云计算隐私保护研究

刘景森　郭永伟　陈　阳
（河南大学计算机与信息工程学院　河南开封　475000）

【摘　要】云计算作为一种新的IT应用模式，数据安全和隐私保护对云计算的安全与普及至关重要，也是用户关注的一个焦点。K-匿名算法是目前数据发布环境下隐私保护的主要技术之一。文章分析了当前云计算在隐私保护方面存在的风险，提出了一种K-匿名算法在云计算中的应用方法，使得用户在向云服务提供商请求服务时可以隐藏个人敏感信息。

【关键词】云计算；隐私保护；K-匿名算法

Privacy-Preserving Schemes for Cloud Computing

Liu Jing-sen　Guo Yong-wei　Chen Yang
(College of Computer and Information Engineering,Henan University　HenanKaifeng 475000)

【Abstract】Data security and privacy protection are vital for the security and popularization of cloud computing which is known as a new application pattern of IT, and they have become the focus of users' concern. K-anonymity algorithm is one of the main technologies for privacy protection in the current data dissemination environment. After analyzing the current issue of privacy protection existing in the clouding computing, this paper presents a new approach applying k-anonymity algorithm to cloud computing. Through this way, users are allowed to hide their personal sensitive information when requesting services from cloud service providers.

【Keywords】cloud computing；privacy protection；k-anonymity

1　引言

继分布式计算、网格计算之后，云计算是新一代信息技术和产业的重要发展方向，是新的网络应用模式。其核心思想是资源租用、应用托管、服务外包。在云计算环境下，IT行业的按需服务真正得到了体现。云计算具有非常广泛的应用前景，然而在云计算提供方便易用和低成本特性的同时，也带来了新的危机，安全方面的问题首当其冲。随着云计算的不断发展，用户日益增多，用户数据的安全、用户隐私信息的保护问题、数据的异地存储以及云计算平台自身的稳定性等诸多安全和云计算监管方面的问题，直接关系到了云计算业务被用户的接受程度，进而成为了影响云计算业务拓展的最重要因素。因此，要让不同用户大规模应用云计算技术与平台，放心地将自己的数据信息交付于云服务提供商管理，就必须全面地分析并着手解决云计算所面临的各种安全问题。

2　云计算数据和用户隐私风险分析

云计算的基本原理是，通过使计算分布在大的分布式计算机上，而非本地计算机或远程服务器中，企业数据中心的运行将更与互联网相似这使得企业能够将资源切换到需要的应用上，根据需求访问计算机和存储系统。

云计算应用环境具有多租户、动态性、虚拟化等特点，其数据和用户隐私安全问题与传统信息安全问题有着不同的特点。在传统网络模式下，可以使用防火墙、网闸技术、数据交换网技术等边界网络防护手段。由于云计算环境的边界不确定性，这些传统的边界网络防护手段在云计算下的应用受到了极大的限制。对于用户，最重要的安全目标就是数据安全与隐私保护，防止云服务提供商恶意泄露或出卖用户隐私信息，或者对用户数据

进行搜集和分析,挖掘出用户隐私数据。对于云服务提供商而言，数据的安全性不仅是对用户的可靠保证,也是自身业务的最基本需求。

云计算数据和用户隐私风险具体有几种。

(1)数据隔离风险。用户对与云计算存储的不可控性,不同的用户数据之间应该做到有效的隔离和加密保护,防止用户的数据遭到非法访问,威胁用户数据的安全。

(2)数据完整性风险。由于恶意攻击和病毒感染,用户数据面临的风险主要体现为存储数据的完整性、传输数据的完整性等遭到破坏带来的风险。

(3)数据残留风险。不完全的数据删除,硬盘设备的维修和报废都有可能导致机密数据的泄露所带来的风险。

(4)用户隐私风险。用户身份相关的关键数据,如用户口令、姓名、银行卡号等泄露所带来的风险。

3 云计算数据和用户隐私保护方法

3.1 传统隐私保护方法

数据挖掘时代,为人们提供了十分强大的发掘信息的功能,同时也给个人的隐私带来了巨大的问题。目前,解决该问题的主要方法有几种。

(1)匿名保护。某些机构为了保护个人的隐私信息,通常都是对姓名、个人社会保障代码等能够清楚标示个人信息的显示标示符进行加密或者是删除，但是这并不足以阻止攻击者获取信息，攻击者通过所发布数据中的其他信息,例如民族、性别、生日、邮编等,和其他渠道获取的信息进行交叉对比,最终能够挖掘出用户的隐私信息。

(2)在对数据进行清理时,对原始数据进行扰乱,扭曲,随机化之后再进行挖掘。这种方法虽然能够尽可能的保持结果里面的整体特性,但是这中方法的代价就是数据的完整性、真实性遭到破坏。

(3)基于密码学的隐私保护技术,主要有安全多方计算、盲签名等。该方法需要很多的资源。

为了解决以上三种方法的不足,1998 年 Samarati P 和 Sweeney L 提出了 K- 匿名算法。该算法要求公布后的数据信息中必须存在一定数量的不可区分的个体信息,使攻击者无法判别出隐私信息具体属于哪一个个体信息,从而防止了个人隐私的泄露。

3.2 K- 匿名算法

显示标识符指能够清楚标识用户隐私信息的属性,如用户身份证号、个人社会保障号、姓名等,在用户数据表中删除显示标识符可以在一定程度上达到保护个人隐私信息的目的。但事实上,原始数据中通常还包含邮编、性别、生日、地址等非显示标识符,攻击者可将非显示标识符和其他渠道获得的信息进行链接对比,识别出主体身份信息。例如,某些患者不想其他人知道他的病情,但是攻击者可以从表 1 中根据非显示标示符来获取数据信息,再通过其他部门或者商业机构中获取其他的数据信息,进行链接对比,从而能够得到用户的隐私信息。

表 1 医疗信息表

姓名	民族	出生日期	性别	地址	疾病
	汉	1964-05-12	男	湖北省襄樊市	流感
	汉	1964-02-15	男	湖北省武汉市	癌症
	汉	1968-04-06	男	河南省郑州市	肺结核
	汉	1968-10-11	男	河南省开封市	癌症
	回	1965-12-05	女	甘肃省兰州市	癌症
	回	1965-05-06	女	甘肃省天水市	胃炎

表 2 医疗信息表的一个 2-匿名化表

民族	出生日期	性别	地址	疾病
汉	1964	男	湖北省	流感
汉	1964	男	湖北省	癌症
汉	1968	男	河南省	肺结核
汉	1968	男	河南省	癌症
回	1965	女	甘肃省	癌症
回	1965	女	甘肃省	胃炎

K- 匿名算法的要求是,给定的数据表 $T(A_1,A_2,\cdots,A_n)$,其中准标示符(QI,Quasi Identifier)为 $QI(A_i,\cdots,A_j)$,$A_i,\cdots,A_j \in A_1,A_2,\cdots,A_n$,在数据表 T 中的任何一个有序元素组值在 T[QI] 中重复 k 次以上。

表 2 满足 K- 匿名要求,准标识符为{民族,出生日期,性别,地址},K=2。表 2 中与准标示符任意一个属性相关联的值至少出现两次,K- 匿名算法主要是通过泛化和隐匿技术实现,能够保持数据的真实性,发布精度比较低的数据信息,使数据表中的每一条记录都至少与该数据表中其他的 K-1 条记录具有完全相同的准标示符属性值,从而降低链接攻击所导致的隐私信息泄露。

3.3 K- 匿名算法在云中的应用

假设终端用户想向云服务提供商(SP)请求服务,SP 为了确保服务发给正确的用户,就会要用户的一些属性。

如果用户发出了他的信息(例如地址、性别、电话号码等)给 SP,这就会变得很危险。因为如果其他企业(不是 SP)获得这些信息,他就可以识别用户。因此,为了避免用户的个人信息被泄露出来,我们使用基于 K- 匿名的方法,在用户的信息发给 SP 之前先对用户数据进行处理。

使用加密技术处理用户的数据并不是有效的,因为如果数据没有修复的话,服务提供者可能无法访问他们。一旦数据被解密,用户就处在风险中。然而,如果我们使用匿名处理这些数据并把这些匿名数据发送给云服务提供商,云服务商可以立即使用这些数据而不用再修复它们,因此这对于在云计算中保护个人隐私也变得更加的灵活和安全。

当客户端想发送数据给 SP 数据集时,SP 首先发送匿名披露集(名为 DSets)到客户端,我们假设披露是匿名和集加密的,因此,它不传达任何独特的属性标识符,以避免唯一标识,客户端获取披露集来验证他的数据是否能满足 K- 匿名算法对应于 SP 的表。

用户发送自己的匿名数据给 SP 之前,用户应该检查添加到 SP 表中的数据是否仍能满足 k- 匿名,如果把该元组添加到表仍然满足 k- 匿名,然后用户就可以添加自己的数据。

4 结束语

随着云计算技术的发展,越来越多的企业愿意应用该技术,如亚马逊、谷歌、惠普等。随着存储个人敏感信息的数据库和软件遍布在互联网的各个地方,如果我们可以解决非常现实的隐私和安全问题,我们能够充分享有云计算的好处。个人或企业的隐私泄露问题不可避免地在云服务的发布和数据共享出现了。所以我们强调在云计算中保护个人隐私的重要性,并提出了一种新的方法,旨在避免用户向云服务提供商请求服务时个人的敏感信息的泄露。

参考文献

[1] Samarati P. Protecting respondents identities in microdata release [J]. Knowledge and Data Engineering, IEEE Transactions on, 2001, 13(6): 1010-1027.

[2] Samarati P, Sweeney L. Generalizing data to provide anonymity when disclosing information[C]. PODS. 1998, 98: 188.

[3] Bayando R, Agrawal R. Data privacy through optimal K-anonymity [C]. Pmc of the 21 st International Conference on Data Engineering. 2005: 217-228.

[4] Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1984 (6): 721-741.

[5] Chen Y, Ma J, Feng Q, et al. Nonlocal prior Bayesian tomographic reconstruction [J]. Journal of Mathematical Imaging and Vision, 2008, 30(2): 133-146.

[6] Machanavajjhala A, Kifer D, Gehrke J, et al. l-diversity: Privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1): 3.

[7] 杨晓春, 刘向宇, 王斌, 等. 支持多约束的 K-匿名化方法[J]. 软件学报, 2006, 17(5): 1222-1231.

[8] Jiang W, Clifton C. Privacy-preserving distributed k-anonymity [J]. Data and Applications Security XIX. Springer Berlin Heidelberg, 2005: 166-177.

[9] 岑婷婷, 韩建民, 王基一, 等. 隐私保护中 K-匿名模型的综述[J]. 计算机工程与应用, 2008, 44(4): 130-134.

[10] 周水庚, 李丰, 陶宇飞, 等.面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009, 32(5): 847-861.

[11] Chao C M, Chen P Z, Sun C H. Privacy-Preserving Clustering of Data Streams [J]. Tamkang Journal of Science and Engineering, 2010, 13(3): 349-358.

[12] 杨晓春, 王雅哲, 王斌, 等.数据发布中面向多敏感属性的隐私保护方法[J].计算机学报, 2008, 31(4): 574-587.

[13] 韩建民, 岑婷婷, 虞慧群. 数据表 k-匿名化的微聚集算法研究[J]. 电子学报, 2008, 36(10): 2021-2029.

[14] Truta T M, Vinay B. Privacy Protection: p-Sensitive k-Anonymity Property[C]. ICDE Workshops. 2006: 94.

[15] Vaidya J, Clifton C. Privacy-preserving data mining: Why, how, and when[J]. IEEE Security & Privacy, 2004, 2(6): 19-27.

[16] 张文科, 刘桂芬.云计算数据安全和隐私保护研究[J].信息安全与通信保密, 2013 (11): 38-40.

[17] Jensen M, Schwenk J, Gruschka N, et al. On technical security issues in cloud computing[C]. Cloud Computing, 2009. CLOUD'09. IEEE International Conference on. IEEE, 2009: 109-116.

作者简介:

刘景森(1968-),男,河南唐河人,西北工业大学,博士,河南大学,教授;发表论文 30 余篇,主持完成省部级以上科研项目 12 项,参加过多个重大国防项目和横向项目的研究与开发;主要研究方向和关注领域:网络信息安全、电子商务。

郭永伟(1988-),男,河南安阳人,河南大学硕士研究生;主要研究方向和关注领域:云计算隐私保护。

陈阳(1988-),女,河南南阳人,河南大学硕士研究生;主要研究方向和关注领域:云计算存储。