_{t+2} phase message transmission prevents the attacker from obtaining D_1 and D_3, while the attacker cannot derive d_1 based on D_0 and D_2. According to equations (10) and (11), we know that it is difficult for the attacker to decrypt the value of SK_i, Even if the attacker decrypts the value of SK_i, if equation (18) does not hold, the SP still cannot decrypt the aggregated information. From Eq. (19), if there is a legitimate user who has not sent a message, the SP will not get the exact ciphertext message and will not get the exact plaintext message, and the attacker cannot get the aggregated information by invading the SP database. Thus, it can resist external malicious data mining intrusion attacks to ensure the privacy of user measurement data.

6. Conclusions

In this paper, we improve the homomorphic encryption algorithm to counter malicious data mining attacks, and simultaneously resist data mining attacks and protect user privacy while accomplishing inter-entity authentication and preventing internal attackers. This paper is based on the existing model and applies it to the smart grid scenario of service outsourcing. This paper is feasible in terms of security and security efficiency.

References

[1] Heydarianforushani E, Golshan M, Shafiekhah M. A comprehensive linear model for demand response optimization problem [J]. Energy, 2020, 209.

[2] Xu B, Wang J, Guo M, et al. A Hybrid Demand Response Mechanism Based on Real-time Incentive and Real-time Pricing [J]. Energy, 2021, 231(1):120940.

[3] Chen W, Zhou A, Zhou P, et al. A Privacy-Preserving Online Learning Approach for Incentive-Based Demand Response in Smart Grid [J]. IEEE Systems Journal, 2019:4208-4218.

[4] Hu J, Vasilakos A V. Energy Big Data Analytics and Security: Challenges and Opportunities [J]. IEEE Transactions on Smart Grid, 2016, 7(5):2423-2436.

[5] Xue K, Yang Q, Li S, et al. PPSO: A Privacy-Preserving Service Outsourcing Scheme for Real-Time Pricing Demand Response in Smart Grid [J]. IEEE Internet of Things Journal, 2019, 6:2486-2496.

[6] He D, Chen C, Bu J, et al. Secure service provision in smart grid communications [J]. IEEE Communications Magazine, 2012, 50(8):53-61.

[7] Abdallah A, Shen X S. A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid[J]. IEEE Transactions on Smart Grid, 2016, 9(1): 396-405.

[8] Gope P, Sikdar B. An Efficient Data Aggregation Scheme for Privacy-Friendly Dynamic Pricing-based Billing and Demand-Response Management in Smart Grids [J]. IEEE Internet of Things Journal, 2018:3126-3135.

[9] Shen H, Zhang M, Shen J. Efficient Privacy-Preserving Cube-Data Aggregation Scheme for Smart Grids [J]. IEEE Transactions on Information Forensics & Security, 2017, 12(6):1369-1381.

[10] Hua S A, Yl A, Zhe X D, et al. An efficient aggregation scheme resisting on malicious data mining attacks for smart grid [J]. Information Sciences, 2020, 526:289-300.