

电力客户数据隐私保护机制研究

邹云峰¹,徐超¹,倪巍伟²,唐刘远¹,沈涛²

(1. 国网江苏省电力有限公司 营销服务中心,南京 210000;2. 东南大学,南京 210000)

Research on data privacy protection mechanism for electric power customers

ZOU Yunfeng¹, XU Chao¹, NI Weiwei², TANG Liuyuan¹, SHEN Tao²

(1. Marketing Service Center, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210000, China;

2. Southeast University, Nanjing 210000, China)

摘要:随着互联网和大数据时代的到来,数据的价值越来越显著,在各个领域发挥的作用越来越重要,也催生了新的研究领域,即数据隐私安全。电网公司采集存储了海量电力客户数据,解决如何在使用过程中保障客户数据隐私安全的问题,变得尤为迫切。通过解析电力客户数据特征和应用场景,提出涵盖业务场景分类、隐私分级、保护策略与保护方法库在内的电力客户数据隐私保护体系建设方案,构建面向电力客户数据的主动隐私保护机制,实现对电力客户数据隐私保护策略和技术方法的主动推荐。实际案例展示了建立的电力客户数据隐私保护系统具有良好的实用性。

关键词:电力客户数据;隐私安全;隐私分级;保护策略;主动隐私保护

Abstract: With the advent of the Internet and the era of big data, the value of data is becoming more and more significant, playing a more and more important role in various fields. It also gives birth to a new research field, that is data privacy security. Power grid companies collect and store a large amount of power customer data. Solving the problem that how to protect the privacy and security of customer data in the process of using has become particularly urgent. By analyzing the characteristics and application scenarios of power customer data, a construction scheme of data privacy protection system for power customer is proposed, which includes business scenario classification, privacy classification, protection strategy and protection method library. An active privacy protection mechanism for power customer data is also built, and the active recommendation of data privacy protection strategies and technology methods for power customers are realized. The practical case shows that the data privacy protection system for power customers established has good practicability.

Key words: electric power customer data; privacy security; privacy classification; protection strategy; active privacy protection

0 引言

互联网和大数据技术的飞速发展,给人类生产生活带来了诸多便利。但在享受互联网和大数据红利的同时,数据隐私安全风险也日益受到社会和各行业的重视。针对数据隐私保护问题,近年来国内外研究者进行了积极的探索,在隐私保护数据挖掘^[1-3]、隐私保护数据发布^[4-5]、外包数据隐私保护^[6-7]等方面,已经开展了卓有成效的工作,提出了基于匿名、加密、差分、隐私信息检索在内的一系列保护技术。电网公司也已建立了电力客户敏感信息及数据安全管理体系,但当前对敏感数据的保护主要限于传统的访问权限控制技术,基于加密与数据脱敏的防护方法也在一些关键数据的保护中得到应用。

相较电力行业数据应用的现状和实施前景,目

前电力客户数据隐私保护机制存在以下不足:

(1) 不同类型数据保护目标不清晰,保护策略单一。缺少针对不同业务场景和隐私保护目标的差异化保护,无法结合数据应用特征对不同类型数据实施强度适当的隐私保护。

(2) 片面关注安全性,缺少对数据可用性的兼顾。目前电力数据多数基于不外泄原则进行保护,保护策略片面,导致数据的可用性与安全性矛盾突出。

(3) 片面依赖脱敏与加密,尤其缺少对大数据环境的适应性。现有的隐私保护方法片面依赖访问权限控制、加密和数据脱敏,在封闭系统内确实可以实现隐私安全保护,但在大数据环境下多源数据融合分析成为常态,孤立的数据脱敏或加密难以抵御多源数据级联攻击,存在隐私泄露隐患。

针对电力客户数据隐私安全保护存在的上述问题。本文解析电力客户数据特征和应用场景,提出涵盖业务场景分类、隐私分级、保护策略与保护方法库在内的电力客户数据隐私保护体系建设方案,构建主动隐私保护机制,实现针对不同业务场景和隐私保护目标的差异化精准保护。

收稿日期:2019-08-21;修回日期:2019-11-19

基金项目:国家自然科学基金项目(61772131)

This work is supported by National Natural Science Foundation of China (No.61772131)

1 现状分析

1.1 国内外数据隐私保护研究现状

自2000年开始数据隐私保护研究以来,国内外研究者在隐私保护数据挖掘、隐私保护数据发布、外包数据隐私保护等方面,提出了基于包括匿名、扰动、差分^[8-9]、隐私信息检索在内的一系列保护技术。

在隐私保护数据挖掘方面,研究者针对交易事务数据、关系表数据等代表性数据模式结构,采用向敏感属性添加特定噪声的方法,提出了一系列支持频繁项集、分类挖掘功能的隐私保护挖掘方法,并提出针对数据水平和垂直分割的隐私保护挖掘解决方案。在隐私保护数据共享发布方面,针对各类数据模式及应用特征,提出了一系列隐私模式和数据隐私保护技术方法。代表性数据隐私模型如表1所示。

表1 代表性数据隐私模型
Table 1 Representational data privacy models

隐私模型	隐藏原理	对象
k -匿名 ^[10]	用高粒度数值替换低粒度数值,实现泛化	关系表
l -diversity ^[11] , t -Closeness ^[12]	k -匿名基础上增加敏感属性值分布类约束	关系表
图差分隐私 ^[13]	添加噪声实现边的DP隐藏	图结构
驻留点 l -匿名 ^[14]	针对驻留点的泛化	轨迹
Semi-Edge Anonymity ^[15]	对节点间连接关系进行规范约束	社交网
k -Isomorphism ^[16]	通过修改边实现网络图中子图重复度	网络图
D-Clustering ^[17]	聚簇规模不小于 k 实现匿名	社交网

基于上述隐私模式,国内外研究者也提出了很多卓有成效的数据隐私保护方法,代表性方法如表2所示。

1.2 电力及其他行业数据隐私保护现状

各行业已认识到隐私问题的严重性,电网公司建立了数据安全管理体系,制订了数据脱敏标准规范,采取了一系列防范措施^[18],包括角色权限控制、数据传输加密、敏感数据脱敏等。苹果公司将差分隐私技术植入iOS10/macOSSierra中,确保用户位置分享、图片搜索和分类过程不泄露自身隐私;谷歌将局部差分隐私技术植入chrome浏览器,进而确保用户网页浏览访问日志中的隐私信息不受外来侵犯;阿里积极推出隐私守护App,保护手机上网用户的隐私不被泄露。国内移动、通信等行业也针对用户隐私泄露问题,制订了行业隐私保护法规和数据脱敏标准规范。

综上所述,目前在数据隐私保护研究方面,国内外研究者针对各类数据模式特征和应用特点,以及数据隐私泄露模式,提出了一系列数据隐私保护解决方案,取得了长足进展,但要实现数据隐私保护的行业应用落地,还需要针对行业应用固有模

表2 代表性数据隐藏方法

Table 2 Representational data hiding methods

隐藏原理	对象结构	描述	数据可用性
用高粒度数值替换低粒度数值,实现泛化	关系表	泛化模糊属性间映射关系	聚集查询
k -匿名基础上增加敏感属性值分布约束	关系表	添加对匿名组内敏感属性取值分布的约束	聚集查询
将数据表准标识符(Quasi_Identifier,QID)和敏感属性分到两张表泛化	关系表	泛化QID属性,分离QID与敏感属性	聚集查询
矩阵变换、保距隐藏	关系表	矩阵变换实现属性值修改	挖掘分析
利用多方安全计算实现分布子图隐藏	社交网络	加密实现子图结构隐藏	挖掘分析
规范节点联系实现隐藏	社交网络	泛化节点间连接关系实现节点和边敏感信息隐藏	挖掘分析
节点边的差分隐私(differential privacy, DP)隐藏	图结构	基于dk-graph生成模型,加入隐私参数实现边差分	节点度相关应用
用区域代替具体位置实现泛化	轨迹	泛化驻留点实现敏感信息保护和可用性提升	区间查询等

式,进行大量适配定制和落地工作。

目前虽然在通讯、互联网等行业,数据隐私保护已得到了初步应用,但隐私理论技术结合实际行业应用落地总体仍属于探索状态。近年来伴随着大数据分析研究的不断深化,电力客户数据在内部应用和对外数据共享的需求日益普遍,迫切需要系统的数据隐私保护机制。

2 电力客户数据隐私保护功能架构

数据隐私保护源于数据使用价值的发现与传递诉求,因此考虑从数据业务场景分类和数据应用模式解析角度研究电力客户数据隐私保护解决方案,构建数据隐私保护功能。电力客户数据隐私保护功能架构如图1所示。

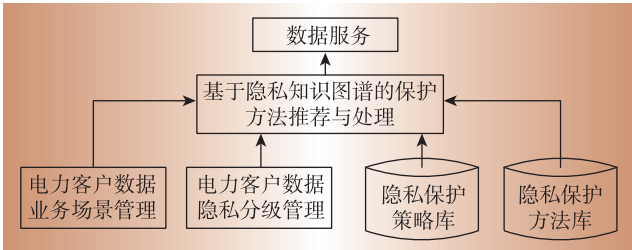


图1 电力客户数据隐私保护功能架构
Fig. 1 Data privacy protection functional architecture of electric power customers

架构由以下部分组成:数据业务场景管理、隐私分级、隐私保护策略库与隐私保护方法库以及保护

方法推荐功能模块。其中,数据业务场景管理实现对电力客户业务领域保护场景的创建管理,包括场景名称、类型、数据类型等;隐私分级实现对待保护数据的保护强度的宏观定级;隐私保护策略库是对隐私保护方法规范的形式化实现;保护方法库存储管理适用于电力客户主要业务场景下各类数据以及数据应用类型的代表性隐私保护算法。推荐功能模块通过内置推荐算法实现业务场景、数据、隐私保护强度、保护方法的自动适配和推荐,并调用方法库中的具体算法,实现对待保护数据的隐私保护处理。

3 电力客户数据隐私保护实现方法策略

3.1 电力客户数据业务场景分类

业务数据的安全性取决于数据所属业务领域的关键程度及其在系统内和系统外的共享需求特点。对电力客户数据进行分类梳理,解析各类电力客户数据应用场景和隐私泄露场景。

数据业务场景分类,以电力客户数据产生与应用环节为主线,兼顾数据类型特征进行分类。数据业务类型具体划分为电费、计量、业扩、线损、用检、稽查、服务七类,应用场景包括单户查询、聚集查询、统计报表、数据分析、数据发布等,泄露场景主要包括查询使用泄露、数据传输泄露、数据发布泄露等。

3.2 电力客户数据隐私分级

数据隐私安全不仅与数据所属业务环节有关,还与数据共享使用的应用类型有关。对于不同业务环节,数据的重要性和保护目标存在差异。即便是相同的数据,在用于不同应用时,数据隐私泄露风险也可能不同,相应的数据保护诉求也存在差异。

基于上述考虑,采用多级隐私分级思想,首先从业务类型重要性角度,对电力客户数据的隐私保护等级进行第1层划分,形成一级隐私、二级隐私、三级隐私。

在每一级隐私内部,继续从电力客户数据应用模式和泄露场景角度,对各类电力客户数据隐私保护等级进行2层划分。结合电力行业大数据应用场景,将数据共享应用模式归纳为聚集查询、数据挖掘、数据共享发布几类。从而实现业务类型、数据应用模式、泄露场景与数据的多级关联,以便基于隐私分级,实现数据隐私的多粒度保护。第2层隐私保护等级也划分为3级,1级最高,3级最低。相应的,对相同数据集,所属业务类型相同,第一层隐私分级相同,例如都为一级,但由于该数据集用于不同类型应用,需要提供不同强度的隐私保护,若用于查询,敏感数据隐私泄露风险相对小,从应用类型角度考虑,只需提供中等强度保护,则保护强度设置为1.2级;若用于挖掘分析,可能造成敏感数据隐私泄露的风险更大,需要提供1.1级保护强度。

电力客户数据隐私分级如图2所示。

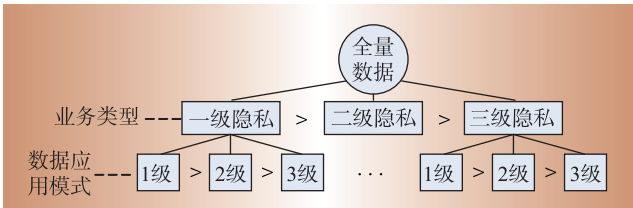


图2 电力客户数据隐私分级

Fig. 2 Data privacy classification of electric power customers

电力客户数据隐私分级为电力客户数据隐私保护的实施提供了保护标准参考,隐私分级也为后续隐私保护策略的制定和保护方法的定制筛选提供宏观指导。

3.3 隐私保护策略与方法库构建

电力客户数据隐私保护策略实现电力客户数据隐私分级与具体隐私保护方法的关联适配,描述各级别隐私建议采用的保护技术方法。结合电力客户数据应用场景,制订了如表3所列的保护策略。

表3 隐私保护策略
Table 3 Privacy protection strategy

类型	策略
1	访问控制策略
2	查询防护策略
3	数据挖掘保护策略
4	数据发布保护策略
5	数据传输安全策略

例如数据发布隐私保护策略中,规定对隐私分级较高的客户数据,建议采用基于差分隐私技术的保护方法,对保护需求较低的客户数据,可采用匿名保护技术等。对低隐私等级的数据采用高强度保护技术,浪费了计算资源,并可能造成数据使用方式和价值方面的损失;若采用达不到隐私级别的保护技术,又不可避免的会带来数据隐私的泄露。

结合电力客户数据主要应用类型,电力客户数据隐私保护方法库中对基于各种保护技术的隐私保护方法进行分类管理。具体分为面向查询的隐私保护方法库、面向数据挖掘的隐私保护方法库和面向数据发布的隐私保护方法库。各类方法库根据保护技术所提供保护效果和数据可用性维持类型进行标签分类。例如,在面向数据挖掘的隐私保护方法库中,分为隐私保护特征提取类方法、隐私保护关联分析类方法、隐私保护聚类方法等共计20余种保护算法,采用的技术涵盖k-匿名、差分、数据扰动等,能够提供不同层级的隐私保护强度。

3.4 保护策略推荐

要实现目标电力客户数据的隐私保护,需要根据待保护数据所属场景,判定数据所属隐私等级,

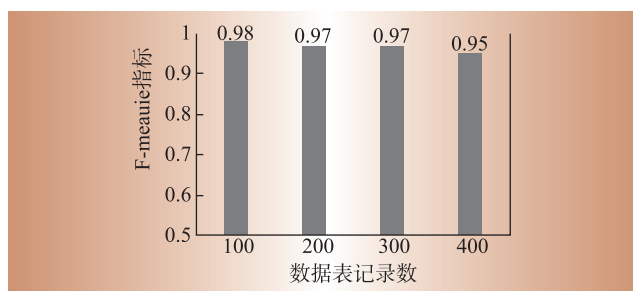


图6 聚类可用性维持效果

Fig. 6 Cluster utility maintenance effect

5 结束语

本文对电力客户数据隐私保护技术和体系进行研究,通过电力客户数据隐私分级,构建多级隐私保护策略和方法库,实现对电力客户数据隐私保护策略和技术方法的主动推荐,确保电网公司在电力营销大数据应用中数据隐私的安全管理,支撑兼顾数据隐私安全的电力大数据应用,提升电网公司市场竞争力。

参考文献:

- [1] 李学明,刘志军,秦东霞. 隐私保护数据挖掘[J]. 计算机应用研究, 2008, 25(12): 3 550-3 555.
LI Xueming, LIU Zhijun, QIN Dongxia. Privacy preserving data mining[J]. Application on Research of Computers, 2008, 25(12): 3 550-3 555.
- [2] 吴响,俞啸,王换换. 面向数据挖掘的匿名化隐私数据发布系统设计[J]. 电子技术应用, 2016, 42(11): 62-65.
WU Xiang, YU Xiao, WANG Huanhuan. Design of anonymous privacy data publishing system for data mining[J]. Application of Electronic Technology, 2016, 42(11): 62-65.
- [3] 左国才. 基于大数据的分布式隐私保护聚类挖掘算法研究[J]. 智能计算机与应用, 2018, 8(6): 57-60.
ZUO Guocai. Research on distributed privacy protection clustering mining algorithm based on big data[J]. Intelligent Computer and Applications, 2018, 8(6): 57-60.
- [4] 倪巍伟,陈耿,崇志宏,等. 面向聚类的数据隐藏发布研究[J]. 计算机研究与发展, 2012, 49(5): 1 095-1 104.
NI Weiwei, CHEN Geng, CHONG Zhihong, et al. Privacy-preserving data publication for clustering[J]. Journal of Computer Research and Development, 2012, 49(5): 1 095-1 104.
- [5] 岳思,吴伟明,谷勇浩. 数据发布中k-匿名隐私保护技术研究[J]. 软件, 2017, 38(11): 12-17.
YUE Si, WU Weiming, GU Yonghao. Research on k-anonymous privacy protection technology in the data release[J]. Computer Engineering & Software, 2017, 38(11): 12-17.
- [6] 任静涵. 外包数据挖掘隐私保护算法研究和改进[D]. 上海:上海交通大学, 2009.
REN Jinghan. Privacy preserving algorithm research and improvement of outsourcing data mining[D]. Shanghai: Shanghai Jiao Tong University, 2009.
- [7] 李晔. 具有隐私保护的外包数据分类方法研究[D]. 哈尔滨:哈尔滨工业大学, 2018.
LI Ye. Research on outsourcing privacy-preserving data classification method[D]. Harbin: Harbin Institute of Technology, 2018.
- [8] 王迪,袁健,申泽宇. 大数据环境中交互式查询差分隐私保护模型[J]. 计算机应用研究, 2019(6): 1-8.
WANG Di, YUAN Jian, SHEN Zeyu. Interactive queries differential privacy protection model in big data environment[J]. Application Research of Computers, 2019(6): 1-8.
- [9] 陈晓宇,韩斌,黄树成. 基于差分隐私的数据匿名化隐私保护方法[J]. 计算机技术与发展, 2018, 28(7): 99-102, 107.
CHEN Xiaoyu, HAN Bin, HUANG Shucheng. An anonymized data privacy protection method based on differential privacy[J]. Computer Technology and Development, 2018, 28(7): 99-102, 107.
- [10] SWEENEY L. K-anonymity: a model for protecting privacy[J]. Int'l Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5): 557-570.
- [11] MACHANAVAJJHALA, GEHRKE A, KIFER J, et al. L-diversity: privacy beyond k-anonymity[C]. ICDE, 2006: 24.
- [12] LI N, LI T, VENKATASUBRAMANIAN S. T-closeness: privacy beyond k-anonymity and l-diversity[C]. ICDE, 2007: 106-115.
- [13] YUE W, XIN T. Preserving differential privacy in degree-correlation based graph generation[J]. Transactions on Data Privacy, 2013, 6(2): 127.
- [14] ZHENG H, XIAO F, HAI B, et al. You can walk alone: trajectory privacy - preserving through significant stays protection[J]. DASFAA(1), 2012: 351-366.
- [15] MING X, LEI C, HONG M. Graph publication when the protection algorithm is available[J]. Data & Knowledge Engineering, 2013, 88: 267-285.
- [16] CHENG J, FU A W C, LIU J. K-isomorphism: privacy-reserving network publication against structural attacks[J]. In: SIGMOD, 2010.
- [17] HAY M, MIKLAU G, JENSEN D, et al. Resisting structural re-identification in anonymized social networks[J]. The VLDB Journal, 2010, 19(6): 797-823.
- [18] 杨帆,张琴,刘捷,等. 国网四川电力数据资产安全管理体系构建研究[J]. 电力信息与通信技术, 2018, 16(1): 90-95.
YANG Fan, ZHANG Qin, LIU Jie, et al. Research on the construction of SEPC data assets security management system[J]. Electric Power Information and Communication Technology, 2018, 16(1): 90-95.
- [19] BENJAMIN C, FUNG M, KE W, et al. Privacy-preserving data publishing for cluster analysis[J]. Data & Knowledge Engineering, 2009, 68(6): 552-575.
- [20] ESTER M, KRIEGER H, SANDER J, et al. Proceedings of the second international conference on knowledge discovery and data mining (kdd - 96). 1996 [C]. Menlo, Park CA: AAAI Press, 1996.

作者简介:

邹云峰(1977),男,江西丰城人,硕士,高级工程师,主要研究方向为电力营销及信息化;
徐超(1989),男,山东济南人,硕士,工程师,主要研究方向为电力大数据分析、电力信息安全;
倪巍伟(1979),男,江苏淮安人,博士,教授,主要研究方向为数据挖掘、数据隐私安全保护;
唐刘远(1990),男,江苏连云港人,本科,助理工程师,主要研究方向为电力大数据分析、电力信息安全;
沈涛(1994),男,浙江湖州人,硕士,主要研究方向为数据隐私安全保护。

(责任编辑 徐文红 赵雨昕)