文章编号：2095-4980(2021)03-0485-05

智能电网中面向隐私保护的数据聚合算法

左建业[1,2]

(1.新郑市技工学校，河南 新郑 451150；2.中原工学院 信息商务学院，河南 新郑 451150)

摘　要：随着智能电网(SG)的迅速发展，其安全和效率受到广泛关注。在SG中，居住区域内多个智能电表(SM)设备将感测数据传输至控制中心，使得用户数据需经过一些中间节点才到达控制中心。而通过挖掘用户相关数据(URD)，攻击者能够窃取用户的习惯和行为，因此，需要保护用户的隐私。为此，提出面向隐私保护的数据聚合(PPDA)算法。PPDA算法利用群位置隐私(GLP)掩饰由SM产生的数据，并利用Paillier加密系统对用户数据进行保护。此外，PPDA算法无需任何安全信道。仿真结果表明，提出的PPDA算法能够防御偷听攻击和勾结攻击，并控制了计算开销。

关键词：智能电网；隐私保护；数据聚合；群位置隐私；Paillier 加密

中图分类号：TP393　　**文献标志码：**A　　**doi：**10.11805/TKYDA2019368

A privacy-preserving data aggregation algorithm in Smart Grid networks

ZUO Jianye[1,2]

(1.Xinzheng Technical School, Xinzheng Henan 451150, China;
2.College of Business, Zhongyuan University of Technology, Xinzheng Henan 451150, China)

Abstract: In Smart Grids(SG), measured user data in Smart Meter(SM) in residential area are sent to Control Center(CC) passing through a few intermediate nodes. Hence, privacy preserving of user data is one of the biggest challenges in smart grid researches because the habits and behaviors of users can be disclosed by data mining the User-Related Data(URD). A Privacy-Preserving Data Aggregation(PPDA) algorithm is proposed. In PPDA algorithm, the Group Location Privacy(GLP) approach is taken advantage to mask fine-grained user data generated by SMs, and Paillier encryption system is utilized to preserve the user data privacy. The proposed protocol does not need any secure channel. The security and performance analysis shows that the proposed approach is secure against eavesdropping attack and collusion, and its computational overhead is acceptable compared with that of the previous works.

Keywords: Smart Grid; privacy-preserving; data aggregation; Group Location Privacy; Paillier encryption

相比传统电网，由电网和通信基础设施组成的智能电网，提高了数据传输的效率和可靠性[1]。由于智能电网直接与人民生活息息相关，安全、隐私和可靠性是部署智能电网时必须考虑的问题，也是智能电网领域的研究热点。物理和网络攻击[2]是威胁智能电网的两个主要攻击。物理攻击是指攻击者试图破坏电网设备和通信链路，进而达到攻击智能电网的目的；网络攻击是指攻击者通过偷听、数据破坏等手段攻击电网。此外，智能电网采用有线或无线通信方式，如 Wi-Fi,WI-Max,Zig-bee[3]，这些通信策略容易遭受攻击。因此，攻击者利用它们的弱点进行攻击，如偷听、勾结、拒绝服务(Denial of Service，DoS)、欺骗攻击[4]等。

用户相关数据(URD)[5]的隐私保护是智能电网最基本的要求。实质上，为了监测和控制用户能耗，节点产生数据，再聚合传输至控制中心。通过分析所接收的数据，控制中心能够控制和监测系统。

依据一个或多个中间节点可完成控制中心与节点间的通信。一个智能电网设备将其感测的数据传输至网关(Gateway，GW)；然后，GW 对数据进行预处理，再传输至下一接收点或控制中心。在数据传输的过程中如何保护 URD 的隐私是非常重要的。隐私保护就是通过技术手段使中间节点或控制中心无法掌握细粒度用户数据。原因在于：细粒度用户数据揭露了用户的习惯、行为特征等信息。

收稿日期：2019-09-26；**修回日期：**2020-02-17

作者简介：左建业(1974–)，男，硕士，高级讲师，主要研究方向为电气工程与自动化。email:cjswzyhh@163.com

即使智能电网系统在遭受攻击者的偷听或勾结攻击时，仍需保护 URD 的隐私。换而言之，URD 的隐私数据对攻击应具有一定鲁棒性。数据聚合是防止中间节点甚至控制中心发生对 URD 的细粒度用户数据进行攻击的有效方法。文献[4]研究表明，数据聚合有效提高了网络效率，降低了网络流量。

为此，本文提出面向隐私保护的数据聚合(PPDA)算法。PPDA 算法利用群位置隐私(GLP)方法掩饰由节点产生的细粒度用户数据，并利用 Paillier 加密系统对用户数据进行保护。仿真结果表明，提出的 PPDA 算法能够防御偷听攻击，控制了计算和通信开销。

1 系统模型

考虑如图 1 所示的网络模型，由智能电表(SM)、网关和控制中心构成。假定在居住区域内有 n 个 SMs，令 $\{\mathrm{sm}_1,\mathrm{sm}_2,\cdots,\mathrm{sm}_n\}$ 表示这 n 个 SMs，每个 SM 连通至智能电网，并收集用户数据。第 i 个 SM(sm_i)产生数据，再传输至 GW。通常，SM 在向 GW 传输数据前，先进行预处理，如加密操作。接收了所有 SM 数据后，网关进行聚合，然后再传输至控制中心。控制中心从中提取数据，并存储。

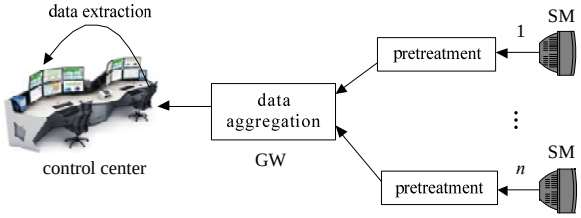

Fig.1 System model
图 1 系统模型

与文献[6–13]类似，本文考虑 3 类攻击：外部攻击、内部攻击和勾结攻击。在外部攻击中，攻击者窃听通信网络，并窃取通信数据；在内部攻击中，攻击者与系统内实体共谋，窃取用户的细粒数据，掌握用户的习惯和行为；在勾结攻击中，攻击者联合一个或多个实体，协同窃取用户的细粒数据，获取用户的隐私。

2 PPDA 算法

PPDA 算法利用 GLP 策略解决智能电网的数据聚合问题。GLP 策略结合了匿名否决网络(AV-net)[14]和 Paillier 加密[15]。文献[15]证实了 Paillier 加密的性能，Paillier 加密属概率公钥算法，其主要由密钥生成、加密和解密三部分组成。

Paillier 加密系统基于 $Z_{N^2}^*$ 上的模运算，其密钥空间为：

$$\kappa=\left\{(g,\lambda,N)\,|\,\lambda=lcm(p-1,q-1)\right\} \tag{1}$$

式中：p,q 为大素数；$N=pq$；lcm 为求最小公倍数函数；λ[15]表示控制中心的私钥；g 为随机整数。

选择随机数 $r\in Z_{N^2}^*$，Paillier 加密系统分别利用式(2)~(3)进行加密、解密：

$$E_{\mathrm{pk}}(m,r)=g^m r^N \bmod N^2 \tag{2}$$

$$D_{\mathrm{sk}}(c)=L\left(c^\lambda \bmod N^2\right)/L\left(g^\lambda \bmod N^2\right) \tag{3}$$

式中：m,c 分别表示明文、密文；函数 $L(x)=(x-1)/N$。

作为经典的加密系统，Paillier 在安全多方计算有重要应用。Paillier 是基于 n 次剩余问题判断的加密方案，其安全性体现于：对于整数 $N=pq$，攻击者要区分 $Z_{N^2}^*$ 中的元素与集合 $\left\{g^N\,|\,g\in Z_{N^2}^*\right\}$ 中的元素是困难的。

为此，PPDA 算法利用 AV-net 协议掩饰节点产生的数据；利用 Paillier 加密系统对聚合的数据进行加密。此外，PPDA 算法引用循环群 G，其阶数为 q。令 $Z_{N^2}^*$ 表示 Paillier 群，g 表示从群 G 内随机选择的产生因子，并且其满足计算 Diffie–Hellman(CDH)问题的不可逆性[15]。

利用式(4)从群 G 中计算 Paillier 产生因子 g_{s}：

$$g_{\mathrm{s}}=1+N \bmod N^2 \tag{4}$$

网络内所有的通信实体均认可群 G 和产生因子 g。此外，所有实体都知晓控制中心的公密钥(g_{s},N)。

2.1 数据产生阶段

令 m_i 表示 sm_i 在时刻 T 产生的数据。首先，多个 SMs 联合 AV-net，对数据进行泛化：

1) 每个 sm_i 从循环群 G 内产生一个随机数 x_i，并计算参数 g^{x_i} 值，同时将其 g^{x_i} 值向所有 SM 广播。

2) 随后，每个 SM 遵照 AV-net 协议，并依据式(5)计算 g^{y_i}：

$$g^{y_i}=\frac{\prod_{j=1}^{i-1}g^{x_i}}{\left(\prod_{j=i+1}^{n}g^{x_j}\right)} \tag{5}$$

3) 当获取了 g^{x_i} 和 g^{y_i} 值后，$\mathrm{SM}^{\mathrm{sm}_i}$ 就计算 $g^{x_iy_i}$，将其保存作为 AV-net 的 Mask，并利用 $g^{x_iy_i}$ 掩饰 SM 数据。

一旦获取了 Mask $g^{x_iy_i}$，SM 就对数据进行掩饰。具体过程如下:

1) 随机选择一个 SM(假定是 sm_α)。sm_α 先产生一个随机数 $r\in Z_{N^2}^*$，再利用式(3)对 sm_α 所产生的数据 m_α 进行掩饰，进而形成掩饰后的数据 C_α:

$$C_\alpha=g^{x_\alpha y_\alpha}g_{\mathrm{s}}^{m_\alpha}r^N \bmod N^2 \tag{6}$$

2) 除 sm_α 外，每个 $\mathrm{SM}^{\mathrm{sm}_i}$ 利用它自己的 AV-net 掩饰码对其数据进行掩饰，进而形成掩饰后的数据 C_i:

$$C_i=g^{x_iy_i}g_s^{m_i} \bmod N^2,\quad i=1,2,\cdots,n,i\neq\alpha \tag{7}$$

最后，所有 SM 将自己掩饰后的数据传输至网关。

2.2 数据聚合阶段

GW 收取了所有 SM 数据后，依据式(8)进行聚合:

$$C=C_\alpha\prod_{i\neq\alpha}C_i \bmod N^2=g^{\sum x_iy_i}g_{\mathrm{s}}^{\sum m_i}r^N \bmod N^2=g_{\mathrm{s}}^{\sum m_i}r^N \bmod N^2 \tag{8}$$

式中 $\sum x_iy_i=0$[16]。

将数据聚合后，GW 就将聚合后的数据 C 传输至控制中心。

2.3 提取数据阶段

收取所有 SM 发送的数据后，控制中心利用自己的私钥解密，进而提取数据，如式(9)所示:

$$Dec(C)=\sum m_i=\frac{L\left(C^\lambda \bmod N^2\right)}{L\left(g_{\mathrm{s}}^\lambda \bmod N^2\right)} \bmod N \tag{9}$$

PPDA 算法的执行流程如图 2 所示。

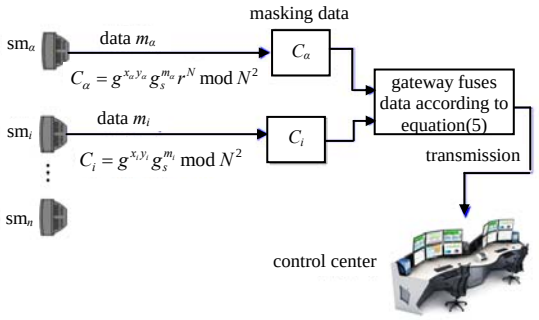

Fig.2 Execution process of PPDA algorithm
图 2 PPDA 算法的执行流程

3 性能分析

3.1 安全性能分析

3.1.1 外部攻击

PPDA 算法未采用任何安全信道，外部攻击者可偷听通信链路，并获取通信链路传输的数据。但这些数据都进行了处理，攻击者无法从这些数据中学习到 URD。

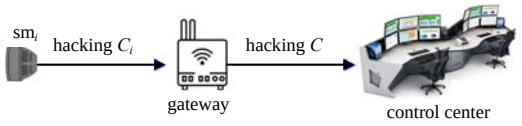

Fig.3 Diagram of external attack
图 3 外部攻击示意图

如图 3 所示，攻击者窃取了 SM 至网关的数据 C_i。若攻击者想从 C_i 内提取 m_i，需解决 CDH 问题[14]。而网关对传输控制中心传输数据 C 进行了加密。即使攻击者从网关至控制中心链路获取了加密数据 C，但其也无法窃取聚合数据。

3.1.2 内部攻击

内部攻击可能能够共谋网络实体。假定攻击者共谋了 sm_i，并获取 sm_i 的数据 m_i。但攻击者无法获取其他 SM 的细粒度数据。如果攻击者共谋了网关，但网关所获取的数据已是经过各 SM 进行掩饰的，即使从网关获取，也无法窃取数据。

3.1.3 勾结攻击

在勾结攻击中，恶意 SM 勾结一起，窃取诚实 SM 的数据。在全勾结攻击中，所有 SM 勾结一起共同谋取某一个 SM 的数据。在这种情况下，攻击者能够取消 AV-net 掩饰码，并窃取诚实 SM 的数据。

然而，在非全勾结(部分勾结)中，只有部分恶意 SM 联合谋取诚实 SM 数据。由于 AV-net 协议能够防御部分勾结[14]，攻击者无法通过部分勾结获取诚实 SM 数据。

3.2 计算开销

表 1 为各类操作的时间标识符。同时，选择文献[6–7]、[9–11]和[13]算法作为参照，并比较它们与 PPDA 算法的计算开销。

依据双线性对数，对阶为素数 p 的循环群 G 上的指数运算量进行分析，得出表 2。表 2 显示了 PPDA 算法内 SM、网关和控制中心所产生的计算开销。从表 2 可知，PPDA 算法的单个 SM 处所产生的开销为 $2T_{\text{exp}}+nT_{\text{mul}}$。而文献[6]、[10]和[13]算法的单个 SM 处所产生的开销分别为 $8T_{\text{EC-mul}}+T_{\text{pair}}+(n+2)T_{\text{EC-sum}}+T_{\text{Hash}}$，$3T_{\text{exp}}+3T_{\text{Hash}}+T_{\text{AES-E}}$，$3T_{\text{exp}}+T_{\text{Hash}}+4T_{\text{mul}}+T_{\text{add}}$。从这些数据可知，PPDA 算法的单个 SM 处的开销远低于文献[13]提出的算法。

表 1 标识符
Table1 Identifier

identifier	physical meaning
T_{mul}	modular multiplication time
T_{exp}	modular exponentiation time
T_{add}	modular summation time
T_{Hash}	Hash operation time
$T_{\text{AES-E}}$	execution time of AES encryption algorithm
$T_{\text{AES-D}}$	execution time of AES decryption algorithm
T_{log}	operation time for solving discrete logarithms
T_{pair}	operation time of bilinear pairing
$T_{\text{EC-mul}}$	operational time for elliptic curve multiplication
$T_{\text{EC-sum}}$	operational time for elliptic curve discovery
n	number of electricity meters in the residential area
d	number of servers in CC

表 2 计算开销
Table2 Computational overhead

scheme	calculation overhead at a single meter	calculation overhead at the gateway	calculation overhead at the control center
reference[6]	$8T_{\text{EC-mul}}+T_{\text{pair}}+(n+2)T_{\text{EC-sum}}+T_{\text{Hash}}$	$3T_{\text{EC-mul}}+T_{\text{pair}}+2T_{\text{EC-sum}}+T_{\text{Hash}}$	$3T_{\text{pair}}+T_{\text{log}}+nT_{\text{EC-sum}}$
reference[7]	$T_{\text{exp}}+5T_{\text{mul}}+2T_{\text{add}}+2T_{\text{Hash}}+T_{\text{AES-E}}$	$T_{\text{exp}}+(n+1)T_{\text{mul}}+T_{\text{AES-D}}+4T_{\text{Hash}}$	$3T_{\text{pair}}+T_{\text{log}}+nT_{\text{EC-sum}}$
reference[9]	$2T_{\text{exp}}+T_{\text{Hash}}$	nT_{mul}	$3T_{\text{pair}}+(2d+2)T_{\text{mul}}$
reference[10]	$3T_{\text{exp}}+3T_{\text{Hash}}+T_{\text{AES-E}}$	$nT_{\text{mul}}+T_{\text{AES-E}}+T_{\text{AES-D}}+2T_{\text{Hash}}$	$2T_{\text{exp}}+T_{\text{Hash}}+T_{\text{AES-D}}+T_{\text{mul}}+T_{\text{log}}$
reference[11]	$2T_{\text{exp}}$	$3T_{\text{exp}}+nT_{\text{mul}}$	$3T_{\text{exp}}+2T_{\text{mul}}+T_{\text{log}}$
reference[13]	$3T_{\text{exp}}+T_{\text{Hash}}+4T_{\text{mul}}+T_{\text{add}}$	nT_{mul}	$T_{\text{exp}}+T_{\text{Hash}}+(4n+3)T_{\text{mul}}+(5n+1)T_{\text{add}}$
PPDA	$2T_{\text{exp}}+nT_{\text{mul}}$	nT_{mul}	$3T_{\text{exp}}+2T_{\text{mul}}+T_{\text{log}}$

网关的计算开销最低，只有 nT_{mul}，文献[6–7,10]的计算开销达到 $3T_{\text{EC-mul}}+T_{\text{pair}}+2T_{\text{EC-sum}}+T_{\text{Hash}}$，$T_{\text{exp}}+(n+1)T_{\text{mul}}+T_{\text{AES-D}}+4T_{\text{Hash}}$，$T_{\text{AES-E}}+nT_{\text{mul}}+nT_{\text{AES-D}}+2T_{\text{Hash}}$。相比于同类算法，PPDA 算法控制了计算量。

此外，观察表 2 网关处的开销和控制中心处的开销可以发现，这两处的开销并非都与用户数 n 相关，但至少有一处与用户数 n 相关。如，文献[10]所提出算法的网关处计算开销与用户数 n 相关，但其控制中心处的开销没有与用户数 n 相关；文献[7]网关处的开销和控制中心两处的计算开销都与用户数 n 相关。

PPDA 算法的网关处开销为 nT_{mul}，其控制中心处的开销为 $3T_{\text{exp}}+2T_{\text{mul}}+T_{\text{log}}$。这说明控制中心处的开销与用户数无关，体现了 PPDA 算法对数据聚合的效果。

3.3 通信开销

表 3 为 PPDA 算法 SM 至 SM 间、SM 至网关间和网关至控制中心间通信开销的成本。

表 3 通信成本
Table3 Communication cost

scheme	communication cost from meter to meter	communication cost from meter to gateway	communication cost from meter to control center
reference[6]	$n[544(n-1)]$	1 600n	1 600
reference[7]	none	2 368n	2 208
reference[9]	none	2 048n	2 048
reference[10]	none	2 196n	2 176
reference[11]	none	2 048n	2 048
reference[13]	none	2 048n	2 048
PPDA	$n[2\,048(n-1)]$	2 048n	2 048

1) SM 至 SM 间的通信

PPDA 算法在数据产生阶段，每个 SM 需要计算 g^{x_i}，并向所有 SM 广播。若考虑 N=1024 bit，每个 SM 通信成本为 $2\,048\times(n-1)$ bit；文献[6]引用了 NIST-P192，数据长度为 192 bit。因此，文献[6]内每个 SM 需向其他 SM 发送 $544\times(n-1)$ bit。而文献[7,9–10,11,13]的 SM 至 SM 的通信开销为零。

2) SM 至网关的通信

PPDA 算法中每个 SM 需将 $C_i=g^{x_i y_i}g_s^{m_i} \bmod N^2$ 传输至网关。因此，每个 SM 需向网关发送 2 048n bit。类似地，文献[9,11,13]算法的 SM 至网关的通信开销与 PPDA 算法相似。在文献[7]中，每个 SM 需向网关发送约 2 368n bit。文献[10]算法中，每个 SM 使用 AES 加密数据，并传输至 GW。因此，它的通信成本为 2 196n bit。相比于文献[7,10]，PPDA 算法有效控制了 SM 至网关的通信成本。

3) 网关至控制中心的通信

PPDA 算法的网关需将聚合数据 C 传输至控制中心，通信成本为 2 048 bit，与文献[9,11,13]的通信成本一样，但低于文献[7]、文献[10]。

4　结论

针对智能电网中用户数据的隐私问题提出 PPDA 算法。PPDA 算法联合 AV-net 协议和 Paillier 加密系统，保护用户的隐私。具体而言，通过 AV-net 协议泛化用户数据，网关利用 Paillier 加密系统对数据进行加密。性能分析表明，提出的 PPDA 算法能够防御外部攻击、勾结攻击，并控制了计算开销。目前，本文只对 PPDA 算法进行了理论分析，并没有进行实例算法实验，这将是后期研究工作内容。

参考文献:

[1] 江荣. 智能电网安全与隐私保护相关问题研究[D]. 长沙:国防科学技术大学, 2014. (JIANG Rong. Research on security and privacy protection of smart grid[D]. Changsha,Hunan,China:National Defense University of Science and Technology, 2014.)

[2] MAHMUD R,VALLAKATI R,MUKHERJEE A,et al. A survey on smart grid metering infrastructures:threats and solutions[C]// IEEE International Conference on Electro/Information Technology(EIT). Dekalb,IL,USA:IEEE, 2015:386-391.

[3] PARIKH P P,KANABAR M G,SIDHU T S. Opportunities and challenges of wireless communication technologies for smart grid applications[C]// IEEE in Power and Energy Society General Meeting. Minneapolis,MN,USA:IEEE, 2010:1-7.

[4] 吴丹丹,吕鑫. 分布式结构下基于用户协作的匿名区域构建算法[J]. 计算机科学, 2019,46(4):158-163. (WU Dandan, LYU Xin. Location anonymous algorithm based on user collaboration under distributed structure[J]. Computer Science, 2019,46(4):158-163.)

[5] YAN Y,QIAN Y,SHARIF H,et al. A survey on cyber security for smart grid communications[J]. IEEE Communications Surveys & Tutorials, 2012,14(4):998-1010.

[6] LIU Y,GUO W,FAN C,et al. A practical privacy-preserving data aggregation(3PDA) scheme for smart grid[J]. IEEE Transactions on Industrial Informatics, 2019,15(3):1767-1774.

[7] LU R,HEUNG K,LASHKARI A H,et al. A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT[J]. IEEE Access, 2017(5):3302-3312.

[8] ABDALLAH A,SHEN X. A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid[J]. IEEE Transactions on Smart Grid, 2018,9(1):396-405.

[9] CHEN L,LU R,CAO Z. PDAFT:a privacy-preserving data aggregation scheme with fault tolerance for smart grid communications[J]. Peer-to-Peer Networking and Applications, 2015,8(6):1122-1132.

[10] BAO H,LU R. A lightweight data aggregation scheme achieving privacy preservation and data integrity with differential privacy and fault tolerance[J]. Peer-to-Peer Networking and Applications, 2017,10(6):106-121.

[11] BAO H,LU R. A new differentially private data aggregation with fault tolerance for smart grid communications[J]. IEEE Internet of Things Journal, 2015,2(3):248-258.

[12] DIMITRIOU T,AWA M K. Secure and scalable aggregation in the smart grid resilient against malicious entities[J]. Ad Hoc Networks, 2016(50):58-67.

[13] LI S,XUE K,YANG Q.et al. PPMA:privacy-preserving multisubset data aggregation in smart grid[J]. IEEE Transactions on Industrial Informatics, 2018,14(2):462-471.

[14] 付少锋,李龙海,焦晓鹏. 基于双线性对的匿名否决协议[J]. 计算机工程, 2011,37(22):105-106,109. (FU Shaofeng,LI Longhai,JIAO Xiaopeng. Anonymous veto protocol based on bilinear pairings[J]. Computer Engineering, 2011,37(22): 105-106,109.)

[15] 高胜,马建峰,姚青松,等. LBS 中面向协同位置隐私保护的群组最近邻查询[J]. 通信学报, 2015,36(3):67-73. (GAO Sheng,MA Jianfeng,YAO Qingsong,et al. Towards cooperation location privacy-preserving group nearest neighbor queries in LBS[J]. Journal on Communications, 2015,36(3):67-73.)

[16] CHEN L,LU R,CAO Z,et al. Multifunctional data aggregation in privacy-preserving smart grid communications[J]. Peer-to-peer Networking and Applications, 2015,6(5):777-792.