Smart Grid 智能电网, 2014, 4, 123-128
Published Online June 2014 in Hans. http://www.hanspub.org/journal/sg
http://dx.doi.org/10.12677/sg.2014.43019

Hans汉斯

Privacy Protection of User List in Smart Grid Based on Cloud Computing

Mengyin Ren, Qiqi Mao, Ting Ma, Hong Wen

National Key Laboratory of Science and Technology on Communications, UESTC, Chengdu
Email: 474029586@qq.com

Received: Apr. 22nd, 2014; revised: May 20th, 2014; accepted: Jun. 2nd, 2014

Copyright © 2014 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

With the rapid development of intelligent power system, the requirements of storage and computing massive data are increasing. It becomes inevitable to combine power system with cloud computing. However, access to cloud computing system also brings data security issues. This paper is focused on the protection of user privacy in smart meter system based on data combination privacy and trusted third party. We introduce data chunk storage and chunk relationship confusion to protect user privacy. We also propose a chunk information list system for inserting and searching data.

Keywords

Power System Cloud Computing, Privacy Protection, Data Chunk

基于云计算的智能电表用户表单隐私保护

任梦吟，毛琪琦，马　婷，文　红

电子科技大学通信抗干扰国家级重点实验室，成都
Email: 474029586@qq.com

收稿日期：2014年4月22日；修回日期：2014年5月20日；录用日期：2014年6月2日

摘　要

智能电力系统的快速发展，对于海量数据的存储及计算的要求越来越高，使其与云计算的结合成为必然。但是接入云计算系统不可避免的带来了数据安全问题。本文针对智能电表系统的用户信息隐私保护，基于数据组合隐私机制，引入可信第三方，运用数据分块存储及分块关系混淆技术保护智能电表系统的用户隐私。并提出分块信息表单系统，在分块存储环境下实现数据插入与数据查询。

关键词

电力云计算，隐私保护，数据分块

1. 引言

随着电力系统的迅速发展，智能电力设备的大量建成，海量数据的实时采集并涌入电网数据中心，给数据中心造成巨大的存储计算负担，使得智能电网[1]与云计算[2]的结合成为必然。云计算将数据存储在网络中的虚拟设备上，并且能够分布式、并行的利用大量的处理器资源计算海量的数据。同时，云计算还有高可靠性、支持定制与扩张、费用低廉等优点，这些对于电力系统都很具有吸引力。

但是，云计算的引入，不可避免的带来了隐私安全问题[3]，用户的信息若不加保护，会轻易的被挖掘利用。涉及隐私的信息泄露可能会对用户带来巨大的困扰、造成严重的后果。本文针对智能电表系统的用户信息隐私保护，基于数据组合隐私机制，引入可信第三方，运用数据分块存储及分块关系混淆技术保护智能电表系统的用户隐私。并提出分块信息表单系统，在分块存储环境下实现数据插入与数据查询。

2. 隐私保护

隐私是个体、机构不愿意被他人知晓的信息。通常可以将隐私理解为敏感数据，如病患所患疾病、银行卡号等。智能电表系统的用户隐私主要表现为用户的个人信息，如身份证号、联系方式，及用户的用电习惯等。

云计算环境中，服务提供商不一定是完全可信的，即服务提供商可能会尝试窥探并获取用户的隐私信息。同时，外部的一些居心叵测的人，通过数据挖掘[4]技术也可以获取到用户隐私。这些盗窃者可能会将隐私信息出售给广告公司以谋取私利，并对用户的生活造成困扰。

目前，对于隐私保护方面的研究主要集中在数据失真、数据加密和数据匿名[5]技术上。数据失真技术通过增加噪声、进行凝聚等操作给敏感数据添加扰动，使攻击者不能获取到真实数据。数据失真技术的计算开销较小，但十分依赖原始数据，不适合大范围的应用。数据加密即使用加密算法对敏感数据进行加密处理，数据加密可以有效的保护用户隐私，但加解密时较大的计算与通信开销使其在应用于基于云的大量数据处理时效率较低。数据匿名技术有选择的发布敏感数据和可能披露敏感数据的信息，但将敏感信息泄露概率限制在一个可容忍的范围内。数据匿名技术的计算开销不大，但伴有一定程度的数据缺损与隐私泄露。

针对智能电表用户表单信息中用户私人信息是主要需要保护的信息，而用电量等为次保护信息，因此其表现为部分隐私保护的需求，本文以电网中一份用户信息表单为例，在云存储环境下，参照文献[6]提出的数据组合隐私技术，提出了保护用户隐私信息的方法。

3. 数据组合隐私

3.1. 相关定义

数据组合隐私(Data Combination Privacy, DCP)，指个体不希望暴露的一系列数据属性的组合，即这些属性的数据值组合可以确定某个特定的个体。对于一个给定的数据存储 D，一个组合隐私泄露的概率为 $P_{DCP} = P(DCP/D)$，规定隐私保护阈值 $T(0 \le T \le 1)$，当任意组合数据隐私满足 $P_{DCP} \le T$ 时，我们认为该数据存储 D 保护了用户数据组合隐私。

隐私约束(Privacy Constraint, PC)，PC{AS(Attribute Set, AS $\subseteq$ A = {A_1,A_2,⋯,A_n})，PP(Privacy Policy)}，指对于用户数据属性集合 A 的一个子集 AS，采用相应的隐私策略 PP。隐私策略有两种：相容与不相容。其中，相容指 AS 集合中所有属性同时出现不会导致数据组合隐私泄露，不相容则会导致隐私泄露。

数据分块(Data Chunk, DC)，指基于隐私约束的规则，将原始数据属性集合划分成一些不相交的子集 $AS\left(\bigcup AS_i = A, AS_i \ne \varnothing\right)$，按照这些子集将原始数据表单分块，并通过一些手段，混淆不同分块之间数据的关联关系。

数据分块云存储(Data Chunk Cloud Storage, DCCS)，指将数据分块后，存储于云端不同位置，为了保护数据分块的相关信息，引入可信第三方。由可信第三方实现数据的分块，并决定存储的位置。

3.2. 分块信息表单系统

为了实现数据正确重构，我们添加了数据分块 ID(DCID)与数据记录 ID(Data Record ID, DRID)。

在可信第三方中建立分块信息表单(Chunk Information List, CIL)，对于每一个数据分块，给其设置一个独一无二的 DCID，以 DCID 为表单名存储到指定位置，并在 CIL 中添加记录，分块信息表单属性如表 1 所示。

分块 DC_i 中的每一条记录，通过特定的映射算法，根据该条记录在原始表单中的 $ID_{original}$ 计算出 $DRID_{DC_i}\left(DRID_{DC_i} = F_{DC_i}\left(ID_{original}\right)\right)$。映射算法应该满足：

1) 原始表单中的每一条记录，在不同的分块中，DRID 都不同；

2) 可以通过反映射算法，由 $DRID_{DC_i}$ 计算出 $ID_{original}$。

映射算法相关的信息也存储于可信第三方中。同时，为了减少通过一一对应的方法猜测记录关联关系的可能性，将每一个分块中的记录的顺序打乱，从而实现分块记录关系的混淆。数据组合隐私保护主要的时间开销在于映射算法。

4. 用户信息表单隐私保护应用

表 2 为电网系统中一份智能电表用户信息表单属性，下面就其数据保护进行分析。

4.1. 数据分块

1) 隐私约束

表单中，账号、姓名、联系方式、地址和电表 IP 属于显式标识符[7](能唯一标识单一个体的属性)，

Table 1. Chunk information list attributes
表 1. 分块信息表单属性

DCID	原始表单	包含属性	存储位置

电表读数(1 h)和读数时间属于敏感属性(包含隐私数据的属性)。在制定隐私约束规则时，设定显示标示符与其他属性均为不相容策略，敏感属性之间也为不相容策略。

2) 数据分块

按照制定的隐私约束，运用文献[6]中的算法，得到的数据分块结果为：

$DCID_1${账号}

$DCID_2${姓名}

$DCID_3${联系方式}

$DCID_4${地址}

$DCID_5${电表 IP}

$DCID_6${性别，电表读数(1 h)，上一月消费，账户余额}

$DCID_7${读数时间}

3) 分块信息表单

对于智能电表用户信息表单，存储于可信第三方中的分块信息表单 CIL 如表 3 所示。

4.2. 数据插入

随着用户的增加，需要对表单实时的更新，插入新用户的相关信息。对于分块存储的表单，必须通过可信第三方的协助实现记录的插入。

设智能电表用户信息表单为 L，需在其中插入记录 R，数据插入算法步骤如下：

1) 在可信第三方中搜索分块信息表单 CIL 中满足原始表单为 L 的所有记录，记录集合为 RS；

2) 根据原始表单 L 的分块情况对应的将记录 R 进行分块；

3) 查询 L 的第一块分块 DC_1，获取分块存储位置 P_1，在 P_1 中查找 $DCID_1$ 表单，获取表单现有的记录数 N，则新插入的数据的 $ID_{original}$ 为 N + 1；

4) 对于每一个分块 DC_i，根据可信第三方中的映射算法与记录 R 的 $ID_{original}$，计算记录 R 的分块 DC_{Ri} 对应的 $DRID_i$，并将记录{$DRID_i$, DC_{Ri}}插入到表单 $DCID_i$ 的随机位置。

Table 2. Smart Grid user information list attributes
表 2. 智能电表用户信息表单属性

账号	姓名	性别	联系方式	地址	电表 IP	电表读数(1 h)/度	读数时间	上一月消费	账户余额

Table 3. Chunk information list of Smart Grid user information list
表 3. 智能电网用户信息的分块信息表单

DCID	原始表单	包含属性	存储位置
$DCID_1$	智能电表用户信息	账号	P_1
$DCID_2$	智能电表用户信息	姓名	P_2
$DCID_3$	智能电表用户信息	联系方式	P_3
$DCID_4$	智能电表用户信息	地址	P_4
$DCID_5$	智能电表用户信息	电表 IP	P_5
$DCID_6$	智能电表用户信息	性别、电表读数(1 h)、上一月消费、账户余额	P_6
$DCID_7$	智能电表用户信息	读数时间	P_7

4.3. 数据查询

用户具有通过电网系统查询权限范围内账号相关数据的权力，若查询的数据表单被分块存储，则必须运用重构算法。

设智能电表用户信息表单为 L，用户需要在 L 中通过值为 x 的 A_i 属性查询其 A_j 属性对应值。算法步骤为：

1) 在可信第三方中搜索分块信息表单 *CIL* 中满足原始表单为 *L* 的所有记录，记录集合为 *RS*；

2) 从搜索到的记录中查找出包含 A_i 属性的记录(只有一条)，得到对应的 $DCID_i$ 与存储位置 P_i；

3) 从获取到的存储位置 P_i 中查找 $DCID_i$ 表单，搜索 A_i 属性为 x 的记录，得到记录的 $DRID_i$；

4) 结合 $DCID_i$ 和可信第三方中对于表单 L 的反映射算法，计算出对应于原始表单的 $ID_{original}$；

5) 从记录集合 *RS* 中搜索包含 A_j 属性的记录(只有一条)，得到 $DCID_j$ 与存储位置 P_j；根据映射算法计算出记录在 $DCID_j$ 中 $ID_{original}$ 对应的 $DRID_j$；

6) 从存储位置 P_j 中查找 $DCID_j$ 表单，搜索 *DRID* 为 $DRID_j$，属性为 A_j 的记录对应的值 *y*，返回 *y* 值。

以上步骤完成了单条记录的查询。若要重构全部表单，则可以以表单对应的任意一个分块 DC_k 中所有记录为查询的关键字，并在第五步时，搜索除 DC_k 以外的所有分块。

4.4. 小结

数据组合隐私保护的时间开销主要由映射算法及连接开销决定，其中连接开销与分块数有关。所以在表单属性相容策略较少，分块数等于或接近于属性数时，数据组合隐私保护将不再具有优势。

在智能电网用户信息表单中，主要保护的是用户的个人信息及个人信息与电量数据之间关系，而不一定需要隐藏电量数据，所以在使用数据组合隐私保护时，相容策略比较多，分块数小于属性数，与传统的数据加密相比，具有较低的时间开销。而对于需要隐藏全部信息的表单，如教育系统中学生成绩表单，分块数等于属性数，则不适合使用数据组合隐私保护方法。

5. 总结

本文主要以电网系统下的一份用户表单为例，运用数据组合隐私保护技术，通过数据分块存储与分块关系混淆，可以有效的防范数据挖掘攻击及云服务商监守自盗，保护了用户隐私安全；提出分块信息表单系统，通过可信第三方可以实时的进行数据插入和数据查询操作。

在云存储、大数据环境下，映射算法的复杂度是影响效率的主要因素，这也是后续研究的主要方向。

基金项目

自然科学基金项目(编号：61271172)、高等学校博士学科点专项科研基金(编号：20120185110030 和 20130185130002)、四川省国际合作研究项目(编号：2013HH0005)和国家教育部回国人员科研启动基金联合资助。

参考文献 (References)

[1] 余贻鑫, 栾文鹏 (2009) 智能电网. *电网与清洁能源*, **1**, 7-11.

[2] 王利赛, 杨明玉, 孙月琴, 王栋, 张永浩 (2011) 电力云研究综述. *Electric Power IT*, **5**, 20-23.

[3] 张逢喆 (2010) 公共云计算环境下用户数据的隐私性与安全性保护. 博士论文, 复旦大学, 上海.

[4] 王惠中, 彭安群 (2011) 数据挖掘研究现状及发展趋势. *工矿自动化*, **2**, 29-32.

[5] Sweeney, L. (2002) k-Anonymity: A Model for Protecting Privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, **10**, 557-570.

[6] 张坤, 李庆忠, 史玉良 (2011) 面向 SaaS 应用的数据组合隐私保护机制研究. *计算机学报*, **11**, 2044-2054.

[7] 周水庚, 李丰, 陶宇飞, 肖小奎 (2009) 面向数据库应用的隐私保护研究综述. *计算机学报*, **5**, 847-861.