云环境下用户隐私属性及其分类研究*

季一木[a,b]，匡子卓[b]，康家邦[b]，孙延鹏[b]，潘俏羽[b]
（南京邮电大学 a. 计算机学院，b. 物联网学院，南京 210023）

摘　要：为了保护用户隐私，在云计算环境对用户隐私的影响基础上，通过对用户隐私的分类研究，设计出一种云环境下用户隐私分类方法。对其中隐私的属性进行研究，利用 XML 进行属性描述。最后通过设计的电信行业 CRM 用户隐私保护系统来对云环境下用户隐私分类器的应用作出验证。

关键词：云计算；隐私分类；属性
中图分类号：TP391　　**文献标志码**：A　　**文章编号**：1001-3695(2014)05-1495-04
doi:10.3969/j.issn.1001-3695.2014.05.051

Research on privacy attibutions and classification in cloud computing

JI Yi-mu[a,b], KUANG Zi-zhuo[b], KANG Jia-bang[b], SUN Yan-peng[b], PAN Qiao-yu[b]
(*a. College of Computer, b. College of IoT, Nanjing University of Posts & Telecommunications, Nanjing* 210023, *China*)

Abstract: In order to protect users' privacy in a better way, this paper focused on the impacts made by the cloud computing to the users' privacy, including designing different privacy classifiers on the basis of traditional telecom users' privacy, adding the new privacies under the environment of Internet and cloud computing. The attribute would be researched and also be presented by XML. Finally, the designed customer relationship management system of telecommunication verified the application of users' privacy classifiers.

Key words: cloud computing; privacy classification; attributions

0　引言

随着科学技术的飞速发展，人们的日常生活对网络的依赖越来越强。互联网以及与之相关的产业发展日新月异，云计算(cloud computing)作为一种新的服务模式，受到各方的关注。云计算概念起源于 1999 年 Salesforce.com 提出的通过一个网站向企业提供企业级的应用。其发展的重要时期是在 2002 年 Amazon(亚马逊)提供一组包括存储空间、计算能力甚至人力智能等资源服务的 Web service。2005 年，亚马逊又提出了弹性计算云(elastic compute cloud，EC2)，允许小企业和私人租用亚马逊的计算机来运行它们自己的应用。直到 2008 年，几乎所有的主流 IT 厂商开始谈论云计算，这里既包括硬件厂商(IBM、HP、Intel、思科、SUN 等)、软件厂商(微软、Oracle、VMWare 等)，也包括互联网服务提供商(Google、亚马逊、Salesforce 等)和电信运营商(中国移动、中国电信、AT&T 等)，当然还有一些小的 IT 企业也将云计算作为企业发展战略。云计算具有高动态、高可靠性、以用户为导向的特点[1,2]，即对资源能进行合理的掌控与分配。这种资源管理模式的前提是数据的存储和安全完全由云计算提供商负责，如 Google doc。这种服务模式对于用户的隐私保护存在极大风险[3]，云服务提供商能够对用户的隐私信息进行直接侵犯或者对用户数据进行搜集和分析，挖掘出用户隐私数据。

1　云环境下用户隐私研究综述

云计算技术作为一种商业模式，具有扩展性好和成本低等特点[4]，已经在部分 IT 企业中展开研发和应用，适用于电信海量的用户信息和业务数据[5]。云计算在给人们带来巨大便利的同时，其隐私保护问题已成为社会各界关注的焦点，并严重威胁着网络社会与网络经济的健康发展。据波士顿咨询公司的一项调查显示，隐私比成本易用性和安全性等更为用户所关心。另据木星通信公司估计，2002 年，这种对网络隐私的担心造成了高达 180 亿美元的经济损失；到 2006 年，网络隐私将成为制约电子商务发展的最大障碍。据世界隐私论坛近日发布的一份报告声称，如果企业期望通过利用云计算服务来降低 IT 成本和复杂性，那么首先应保证这个过程中不会带来任何潜在的隐私问题[6]。

云安全联盟的白皮书[7]提出了云安全中最严重的安全威胁之一是数据丢失和数据泄露，而这个安全威胁涉及到三个领域的问题：a)信息生命周期的管理领域；b)加密和密钥管理领域；c)身份和访问管理领域。由此看来，存储在云服务端的数据很有可能被非可信的云计算服务提供商窃取，从而导致用户身份等隐私信息的泄露。

2　云计算面临的隐私安全隐患分析

尽管云计算产业的市场前景非常巨大，但是对于应用云计

收稿日期：2013-04-28；**修回日期**：2013-06-14　**基金项目**：江苏省工业支撑项目(BE2010057)；江苏省自然科学基金资助项目青年基金(BK20130876)；中国博士后科学基金第 54 批面上资助项目(2013M541702)

作者简介：季一木(1978-)，男，安徽无为人，副教授，博士，主要研究方向为云计算及其安全关键技术(jim_love@163.com)；匡子卓(1991-)，男，硕士研究生，主要研究方向为云安全中的用户隐私保护；康家邦(1987-)，男，硕士研究生，主要研究方向为云安全；孙延鹏(1989-)，男，硕士研究生，主要研究方向为云桌面和通信协议；潘俏羽(1988-)，女，硕士研究生，主要研究方向为云服务.

算服务的企业用户存在的潜在隐私安全风险也是不容忽视的。有关云计算隐私安全风险,这里将从基础设施层、传输层、应用层三个层次,从技术和业务相结合的角度进行分析。

2.1 基础设施层存在的安全隐患

在云计算时代,信息系统主要由云计算服务提供商负责,信息安全威胁的对象由各个组织转移到云计算服务提供商,信息安全威胁的目标将变得更为集中。传统数据中心硬件都相对独立可靠,但是在云计算环境下,虚拟化使硬件界限不再那么明显。如果缺乏系统的评估、科学的分析以及严谨的操作,不仅会导致硬件平台无法发挥出应有的效能,还会被黑客攻击而导致应用系统的崩溃。另外,云计算是基于分布式网络,计算机被看做是云环境下的一个节点。如果将该分布式网络接入互联网后,没有采取有效的安全保护措施,则"云"中的计算机就有可能利用用户上网留下的痕迹窥探用户上网的活动和行为。因此,基础设施层的用户虚拟化节点,主要是通过远程访问操作,这为基础设施层的节点留下了安全隐患。一旦虚拟化环境被攻破,存储和计算等资源就暴露了,存储虚拟化、用户实例可能被修改,从而使云环境受到破坏。

2.2 传输层存在的安全隐患

云计算就是通过其搭建的融合网络和系统平台,以服务的形式向用户提供应用需求。云计算提供的服务形式主要有 IaaS(基础设施服务)、PaaS(平台服务)和 SaaS(软件服务)[8~10],它们对应的成熟的服务平台是 Eucalyptus、Hadoop 和 Saleforce。用户在使用这些云计算平台提供服务的同时,网络传输和通信是必不可少的环节,因此通信数据包就有可能存在非法窃取、攻击、修改和破坏等问题。此外,云计算服务所涉及的数据基本都存放在云中,一旦由于技术原因出现服务中断,则整个服务过程也将会被中断。由此可见,在云计算普及过程中,网络传输的稳定性和安全性也决定着用户隐私是否可被能暴露的因素之一。

2.3 应用层存在的安全隐患

2007 年 1 月,安全专家就发现了 Google 云桌面存在安全漏洞,恶意用户不仅可以远程持续地获取 Google 桌面用户的敏感信息,甚至可以控制用户的整个电脑系统。2010 年 12 月,微软 BPOS(商务办公在线套装软件)发生了首次重大的云计算数据破坏事件,致使其数据被非授权用户下载。由此可见,互联网在大大方便人们的生活和工作的同时,也成为个人隐私的泄露渠道。

应用层是云计算出现隐私安全问题较集中的层次,因为应用层业务与最终用户直接接触,具体体现在两个方面[11]:a)用户管理。云计算平台面向的用户很多,包括云计算提供商、运维人员和普通客户。云计算提供商的主要职责是保证云计算服务中的用户数据不被非法窃取与利用。此外,云计算中的客户位于世界各个地方,其在线数和注册数是随着时间不断变化的,需要保证"特权用户"被客户识别,并安全登录。运维人员则要负责对数据进行分级存储和管理,并做好定期备份工作。b)存储管理。在云计算服务中,一旦用户将数据存储到云中,云服务提供商就对云中的存储数据具有各种操作权限,一些不法提供商就可能以不正当的手段从云计算提供商处获取用户的隐私资料。在数据存储时,由于升级等原因,云计算服务提供商还可能会将用户数据从一个服务器转移到另一服务器中,造成数据存储位置的不确定性,这也增加了用户隐私数据的不安全性。

在云计算环境下,云计算的安全主要由云计算提供商负责,云计算作为当前新兴的计算模式,其对用户信息隐私的信息保护已成为人们关注的焦点。云计算通常以用户为导向,对资源进行合理的掌控和分配,但是这种管理模式是以数据的存储和安全完全由云计算提供商负责为前提的,即身份信息、敏感信息由云计算提供商存储和管理,其安全程度与服务提供商的可信赖度和保护隐私的技术水平有极大的关系。正因为云计算服务通常对客户并不完全透明,一旦有意外发生,客户也并不清楚其数据面临的情况,因此加强对他们的监管是十分必要的。综上所述,云环境下影响用户隐私安全泄露可能的指标如表 1 所示。

表 1 云环境下影响用户隐私的安全指标

	基础设施安全隐患(IaaS 服务)		传输交互安全隐患(PaaS 服务)		业务定制访问安全隐患(SaaS 服务)	
1	数据存储	泄露用户隐私数据	传输协议	抓取数据包并破解	用户认证	身份安全性
2	存储虚拟化	稳定性、安全性	跨网络	不同域之间的安全策略更复杂	数据操作	数据安全性
3	带宽	网络时延 QoS	跨平台	不同平台的安全访问控制策略不同	业务处理	平台安全性

3 云计算环境下用户隐私分类

3.1 云计算环境下用户隐私分类概述

云用户数据可以分为公共信息、私有信息和限制信息[12]三个模块。隐私保护就是保护私有和限制的信息不被他人所获取,其中隐私就是个人、机构等实体不愿意被外部知晓的信息。在具体应用中,隐私即为数据所有者不愿意被披露的敏感信息,包括敏感数据以及数据所表征的特性。文献[13,14]从隐私所有者的角度将隐私分为个人隐私和共同隐私两类。文献[15]基于属性关联和敏感参数将云用户数据分为认证信息、基本信息和元数据三类,并对这些属性进行了描述。文献[16]提出了数据组合隐私的概念,数据组合隐私为个体不希望暴露的一系列数据属性的组合,通过这些数据值组合起来又可以确定特定个体。文献[17]将用户数据记录属性分为显式标志符、准标志符、敏感信息和非敏感属性。

传统的用户隐私包括用户的基本资料、登录密码、相关受理业务、绑定的银行账号和个人偏好等,而云环境下的用户隐私还包括电子财务数据、位置隐私、浏览踪迹、服务端记录信息、软件使用习惯和操作状态、敏感数据等。在云计算服务模式下,数据提供者和数据访问者不再是简单的对应关系,隐私已不同于传统的隐私保护。怎么保护个人隐私不被泄露是云计算面临的一个技术挑战,如何管理云计算中各种成员的身份以及如何保护其身份信息不被云计算服务提供商泄露,成为云计算中隐私保护的关键。

在了解传统用户隐私的基础上,加上云环境新的特征,如网络开放、远程操作和多用户协同等,本文设计了一个云环境下的用户隐私分类器,用于对用户隐私进行度量和归类,以及对用户隐私属性进行挖掘和描述。

3.2　云计算环境下用户隐私属性分类视图

3.2.1　基于信息来源的用户隐私分类

云环境下用户隐私可以从个人基本信息、业务数据和 Web 环境涉及到的用户隐私三个维度进行划分,如图 1 所示。隐私分类器的作用就是将以上类别中的隐私进行分类和度量。因为云环境是一个开放的 Web 环境,所以除了考虑个人的信息和数据隐私外,还要考虑网络信息的隐私。可以将隐私的度量从等级上分为 S 级,A 级和 B 级三类。S 级是绝密的隐私信息,A 级是较为保密的隐私信息,B 级是保密性较低的隐私信息。如个人登录身份和密码、电话号码、身份证号、电子财务、用户个人偏好、位置隐私、网页活动内容、浏览踪迹以及服务端记录日志都是 S 级绝密隐私,需要使用更加安全的隐私保护技术去保证数据的绝密性,如使用访问控制、身份认证和数据加密等综合方法。A 级别的隐私包括家庭住址、邮箱地址、电话和短信记录、工作单位、工作类别和软件操作状态,这些可以采用相对安全的隐私保护技术,如加密机制就可以起到保护作用。B 级包括个人信用、个人业务、浏览流量、消费账单,这些可采用较为简单的隐私保护技术,如身份认证,只要通过身份认证的用户就可以获得这类信息。在进行隐私保护之前,首先要明确用户隐私属于哪一类,用户隐私又有哪些属性。当然,这种划分也不是一成不变的,可以根据不同行业、不同业务和不同安全需求进行扩展。

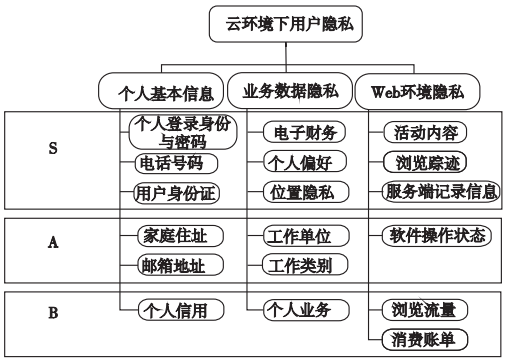

图1　按照信息来源的属性分类

3.2.2　基于保护等级的用户隐私分类

如果对用户隐私按照等级直接进行划分,可分为高、中、低三个等级,然后对应等级给相应隐私采取保护技术,其分类设计如图 2 所示。其中,个人登录身份与密码、电话号码、用户身份证、电子财务、个人偏好、工作类别是等级较高的隐私信息;邮箱地址、家庭住址、位置隐私、消费账单、个人信用、工作单位是中级隐私;活动内容、浏览踪迹、服务器记录信息、软件操作状态、浏览流量、个人业务则是等级较低的隐私。

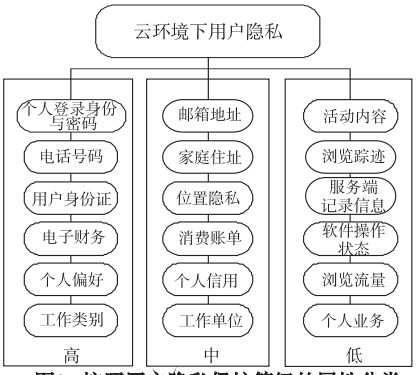

图2　按照用户隐私保护等级的属性分类

通过两种不同角度,对云环境下用户隐私保护属性进行分类,针对不同级别 S(高)、A(中)和 B(低)的分类采用不同的保护策略,如表 2 所示。

表 2　用户隐私属性按等级分类的保护方法

等级	云环境下用户隐私保护等级			其他(多数互联网 Web 服务)
方法	S(高)	A(中)	B(低)	
保护方法、手段、策略	访问控制 身份认证 数字签名 数据加密	访问控制 身份认证 数字签名	访问控制 身份认证	身份认证

3.3　基于 XML 的用户隐私属性描述方法

用户的隐私属性可以基于 XML 进行描述,XML 易于扩展,而且容易与其他开发工具进行交互和集成处理,非常适合 Web 传输。EPAL[18] 就是一种基于 XML 的形式化语言隐私保护标准和规范。基于 XML 的属性描述方法也具有较强的规范性,可以明确地表示电信用户隐私属性中各个部分的意义。云环境下用户(包括个人和企业等)基本隐私属性基于 XML 的描述规范如下:

```
〈Basic personal information〉
    〈username〉Zhangsan〈/username〉        //用户名称
    〈sex〉Male〈/sex〉        //性别
    〈nationality〉 the Han nationality 〈/nationality〉        //民族
    〈age〉20〈/age〉        //年龄
    〈password〉123456〈/password〉        //密码
    〈telephone〉15666666666〈/telephone〉        //电话号码
    〈QQ〉362888888〈/QQ〉        //QQ 号码
    〈e-mail〉362888888@ qq. com〈/e-mail〉        //电子邮箱
    〈address〉Nanjing * * * * * * * * * 〈/address〉        //家庭地址
    〈personal credit〉☆☆☆〈/personal credit〉        //个人信用
    〈/ Basic personal information 〉
```

不同领域和行业的业务都可用 XML 进行属性描述。以电信客户关系管理系统中的语音话费详单查询业务为例,就涉及到用户的电话号码、拨出/接听号码、通话日期、通话时长、收费类别、费用合计等属性。业务、业务对应的属性以及属性的描述流程如图 3 所示。

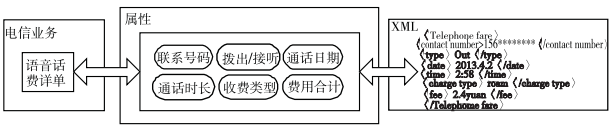

图3　电信业务相关属性的建模和表示流程

4　云环境下电信行业 CRM 用户隐私保护原型系统设计

图 4 为基于 Hadoop 的用户隐私保护原型系统的电信 CRM 业务系统,采用分层的思想,自顶向下,每层都透明地调用下层接口,最顶层为交互层,用于用户和系统之间的交互;最底层为分布式计算层,使用 Hadoop 来实现文件分布式存储和并行计算功能;使用分层,各层之间变得独立,易于系统的扩展。

图 4 中的用户隐私属性分类层用于根据电信用户隐私属性对底层数据进行分类。根据属性,可以将用户隐私分为个人基本信息、业务敏感数据和 Web 环境;用户隐私安全保护层,通过调用用

户隐私分类后的隐私信息进行保护。首先针对不同的属性分类,制定相应的属性访问控制策略,并根据不同的业务类型,采取相应的 ABE 属性加密方法,最后使用 ABS 属性签名方法对加密的数据进行签名,从而达到用户隐私属性高级别保护的目的。

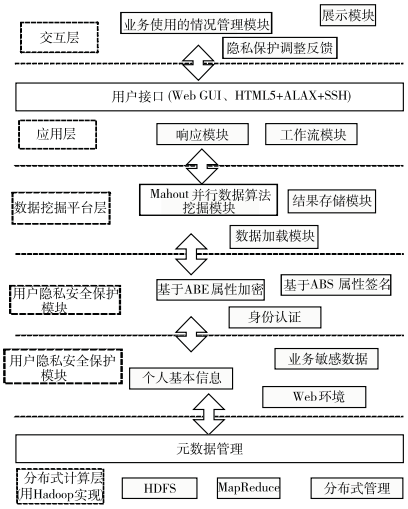

图4 云环境下电信CRM业务系统隐私保护原型系统框架

5 结束语

云计算的应用几乎涉及了信息管理与服务的各个领域,云计算中的隐私问题若得不到很好地解决,则将会直接影响到用户的信息安全和选择云计算的信心。随着电子商务、电子政务、网络教育的兴起与蓬勃发展,人们在网上的活动越来越多,如网上浏览查询、聊天、购物、学习、收发邮件等。在这些网上活动中,将涉及到大量的个人隐私信息,虽然网络导致的隐私泄露已引起社会各界的广泛关注,社会也采取和制定了许多法规、自律政策和技术等措施来加强对网络隐私的保护,但这些措施的作用非常有限,隐私保护仍然是一个亟待解决的挑战性问题,进一步将深入研究云计算环境下的用户隐私属于的形式化表示和隐私保护原型系统的开发。

参考文献:

[1] SU Mu-chun, CHOU C H. A modified version of the K-means algorithm with a distance based on cluster symmetry[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(6): 674-680.

[2] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[J]. Communication of the ACM-50th Anniversary Issue: 1958-2008. 2008, 51(1): 107-113.

[3] ARMBRUST M, FOX A, GRIFFITH R, *et al*. Above the clouds: a Berkeley view of cloud computing, UCB/ EECS-2009-28 [R]. Berkeley: University of California at Berkeley, 2009.

[4] WANG Li-zhe, TAO Jie, KUNZE M, *et al*. Scientific cloud computing: early definition and experience [C]// Proc of the 10th International Conference on High Performance Computing and Communications. Washington DC: IEEE Computer Society, 2008: 825-830.

[5] KAUFMAN L. Data security in the world of cloud computing[J]. IEEE Security and Privacy, 2009, 7(4): 61-64.

[6] 张军,熊枫. 网络隐私保护技术综述[J]. 计算机应用研究,2005, 22(7):9-11.

[7] Cloud Security Alliance. Security guidance for critical areas of focus in cloud computing v2.1[EB/OL]. (2011-12-24). http://www.cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf.

[8] GODSE M, MULIK S. An approach for selecting software-as-a-service (SaaS) product cloud computing[C]//Proc of IEEE International Conference on Cloud Computing. 2009: 155-158.

[9] LIU Feng, GUO Wei-ping, ZHAO Zhi-qiong, *et al*. SaaS integration for software cloud[C]//Proc of the 3rd IEEE International Conference on Cloud Computing. 2010: 402-409.

[10] ZHANG Yong, LIU Shi-jun, MENG Xiang-xu. GridSaaS: a grid-enabled and SOA-based SaaS application platform[C]//Proc of IEEE International Conference on Services Computing. 2009: 521-523

[11] 李晓飞. 云计算环境下的用户隐私问题浅析[J]. 南昌教育学院学报,2013,28(2):194-196.

[12] SOOD S K. A combined approach to ensure data security in cloud computing[J]. Journal of Network and Computer Applications, 2012, 35(6): 1831-1838.

[13] CLINFTON C, KANTARCIOGLUGLU M, VAIDYA J. Defining privacy for data mining[C]//Proc of National Science Foundation Workshop on Next Generation Data Mining. 2002.

[14] 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):848-858.

[15] WAQAR A, RAZA A, ABBAS H, *et al*. A framework for preservation of cloud users' data privacy using dynamic reconstruction of metadata [J]. Journal of Network and Computer Applications, 2013, 36(1): 235-248.

[16] 张坤. 面向多租户应用的云数据隐私保护机制研究[D]. 济南:山东大学,2012:20-38.

[17] 兰丽辉,鞠时光,金华,等. 数据发布中的隐私保护研究综述[J]. 计算机应用研究,2010,27(8):2823-2827.

[18] Enterprise privacy authorization language (EPAL 1.2)[EB/OL]. http://www.w3.org/Submission/EPAL/..

(上接第 1494 页)

[3] MELARA A J. Performance analysis of the Linux firewall in a host [D]. [S.l.]: California Polytechnic State University, 2002.

[4] Arbor Networks Inc. Worldwide infrastructure security report[EB/OL]. http://www.arbornetworks.com/report.

[5] SALAH K. Queueing analysis of network firewalls[C]// Proc of IEEE Global Telecommunications Conference. 2010: 1-5.

[6] SALAH K, QAHTAN A. Implementation and experimental performance evaluation of a hybrid interrupt-handling scheme[J]. International Journal of Computers Communications Control, 2009, 32(1): 179-188.

[7] LAW A, KELTON W. Simulation modeling and analysis[M]. 2nd ed. [S.l.]: McGraw-Hill, 1991.

[8] LIU A X, GOUDA M G. Diverse firewall design[J]. IEEE Trans on Parallel and Distributed Systems, 2008, 19(9): 1237-1251.

[9] SALAH K, SATTAR K, SQALLI M, *et al*. A potential low-rate DoS attack against network firewalls[J]. Security and Communication Networks, 2011, 4(2): 136-146.

[10] Distributed Internet traffic generator[EB/OL]. (2008). http://www.grid.unina.it/software/ITG.

[11] ZANDER S, KENNEDY D, ARMITAGE G. KUTE: a high performance kernel-based UDP traffic engine[EB/OL]. (2005). http://caia.swin.edu.au/reports/050118A/CAIA-TR-050118A.pdf.