云数据安全存储技术

冯朝胜[1),2)] 秦志光[2)] 袁 丁[1)]

[1)](四川师范大学计算机科学学院 成都 610101)

[2)](电子科技大学计算机科学与工程学院 成都 610054)

摘 要 云计算因具有资源利用率高、节约成本等诸多优点而将成为未来的主流计算模式.然而,包括隐私保护在内的数据安全存储问题却成为云计算推广的巨大障碍.该文首先列举了云计算在数据安全上面临的主要挑战,指出了云计算的租用商业模式和其采用的两种关键技术——虚拟化技术和多租户技术是云存储存在诸多安全问题甚至安全悖论的根本原因.从加密存储、安全审计和密文访问控制 3 个方面对云数据安全存储的最新研究进展分别进行了评述.在加密存储上,介绍了云数据安全存储框架和主要的安全存储技术;在安全审计上,分析了外包数据安全审计,特别是公开审计面临的主要难题,介绍了包括云数据在内的外包数据完整性公开证明的主要模型和方法,并指出了它们的优势和不足;在云密文的访问控制上,详述了基于属性的云密文访问控制方法,并指出了这些方法的优劣.最后指出了云数据安全存储研究面临的主要问题并预测了相关研究的未来发展趋势.

关键词 云计算;数据存储;数据加密;安全审计;密文访问控制

中图法分类号 TP393 **DOI 号** 10.3724/SP.J.1016.2015.00150

Techniques of Secure Storage for Cloud Data

FENG Chao-Sheng[1),2)] QIN Zhi-Guang[2)] YUAN Ding[1)]

[1)](*School of Computer Science, Sichuan Normal University, Chengdu* 610101)

[2)](*School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu* 610054)

Abstract Cloud computing will become the main computing model in the future due to its advantages such as high resource utilization rate and high cost performance. How to securely store data including privacy data, however, becomes a huge impediment to its development. In this paper, the challenges, which cloud computing is confronted with, are listed first. The renting mode of cloud computing and its two key techniques, i.e. the virtualization and multi-tenant, are identified to result in these problems. And then the recent studies on cloud storage are reviewed in terms of cipher storage, security audit and cipher access control. The focuses involve in the framework and key techniques of cloud data storage, the problems and methods of security audit, and attribute-based access control methods as well as advantages and disadvantages of these techniques and methods. At last, the problems that the study of secure cloud storage is confronted with are identified, and further, the study trend of secure storage for cloud data is analyzed and predicted.

Keywords cloud computing; data storage; data encryption; security audit; cipher access control

---

收稿日期:2013-06-25;最终修改稿收到日期:2014-09-04.本课题得到国家自然科学基金(61373163)、国家科技重大专项课题(2011ZX03002-002-03)、国家科技支撑计划课题(2014BAH11F01,2014BAH11F02)、可视化计算与虚拟现实四川省重点实验室课题(PJ2012002)资助.**冯朝胜**,男,1971 年生,博士后,教授,硕士生导师,中国计算机学会(CCF)高级会员,主要研究领域为云计算、隐私保护、数据安全.E-mail:csfenggy@sicnu.edu.cn.**秦志光**,男,1956 年生,博士,教授,博士生导师,主要研究领域为信息安全、分布式计算.**袁 丁**,男,1967 年生,博士,教授,主要研究领域为数据安全和密码学.

1 引　言

云计算[1]因能有效解决信息系统计算及存储能力不足、资源利用不充分、IT 设施投入大、系统管理复杂等问题而颇受欢迎. Gartner 的数据显示①，云计算已取代虚拟化技术成为 2011 年全球 CIO 最关注的技术领域，2009 年全球公有云服务市场规模约为 586 亿美元，到 2014 年将达到 1488 亿美元；而 55%的中国被调研企业表示，到 2013 年将会花费超过整体 IT 预算的 10%到云计算项目采购之中. 赛迪顾问发布的《中国云计算产业发展白皮书》②(2011 版)显示：未来 3 年，中国云计算应用将以政府、电信、教育、医疗、金融等行业为重点，市场收入规模将从 2010 年的 167.31 亿元增长到 2013 年的 1174.12 亿元，年复合增长率将达 91.5%. 由于云计算的发展理念符合当前低碳经济与绿色计算的总体趋势，极有可能发展成为未来网络空间的神经系统，它也为世界各国政府所大力倡导与推动.

云计算在迅猛发展的同时，其安全问题，尤其是数据存储的安全性和隐私性问题日益突出[2-3]. Gartner 2009 年的调查结果显示，70%以上受访企业的 CTO 认为近期不采用云计算的首要原因是存在对数据安全性与隐私性的忧虑；而近来，Amazon、Google 等云计算发起者不断被爆出各种安全事故更加剧了人们的担忧. 例如，2009 年 3 月，Google 发生大批用户文件外泄事件③；2010 年 6 月苹果公司发生 iPad 用户隐私数据泄漏事件④. 网络巨头思科公司首席执行官 John Chambers 预言，数据安全将成为云计算前进道路上的“噩梦”.

云计算数据存储的安全性问题已引起学术界和产业界的广泛关注，相关研究已围绕数据保密存储、安全审计和密文访问控制 3 个领域展开[4]，本文对其中的具有代表性的研究进行了分析和评述，对存在的问题进行了归纳，对云数据安全存储研究的未来发展趋势进行了预测.

2 云数据安全存储面临的挑战

2.1 云环境下数据安全面临的挑战

在传统信息系统中，数据安全主要关注数据的加密存储和传输、安全审计和容灾备份；而在云中，除了要关注上述内容外，还有更多关注，云计算的特点决定了⑤要实现集中式的数据存储，必须确保不同用户数据的安全隔离；云端的服务器可能会“宕机”，在这种情况下，如何高效地进行数据安全地迁移很关键；云计算采用租赁方式向用户提供资源，这意味着一个用户使用过的存储区域会被其他用户使用，因而必须解决好数据残留问题.

云环境下数据安全存储面临以下挑战[4-5]：

(1) 数据的加密存储

在传统的信息系统中，一般采用加密方式来确保存储数据的安全性和隐私性. 在云中，似乎也可以这样做，但实现起来却不那么容易. 在基础设施即服务(Infrastructure as a Service，IaaS)云模式中，由于授权给用户使用的虚拟资源可以被用户完全控制，数据加密既非常有必要也容易做到(无论是在公有云或者私有云中). 但在平台即服务(Platform as a Service，PaaS)云模式或者软件即服务(Software as a Service，SaaS)云模式中，如果数据被加密，操作就变得困难. 在云中，对于任何需要被云应用或程序处理的数据，都是不能被加密的，因为对于加密数据，很多操作像检索、运算等都难以甚至无法进行. 数据的云存储面临这样的安全悖论：加密，数据无法处理；不加密，数据的安全性和隐私性得不到保证.

(2) 数据隔离

多租户技术是 PaaS 云和 SaaS 云用到的关键技术. 在基于多租户技术系统架构中，多个租户或用户的数据会存放在同一个存储介质上甚至同一数据表里. 尽管云服务提供商会使用一些数据隔离技术(如数据标签和访问控制相结合)来防止对混合存储数据的非授权访问，但非授权访问通过程序漏洞仍然是可以实现的，比如 Google Docs 在 2009 年 3 月就发生过不同用户之间文档的非授权交互访问. 一些云服务提供商通过邀请第三方或使用第三方安全工具软件来对应用程序进行审核验证，但由于平台上的数据不仅仅针对一个单独的组织，这使得审核标准无法统一.

(3) 数据迁移

当云中的服务器(这里，服务器是指提供 SaaS 和 PaaS 的物理机，对于 IaaS 而言，服务器或者是物理机，或者是虚拟机)“宕机”时，为了确保正在进行

① http://www.gartner.com/it/page.jsp?id=1526414
② http://tech.ccidnet.com/zt/cwb/images/cloudbook.pdf
③ http://www.techcrunch.com/2006/12/28/gmail-disaster-reports-of-mass-email-deletions/
④ http://techcrunch.com/2010/06/15/ipad-breach-personal-data/
⑤ http://www.cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf

的服务能继续进行,需要将正在工作的进程迁移到其他服务器上.进程迁移,实质上就是对与该进程相关的数据进行迁移,迁移的数据不仅包括内存和寄存器中动态数据(或称进程快照),还包括磁盘上的静态数据.为了让用户几乎无法感觉到"宕机"的发生,迁移必须高速进行;为了让进程能在新的机器上恢复运行,必须确保数据的完整性;另外,如果进程正在处理的是机密数据,还必须确保这些数据在迁移过程不会泄露.

(4)数据残留

数据残留是指数据删除后的残留形式(逻辑上已被删除,物理上依然存在).数据残留可能无意中透露敏感信息,所以即便是删除了数据的存储介质也不应该被释放到不受控制的环境,如扔到垃圾堆或者交给其他第三方.在云应用中,数据残留有可能导致一个用户的数据被无意透露给未授权的一方,不管是什么云,SaaS、PaaS 或 IaaS 都有可能.如果一个未授权数据泄露发生,用户可以要求第三方或者使用第三方安全工具软件来对云服务提供商的平台和应用程序进行验证.迄今为止,没有哪个云服务提供商解决了数据残留问题.

(5)数据安全审计

当数据以外包方式存储在云中时,用户会关注两个问题:外包存储的数据确实已存储到云中并归数据所有者所有;除所有者和授权用户外的任何人不能更新数据.这两个问题的解决都离不开安全审计.在数据存放到本地或企业可信域中时安全审计较易实现,而一旦将数据以外包方式存储到云中时,安全审计就变成了难题.显然,用户不可能将数据都下载下来后再进行审计,因为这会导致巨大的通信代价,更可行的思路是:只需取回很少数据,通过某种知识证明协议或概率分析手段,就能以高置信概率判断云端数据是否完整或为用户所有.

2.2 云存储问题的成因

云计算的商业模式和其采用的两种关键技术——虚拟化和多租户是其面临安全挑战的主要原因.

2.2.1 云计算商业模式

云计算商业模式本质上是租用模式,根据租用资源类型可分为计算资源租用模式和存储资源租用模式,分别被称作计算外包和存储外包.

传统模式下,个人用户或企业用户将数据存储在自己的可控信任域(对于个人用户,可控信任域指其使用的终端设备)中,用户可以完全控制自己的数据.当用户采用公有云来存储数据即存储外包时,数据将不再处于自己的可控信任域之内,而是处于云服务提供商的控制域内,如图 1 所示.这种情况下,用户隐私和数据不仅可能泄漏给云服务商,还可能泄漏给包括竞争对手在内的其他用户(因为云服务器提供商可能出于经济目的将用户隐私信息出卖).在私有云中,也存在同样情况:个人用户的隐私信息容易泄漏,只是泄漏范围有极大的缩小:单位内部工作人员.如果仅仅是采用计算外包,那么,对应的云计算模式是 SaaS 和 PaaS,而这两种模式正如前所述是不会(至少目前是)对要处理的数据进行加密的.在这种情况下,用户敏感数据就容易泄漏给云服务提供商和同一机器上的其他租户.

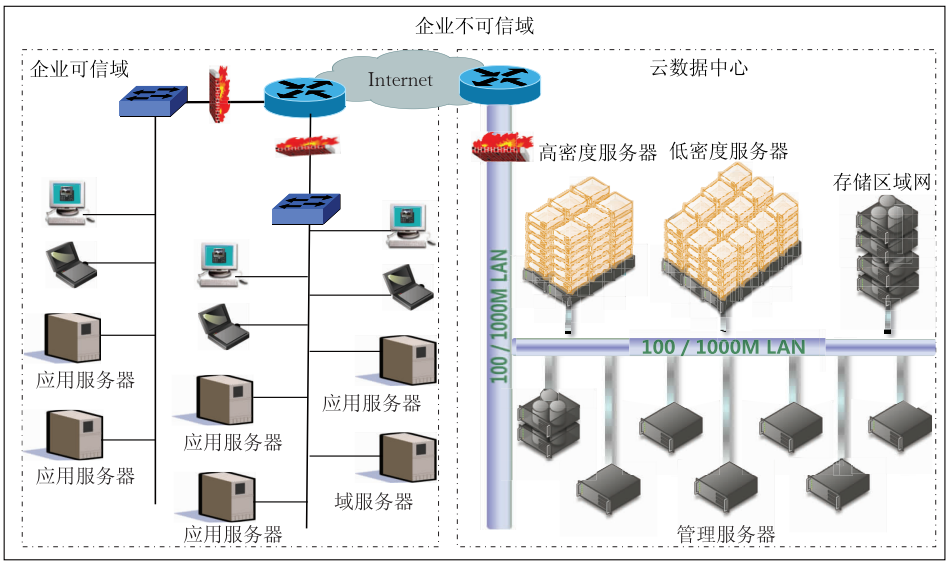


图 1 基于公有云的企业信息系统部署图

2.2.2　虚拟化技术

虚拟化[6]是 IaaS 云采用的关键技术，是资源能动态伸缩和充分利用的关键原因. 通过对 CPU、内存等硬件资源的虚拟化，同一台物理机上可以同时运行多台虚拟机(IaaS 云模式). 尽管这些共享着相同硬件资源的虚拟机在虚拟机监控器 VMM (Virtual Machine Monitor)的控制下彼此隔离，即使是采用了如图 2 所示的数据保护措施. 攻击者通过虚拟机逃逸、流量分析、旁路攻击等攻击手段仍然可以从一台虚拟机上获取其他虚拟机上的数据.

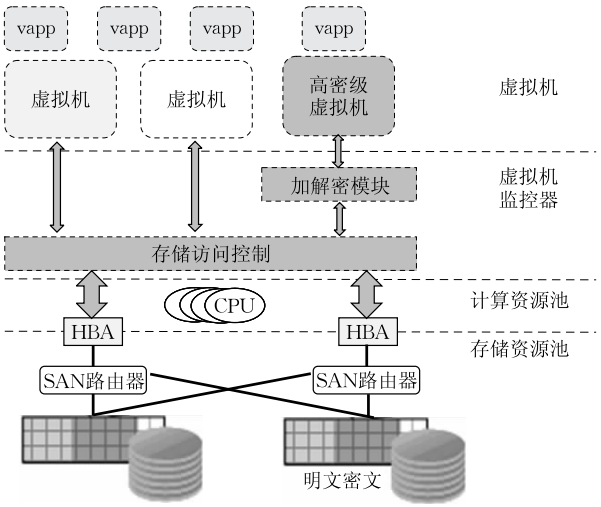


图 2　虚拟化环境下的数据安全存储模型

2.2.3　多租户技术

多租户[6]技术是 SaaS 云模型采用的关键技术. 该技术使云中的同一个应用进程(如 Google Docs)可以同时为多个租户使用(如图 3 所示)，这些租户的数据一般存放在同一张数据表上，采用标签进行区分. 访问控制技术用来确保每个租户只能访问自己的数据而不能访问其他租户的数据. 不过，恶意租户采用漏洞攻击、旁路攻击等方法仍然可以获得其他用户的数据[7].

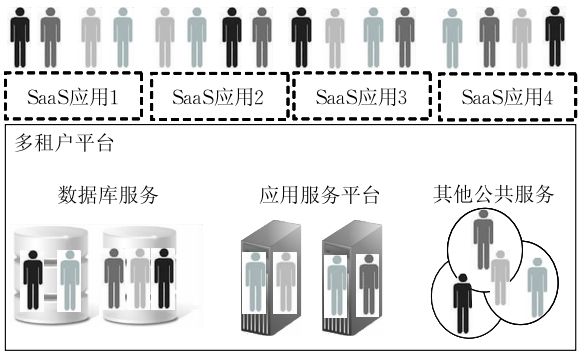


图 3　多租户技术示意图[6]

外包模式、虚拟化技术和多租户技术对云数据安全存储的影响如表 1 所示. 从表中不难看出，云数据安全存储关注的焦点是：加密存储、安全审计和密文访问控制，而这 3 种技术分别用来确保数据安全存储的三大目标：机密性、完整性和可用性. 表中列出的有些解决方案还处于理论探索阶段，如同态加密. 对于数据残留问题，目前还没有发现合适的解决方案.

表 1　外包模式、虚拟化技术和多租户技术对云数据安全存储的影响

外包模式	云服务模型	租用资源(共享)	关键技术	安全挑战	解决方案
计算	SaaS 或 PaaS	CPU＋内存	多租户	云服务提供商用快照获取内存数据	同态加密
				用户数据被同一机器上其他用户窃取(数据隔离问题)	访问控制
存储	IaaS	外存	虚拟化	数据或隐私泄露给云服务提供商或其他用户	加密＋访问控制
				服务提供商擅自删除用户数据(数据安全审计)	安全审计
				用户已删除数据为其他用户获取(数据残留)	无
计算＋存储	SaaS 或 PaaS	CPU＋内存＋外存	多租户	云服务提供商用快照获取内存数据	同态加密
				用户数据被同一机器上其他用户窃取(数据隔离问题)	访问控制
				数据或隐私泄露给云服务提供商或其他用户	加密＋访问控制
				服务提供商擅自删除用户数据(数据安全审计)	安全审计
				用户已删除数据为其他用户获取(数据残留)	无
计算＋存储	IaaS	CPU＋内存＋外存	虚拟化	数据或隐私泄露给云服务提供商或其他用户	访问控制＋加密
				服务提供商擅自删除用户数据	安全审计
				用户已删除数据为其他用户获取(数据残留)	无
				数据安全迁移	快照＋共享外存

3　云数据安全存储研究现状

从已有研究看，云数据安全存储研究主要关注加密存储、完整性审计和密文访问控制这 3 个方面，下面就从这 3 个方面展开研究现状分析.

3.1　云数据加密存储技术研究

加密无疑是保护云中存储的数据的安全性和隐私性的重要方法之一，当前对云数据加密存储的研究主要围绕云数据安全存储框架和安全存储技术进行，被关注的安全存储技术包括同态加密技术、基于 VMM 的数据保护技术、基于加解密的数据安全存

储技术、支持查询的数据加密技术和面向可信平台的数据安全存储技术.

3.1.1 云数据安全存储框架

微软研究院的 Kamara 等人[7]提出了面向公有云的加密存储框架,如图 4 所示.在该框架中,数据处理 DP、数据验证 DV、令牌生成 TG 和凭证生成 CG 是核心组件,这些组件工作在数据所有者的可信域中.数据处理组件负责在数据存储到云中前对数据进行分块、加密、编码等操作;数据验证组件负责验证存储在云中的数据块的完整性;令牌生成组件负责生成数据块访问令牌,云存储服务根据用户提供的令牌提取相应的密文数据;凭证生成组件负责为授权用户生成访问凭证.在访问授权时,数据所有者会将共享文件的令牌和凭证发往授权用户.授权用户使用令牌从云中提取共享文件的密文,使用凭证解密文件.该框架的主要特点有两个:数据由所有者控制;数据的安全性由密码机制保证.该框架除了能解决数据存储的隐私问题和安全问题外,还能解决数据访问的合规性、法律诉讼、电子取证等问题.不过,该框架只是一个宏观的模型,并没有给出具体实现方法.

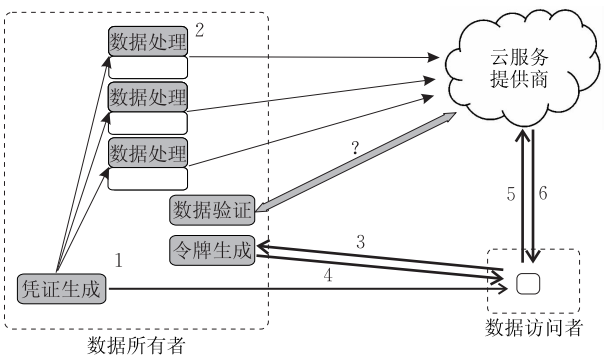


图 4 面向公有云的加密存储框架[7]

文献[8]提出了一种分散式云存储安全架构.该架构采用信息扩散法、分散存储管理、数据自举恢复等技术,分层实现数据的安全存储管理和传输.该方法定期检查数据片受损情况,若存在受损数据,则根据互为冗余的存储设备上的数据加以恢复,从而提高数据的可用性.从数据存储到传输,都建立了相应的保护措施,进行云存储层与其他层间的安全防范,实现了数据的有效防护.该架构并没有具体说明如何保证数据的完整性,数据分片难以确保数据的隐私性和安全性.

3.1.2 云数据安全存储技术

(1)同态加密技术

同态加密是一种加密技术,运用这种技术可以实现对明文上执行指定的代数运算结果等同于在密文上的另一个(可能是不同运算)代数运算结果[①]同态加密,这个特性使得云计算面临的数据存储悖论迎刃而解.同态加密的思想起源于私密同态(privacy homomorphism),它允许在不知道解密函数的前提下对加密数据进行计算.设 S 和 S' 分别为明文空间和密文空间,$a,b\in S$,E 是 $S\to S'$ 上的加密函数.如果存在算法 $PLUS$ 和 $MULT$,使其满足

$$E(a+b)=PLUS(E(a),E(b)),$$

$$E(a\times b)=MULT(E(a),E(b)).$$

这样可以利用 $E(a)$ 和 $E(b)$ 的值计算 $E(a+b)$ 和 $E(a\times b)$,而不需要知道 a,b 的值,称其分别满足加法同态和乘法同态.对于一个加密函数,如果同时满足加法同态和乘法同态,就称其为全同态加密函数;否则,即要么满足加法同态(如 Paillier 算法),要么满足乘法同态(如 RSA),就称其为部分同态加密函数.然而,设计全同态加密函数十分困难.2009 年,IBM 宣布了一项研究成果,称其实现了全同态数据加密方案.研究人员 Craig Gentry 使用称为“理想格(ideal lattice)”的数学对象,使密文数据得到充分操作[9].整个方案主要包括 3 个关键步骤:第 1 步是构建一个受限同态加密算法,该算法支持密文的低阶多项式运算;第 2 步是将解密操作“打散”成更小的子操作,这些子操作可以表示成低阶多项式运算;第 3 步是利用“引导程序”将受限同态加密算法转变成全同态加密算法.该方案密文处理效率很低,离实际应用还有较长时间.

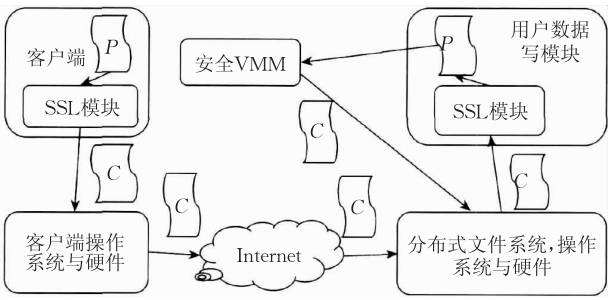
文献[10]设计了一个同态加密算法.该算法通过运用向量和矩阵的各种运算来实现了对数据的加密和解密,并支持对加密字符串的模糊检索和对密文数据的加、减、乘、除 4 种算术运算.该算法执行同态加减运算的效率较高,但在执行密文检索和同态乘除运算时效率很低,且运算代价随向量维度的增加而增加.

(2)基于 VMM 的数据保护技术

鉴于云环境下虚拟机工作在虚拟化平台之上并由虚拟机监控系统或监控器进行管控,文献[11]提出了一种基于 VMM 的云数据机密性保护方法,如图 5 所示.该方法基于 SSL 来保证数据传输的安全,利用 Daoli 安全虚拟监控系统保护数据存储的安全.数据在传输到云端前,用户客户端 SSL 模块会将数据加密.云端的操作系统接收到用户密文数据后,将密文数据提交给分布式文件系统.分布式文件系统的 SSL 模块会将数据解密以进行处理.如果

① http://zh.wikipedia.org/wiki/

用户要将数据保存到分布式文件存储系统，虚拟监控系统会在存储前对数据进行加密；反之，如果用户要从分布式文件存储系统中读取数据，虚拟监控系统会先将数据解密. 该方法显著特点是将云端的操作系统和分布式文件系统进行了隔离，数据加解密由虚拟机监控系统来完成，实现了操作系统和用户数据的隔离. 由于对于操作系统而言数据始终是加了密的密文，当虚拟机操作系统被攻破时，攻击者得到的都是加了密的密文数据，保证了内存数据和硬盘数据的安全性和机密性. 该方案能保证多租户环境下隐私数据不会泄露给其他用户，但数据还是可能会泄漏给云服务提供商.



C: 密文　P: 明文

图 5　基于 Daoli 虚拟监控系统的数据安全存储模型[11]

(3) 基于加解密的数据安全存储技术

公有云中存储的数据一般属于外包数据，存在不少基于传统的加解密技术的研究来确保外包数据的安全. 文献[12]提出了基于代理重加密方法的数据分布式安全存储方案. 数据所有者使用对称的内容加密密钥来加密文件内容，再使用主公钥加密所有的内容加密密钥，只有拥有主私钥的所有者才能解密这些内容加密密钥. 所有者使用其拥有的主私钥和用户的公钥来生成代理重加密密钥. 半可信服务器能使用代理重加密密钥将密文转化成指定授权用户能解密的密文，进而实现访问控制. 该方案的主要问题是存在恶意服务器和任意一个恶意用户勾结就能计算出所有密文数据的解密密钥的漏洞，严重威胁着数据的安全. 除此之外，用户访问权限得不到保护也是该方案的明显不足. 文献[13]提出了一种基于密钥导出方法的非可信服务器数据安全存储方案. 在该方案中，每个文件都用一个对称密钥加密，每个用户都拥有一个私钥. 为进行授权，数据所有者为授权用户创建公开令牌. 授权用户能利用自己的私钥从令牌中导出指定文件的解密密钥. 服务器虽然拥有令牌，但其并不能从令牌中导出解密密钥. 该方案的不足是文件创建操作及用户授权/撤销的复杂性与用户数量成线性关系，这使得系统规模难以扩展. 文献[14]设计了一种将文件安全存储在不可信服务器上的文件加密系统 Plutus. 该系统基于共享属性的相似性对文件进行分组，每个分组关联一个称作锁箱密钥的对称密钥. 使用文件块密钥对文件进行加密，再使用文件所属文件组所关联的锁箱密钥加密文件块密钥. 如果文件所有者想将文件共享给他人，他只需要将文件组的锁箱密钥发送给共享用户即可. 由于密钥管理的复杂性和文件组的数量存在正比关系，所以系统规模很难扩展. 文献[15]提出了一个云环境中外包数据的安全存储与访问控制方案. 为了提高数据存储效率，将数据分块；为了保证数据块的安全，采用多个不同密钥对数据块加密. 该方案的不足是数据所有者要花费很大代价来进行数据加密和密钥管理. 文献[16]实现了采用 SiRiUS 系统，该系统建立在已有的像 NFS(Network File System)这样的文件系统之上，能确保端对端的传输安全. 为了进行访问控制，每个文件都被分配了一个元文件，元文件包含着一个访问控制列表，列表的每一项存放的是受授权用户公钥加密保护的文件加密密钥. 该系统的扩展版本没有采用授权用户的公钥加密文件加密密钥而是采用 NNL(Naor-Naor-Lotspiech)广播加密算法加密文件加密密钥. NNL 中撤销用户权限算法的复杂性与撤销用户数的正比关系使得该系统的复杂性同元文件的大小及加密的负荷也成正比关系，因而该系统难以扩展.

(4) 支持查询的云数据加密存储技术

在数据上传到云数据中心存储前将其加密能确保数据的安全性和隐私性，但这带来一个问题——数据查询变得困难. 如前面所述，如果数据所有者不将数据加密密钥告诉查询用户，用户很难进行数据查询；但如果告知密钥，数据的安全性又得不到保障. 另外，在采用一般的加密方法加密的情况下，索引往往无法建立，这样就导致高效查询无法进行. 针对这些问题，支持查询的数据加密方法[17](Searchable Encryption，SE)被提了出来. 使用 SE，用户将查询关键字或查询条件提交给云中的查询服务器，查询服务器通过检索关键字索引找到符合条件的数据，然后将查询结果返回给用户. 在整个查询过程中，加密的数据不会被解密，查询用户也无需知道数据的加密密钥或解密密钥. SE 要求输入的查询关键字不能有任何错误而且格式必须遵循规定的统一格式. 除此之外，SE 只支持“布尔型”查询，不支持按相关性有排名的查询，这使得 SE 直接应用云存储环境

下的文件查询会面临两大问题：① 用户要将返回的所有文件都一一打开才能确定文件与查询关键字的相关程度；② 返回所有的文件(可能绝大多数都不是用户需要的)将大大增加网络流量，进而增加用户的经济负担.

针对 SE 在查询时要求输入的关键字必须准确且格式必须符合规定这一问题，文献[18]提出了面向云密文数据的模糊查询方法. 该方法不要求输入的查询关键字非常准确也没有苛刻的格式要求，查询先按精确模式进行，在精确匹配失败的情况下将改为模糊模式，在该模式下会将查询关键字集合中与查询关键字最相似的关键字作为查询关键字进行查询. 由于查询是将关键字和事先生成的模糊关键字集合进行匹配，密文文件无需解密，文件的安全性得到保证. 该方法同样只支持“布尔型”查询. 为实现按相关性有排名的查询，文献[19-20]提出了面向云数据的支持排名查询的加密方法. 他们将“相关度”引入到查询索引表的构建之中以实现对排名查询的支持，并利用“一对多有序映射”技术保护敏感数据的“相关度”以确保文件的安全. 该方法的不足是只支持单个关键字的查询；另外，关键数据“相关度”的量化模型是否正确还有待验证. 针对多关键字有排名查询问题，文献[21]采用坐标匹配原则(尽可能多的匹配)来计算多个查询关键字与数据文件的相关性，进而实现多关键字的有排名查询. 在查询的过程中，该方法会遍历整个索引表，因而计算代价较大.

(5) 基于可信平台的数据安全存储技术

考虑到硬件和软件的不可信也是造成云数据存储面临安全挑战的重要原因，文献[22]借助可信计算技术作为硬件上的可信计算基础，借助虚拟机监控器作为软件上的可信计算基础，提出了一种可靠的数据保护与销毁的途径. 可信的虚拟机监控器负责保护用户的敏感数据，并按照用户命令对数据进行彻底销毁，即使云服务器的特权管理员也无法绕过保护机制得到受保护的敏感数据. 该方案要求计算机从硬件到软件都是可信的，显然现实情况无法满足该条件. 文献[23]针对使用数据保护提出了基于二次混淆的隐式分割机制；针对用户身份信息的保护提出了基于可信服务器的云存储架构，实现数据存储和用户个人信息管理隔离. 云服务器利用可信服务器提供的存储认证码判断用户的存储权限，用户的身份信息存储于可信服务器. 该方案在使用数据存储时进行的二次分割和矩阵运算使存储效率低，难以扩展；为保护隐私数据而引入的第三方可能成为数据存储的瓶颈.

表 2 对以上 5 种云数据安全存储技术进行归纳对比. 5 种技术中安全性和功能性最好的是同态加密技术，但该技术还在理论探究之中. 已经提出的秘密同态算法密文处理效率十分低，还无法满足实际的应用需求. 加密位置是一个不容忽视的问题：如果加密位置在客户端，数据的安全性能得到保证，但客户端的计算负荷就会大大加重，而云平台强大的计算能力却没有得到充分利用；如果加密位置在云端，云平台的计算能力能够得到充分发挥，但用户数据却可能泄漏给云服务提供商. 为此，提出两重加密方法. 第一重加密由用户自己完成，为了减少加密负担，可以考虑采用轻量级的加密方法对文件进行加密处理，例如将文件进行秘密分割并生成文件重构文件，将分割后的文件分块和加了密的文件重构文件上传到云端进行强度较大的二次加密. 针对大数据，云端可充分利用 MapReduce 并行编程模型实现海量数据的快速加密.

表 2 云数据安全存储技术安全性比较

技术名称	技术特点	运算支持能力	加密位置	传输安全	内存安全	外存安全	存在的主要问题
同态加密	明文上执行的代数运算结果等同于在密文上的另一个代数运算结果	支持全部运算	客户端	完全解决	完全解决	完全解决	密文处理效率低
基于 VMM 的数据保护技术	操作系统和文件系统只能看到密文	不支持	VMM	部分解决	部分解决	部分解决	特权用户可以解密用户数据；VMM 负荷加重
基于加解密的数据存储技术	采用传统的加密技术	不支持	客户端	完全解决	未解决	部分解决	安全机制复杂且有安全隐患，时空代价太大
支持查询的加密存储技术	加密算法支持密文查询	仅支持查询	客户端	完全解决	完全解决	完全解决	不能支持加减乘除等基本运算
基于可信平台安全存储技术	硬件和软件都可信	支持全部运算	VMM	部分解决	部分解决	部分解决	特权用户可解密用户数据；可信条件难以满足；VMM 负荷加重

3.2 云数据安全审计研究

如前所述，云数据安全审计面临两个难题. 实际上，第 1 个难题是数据持有问题，第 2 问题是数据的完整性保护问题. 云数据安全审计的重难点是数据

的公开审计(或第三方审计)，而一个理想的公开审计方案应具有这样几个特性：额外增加的时空代价小，隐私不会泄露，支持数据的动态变化(即支持数据追加、插入、修改、删除等基本操作)和支持批量审计.

3.2.1　数据持有审计模型

为保证文件在非可信存储系统上存储的安全，Ateniese 等人[24]构建了一个可证数据持有模型 PDP(Provable Data Possession). 在该方案中，使用基于 RSA 的同态标签审计外包数据，以此实现数据的可公开验证性. 然而，由于没有考虑数据的动态存储，该方案要由支持静态存储扩展成支持动态存储，在设计上和安全上都还存在着很多问题. 在随后进行的相关研究中，他们又提出了一个支持动态存储的外包数据存储模型(Scalable PDP)[25]. 该模型只使用到了对称密码机制，显著降低了系统的计算负荷. 该模型支持数据块的更新、删除和追加. 然而，该方案面向的是单服务器环境，在服务器出现故障时，数据就无法使用；该模型还需要对查询次数进行预设，并且不是对所有的动态操作都予以支持，例如插入操作，它就不支持. Erway 等人[26]对可证动态数据持有机制进行了研究，他们扩展了可证数据持有模型. 扩展后的模型 DPDP(Dynamic PDP)利用基于等级的认证跳表，支持存储数据文件的更新，而这个更新过程是可证安全的. 从本质上看，扩展后的模型实际上是可证数据持有模型的一个全动态版本，即它支持所有的动态操作. 为支持更新特别是数据块插入造成的更新，他们试图在标签计算中删除 Ateniese 可证数据持有模型中的索引信息. 为此，在验证之前，要先利用认证跳表结构认证待查数据块或更新块的标签信息. 该模型的不足是执行效率不高. 3 种模型的特点如表 3 所示.

表 3　3 种证据持有审计模型的比较

模型	关键技术	隐私性	动态性	服务器额外计算复杂度	客户端额外计算复杂度	额外通讯复杂度	服务器额外存储复杂度	客户端额外存储复杂度
PDP	基于 RSA 同态标签	支持	不支持	$O(1)$	$O(1)$	$O(1)$	$O(n)$	$O(1)$
Scalable PDP	对称密码机制	支持	不支持插入	$O(1)$	$O(1)$	$O(1)$	$O(n)$	$O(1)$
DPDP	基于等级的认证跳表	支持	支持	$O(\log n)$	$O(\log n)$	$O(\log n)$	$O(n)$	$O(1)$

3.2.2　数据完整性审计模型

Juels 等人[27]提出了一个数据可检索证据模型 POR(Proofs of Retrievability)并给出了严格证明. 在该模型中，采用抽样检查和纠错码来确保存储在存储系统中的数据文件的安全性和可恢复性. 为便于进行检查，将一些称作“哨兵”的特殊数据块随机地嵌入到数据文件之中；为防止暴露这些特殊数据块的位置，将数据文件进行加密存储. 该方案要求查询次数提前预设固定，为“哨兵”进行的预计算使得该方案不支持数据的动态更新，该方案还不支持数据的公开或第三方验证. Shacham 等人[28]采用完全安全证据对数据可检索证据模型进行了改进. 他们使用了公开可验证同态认证符，该认证符采用双线性签名方法 BLS 构建，并且在随机模型中是可证安全的. 基于 BLS 构建方法，数据公开恢复得以实现，多个证据能聚集成一个小的认证值. 该方案的不足是仅仅适用于静态数据文件. Bowers 等人[29]在 Juels 和 Shacham 研究成果的基础上提出了一个数据可检索证据模型改进方案. 在随后的工作中，他们将数据可检索证据模型推广到分布式环境. 然而，他们提出的方案关注的都是静态数据；方案的效率主要由数据文件外包前数据所有者自己进行的预处理过程的效率决定；数据文件的任何变化哪怕是几位的改变都会影响存储的纠删码数据和相应的随机混淆过程，带来巨大的计算和通讯代价.

Wang 等人[30]利用同态令牌和纠删码编码方法提出了一种云存储完整性审计机制. 基于该机制的审计结果不仅可以用来确保数据存储的正确性，还能快速实现数据的定位，即快速确定出现错误数据的服务器. 该机制采用纠删码和冗余存储方法保证数据的可用性，采用同态令牌保证数据完整性. 该机制存在和 POR 一样的不足：计算代价非常大；除此之外，审计次数有限且要预设. 他们在后续研究中指出了云数据公开审计需要关注的问题及可能的解决办法，并提出了一个较为抽象的公开审计机制[31]. 随后，他们将基于公钥的同态认证方法和随机隐藏方法结合起来，提出了具有隐私保护功能的云数据公开审计方案[32-33]，如图 6 所示. 该方案具有 3 个特点：第三方在数据完整性审计时无需获取数据本身；不要求数据所有者随时在线；支持批量审计. 该方法

的不足是过于复杂.针对数据动态更新问题,Wang等人[34]对完整性机制进行了改进,通过将令牌数据同步存储到第三方来实现对第三方审计数据完整性的支持;通过采用 Merkle Hash Tree(MHT)结构存储数据块标志和进行文件组织来实现对数据块插入操作的支持.改进后的方法使基本的文件操作可以正常进行,但不支持批量审计.

典型的外包数据完整性审计模型如表4所示.

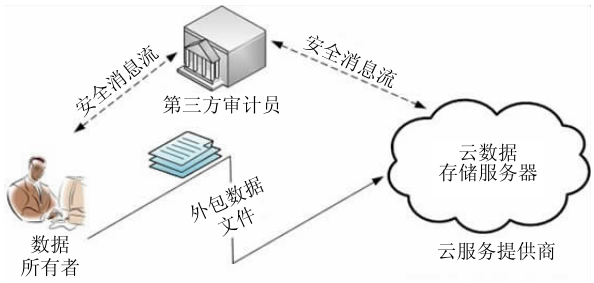


图6 公有云存储环境中数据公开审计模型[33]

表4 4种典型的数据完整性审计模型比较

文献	关键技术	公开审计	批量审计	可恢复性	隐私性	动态性	服务器额外计算复杂度	审计端额外计算复杂度	额外通讯复杂度	服务器额外存储复杂度	审计端额外存储复杂度
[34]	纠删码	不支持	不支持	支持	—	不支持	$O(1)$	$O(1)$	$O(1)$	$O(n\times(m+k)/m)$	$O(t)$
[37]	同态令牌	不支持	不支持	支持	—	部分支持	$O(1)$	$O(1)$	$O(1)$	$O(n\times(m+k)/m)$	$O(t)$
[40]	基于公钥同态认证	支持	支持	不支持	支持	支持	$O(\log n)$	$O(\log n)$	$O(\log n)$	$O(n)$	$O(1)$
[41]	MHT	支持	不支持	不支持	支持	部分支持	$O(\log n)$	$O(\log n)$	$O(\log n)$	$O(n)$	$O(1)$

注:t 为预设的审计次数,对应的纠删码为(m,k)编码.

国内对云数据的安全审计也比较重视.文献[35]基于认证数据结构提出了一种外包数据认证模型,如图7所示.在该模型中,数据集存储在云中的数据外包服务器上,客户仅存储动态结构化数据集的正确性根据.任何时候客户都可以向数据外包服务器发起某一数据的查询验证,通过外包服务器返回的证据和先前计算的正确性根据进行比对就能判断数据是否被修改.该方案的不足是不支持公开审计;动态存储上仅仅支持动态更新.文献[36]提出了一个文本数据的完整性检测方案.在该方案中,数据所有者通过在词、段和篇3个粒度上获取指纹来实现外包数据的完整性检测.该方案不仅需要消耗大量时间来计算指纹,还需要大量空间来存储指纹数据.另外,该方案针对的是文本数据,不支持公开审计.文献[37]提出了一种面向外包数据库的数据完整性检测方法:签名链方法.该方法将验证对象嵌入在外包数据库内部,因而查询验证时无需对 DBMS 作任何功能扩展.文献[38]提出了一种面向外包数据库的基于掩码认证 B 树的完整性检测方法.但这些方法针对的都是结构化数据而且都不支持第三方审计.

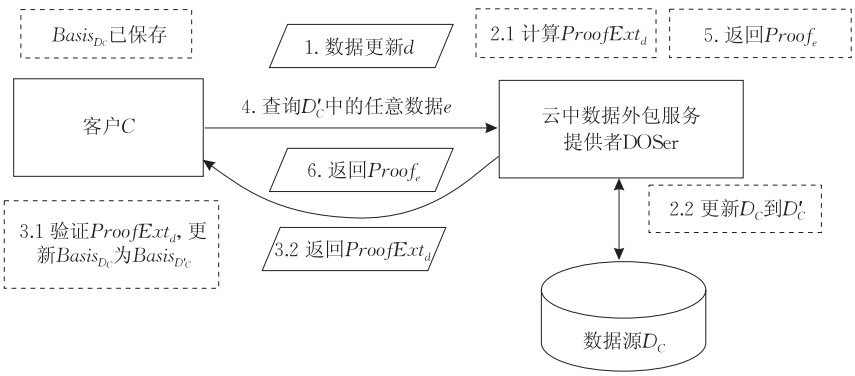


图7 基于认证数据结构的外包数据认证模型[35]

3.3 云密文访问控制研究

为了保证外包到云中存储数据的安全,用户在将数据传输到云端之前会将数据加密.存储位置和处理位置的分置使得密文的访问控制就较为困难.在学术界,密文数据的细粒度访问控制特别是基于属性的密文数据的访问控制比较受关注.

3.3.1 基于属性的密文访问控制

基于属性的密码体制研究始于2005年,其发展了传统的基于身份密码体制关于身份概念,将身份看作是一系列属性的集合.2006年,Sahai 和 Waters[39]首先提出了基于属性的加密方法 ABE(Attribute-based Encryption),其基本思想是密文与私钥分别与一组属性关联,当用户的私钥属性与密文属性相互匹配到达一个门限值时,该用户才能解密密文.2006年,Goyal 等人[40]基于模糊身份加密方案提出了密钥策略的基于属性的加密方案 KP-ABE.该方

案将密钥采用秘密共享的方法隐藏在访问控制结构中. 2007 年，Bethencourt 等人[41]提出了一种密文策略的基于属性的访问控制方法 CP-ABE，该方法使用访问控制结构加密明文. 密文策略的访问控制加密方法包括 4 个步骤：① 初始化 Setup. 生成主密钥 MK 和公开参数 PK；② 加密 $CT=Encrypt(PK,M,T)$：使用 PK 和访问结构 T 加密数据明文 M，加密后的密文为 CT；③ 生成用户私钥 $SK=KeyGen(MK,A)$：使用 MK 和用户属性集 A 生成用户的私钥 SK；④ 解密 $M=Decrypt(CT,SK)$：使用私钥 SK 解密密文 CT 得到明文 M. 对 T 中的每个节点 i，设 k_i 是节点的门限值，选择一个 k_i-1 阶多项式 $q_i(x)$. 随机选择 $s\in Z_q^*$，$q_i(x)$ 的选择满足下述条件：①对根节点 R，$q_R(0)=s$；②对非根节点 i，$q_i(0)=q_{parent(i)}(index(i))$，$index(i)$ 含义为 i 在其兄弟节点中的序号. 对于多项式：$q_i(x)=a_{k_i-1}x^{k_i-1}+a_{k_i-2}x^{k_i-2}+\cdots+a_1x+a_0$，如果有 k_i 个有序数对 $(x_j,q_i(x_j))(j=1,2,\cdots,k_i)$ 满足该多项式，那么根据拉格朗日插值公式由这 k_i 个有序数对就可以唯一地确定 k_i-1 次多项式 $q_i(x)$. 对于节点 i，如果能得知其 k_i 个孩子的值，就能求出其对应的多项式，进而求出 $q_i(0)$. 依照此法，由下至上，就能求出 $q_R(0)=s$，密文也就可在这一个过程被解密.

3.3.2　云密文的访问控制

KP-ABE 和 CP-ABE 算法通过将解密规则蕴含在加密算法之中来避免频繁分发密钥. 但在访问控制策略发生变更时，该方法要求数据所有者对数据重新加密. 由于 ABE 算法的效率较低，重加密的代价可能很大，让数据所有者难以接受. 如何有效地支持动态策略，已成为 ABE 面临的主要难题.

为解决以上问题，Yu 等人[42]将基于属性的访问控制方法、代理重加密方法和懒惰重加密方法结合起来提出了一种面向云存储环境的密钥策略的基于属性的访问控制方案，如图 8 所示. 在该方案中，将文件加密密钥采用秘密共享方法存储在树形访问控制结构中；访问控制结构由授权用户保存，而将文件密文保存在云存储中心，服务器负责访问控制. 非授权用户因没有访问控制结构和用户私钥而无法获得加密的文件；服务器因没有访问控制结构和哑私钥也无法解密文件. 代理重加密使得密钥密文的重加密由云来完成而数据又不会泄露；懒惰加密方法则提高了重加密的效率. 由于密钥分配和数据管理主要由云来完成，减轻了数据所有者的负荷；基于属性的加密方法则实现了细粒度的访问控制. 但该方法仍未解决数据重加密代价太大的问题；另外，该方法没有提及 ABE 随机参数的更换，而随机参数对安全有重大影响. 表 5 从访问控制过程、重加密位置等几个方面对 CP-ABE，KP-ABE 和 Yu 的方案进行了对比.

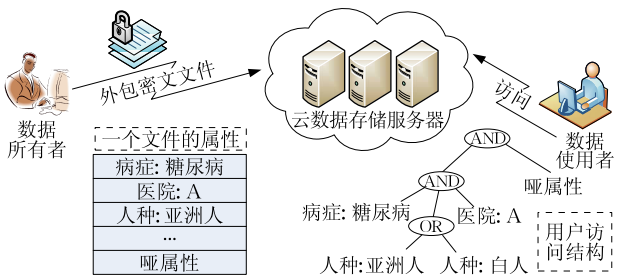


图 8　密钥策略的基于属性的云访问控制模型[42]

表 5　3 种密文访问控制方案比较

方案	访问控制过程				重加密位置	代理重加密	懒惰加密
	初始化	加密	生成用户私钥	解密			
CP-ABE	生成主密钥 MK 和公开参数 PK	$CT=Encrypt(PK,M,T)$	$SK=KeyGen(MK,A)$	$M=Decrypt(CT,SK)$	客户端	不支持	不支持
KP-ABE	生成主密钥 MK 和公开参数 PK	$CT=Encrypt(PK,M,A)$	$SK=KeyGen(MK,T)$	$M=Decrypt(CT,SK)$	客户端	不支持	不支持
Yu 的方案	生成主密钥 MK 和公开参数 PK	$CT=Encrypt(PK,M,A)$；加密算法法支持代理重加密和懒惰加密	$SK=KeyGen(MK,T)$；传送除“dummy”属性外的所有属性的私钥到云端服务器	$M=Decrypt(CT,SK)$	云端	支持	支持

文献[43]提出了一种支持代理重加密的 CP-ABE 算法：CP-ABPRE，但使用 CP-ABPRE 生成代理重加密密钥的代价等同于一次 CP-ABE 加密；而且每次代理重加密都会增大密文体积，不适合频繁地进行重加密. 文献[44]将层次型的基于身份的加密方法和密文策略基于属性的访问控制方法结合起来提出了一种混合的云数据访问控制方案. 该方案将访问控制策略分成两部分：授权用户的身份信息集合和基于属性的访问控制策略，它使用授权用户的身份信息集加密数据并基于属性来进行访问控制，只有身份信息在身份信息集中或其属性满足访问控制策略要求的用户才能获取并解密数据. 该方

案能实现细粒度的访问控制并支持由云存储服务器来负责密钥管理.该方案的不足是没有说明如何实现用户权限的撤销并且针对的是静态数据.文献[45]针对 CP-ABE 方法的不足,对其进行了改进:访问控制结构设计为二叉树,只有 and 和 or 两种关系;秘密共享方式由复杂的多项式运算变成了简单加减运算;密钥密文的部分重加密工作由用户端转移至云端.该改进一定程度上降低了用户端运算负载,加密效率有所提高.但该方法只转移了部分密钥重加密工作,仍然要求用户时刻在线,数据加密依然在用户端进行,没有充分发挥云的计算优势.文献[46]等通过引入密钥分割技术和代理重加密技术对 CP-ABE 方法进行了改进,提出了一种基于 CP-ABE 的密文控制方案.该方案在用户权限撤销时将部分重加密工作转移给云服务器执行,降低了数据所有者的计算代价.该方案同样没有解决数据重加密代价太大的问题.文献[47]提出了一种面向云环境的 CP-ABE 算法.采用数据密钥 k_d 加密数据,使用 RSA 密钥对 K_{sign}/K_{verify} 对加密后的数据进行签名/验证,只读用户应该持有 K_d 和 K_{verify},而读写用户拥有所有 3 个密钥,采用 CP-ABE 算法将数据加密密钥加密.该方法在访问权限只有读写的情况下较简单,而一旦权限类型增多,密钥相应增多,访问控制就会变得非常复杂.

4 问题及未来研究方向

从已有的研究看,云数据的安全存储研究当前还面临诸多问题,主要问题有:

(1)同态加密技术数据处理效率太低.同态加密技术是解决云数据安全存储悖论最有效的方法之一,然而,该技术采用的加密方法和公钥加密方法一样都需要进行大量复杂的指数运算,大大降低了数据的处理效率.目前同态加密技术水平尚不能支持对较多数据的快速处理.

(2)基于虚拟监控器数据保护技术可能带来新的问题.基于监控监控器的数据保护技术能有效防止一个用户的内存和外存数据外泄给其他用户,但却不能防止控制着虚拟监控器的云服务提供商获得数据.另外,在虚拟层增加加解密和相关的其他功能,会增加虚拟层的复杂性和出现漏洞概率.一旦虚拟监控器本身被攻破,所有用户的数据,无论是否加密,都会泄露.

(3)基于加解密的数据存储技术规模难以扩展.从目前的技术条件看,加密是确保云中所存储数据的机密性和隐私性的主要方法.然而,已有的基于加密方式的云存储框架和技术都过于复杂,其复杂性和用户数呈线性关系,极大地限制了这些方法的可扩展性和规模.

(4)可信计算技术尚不成熟.可信计算技术作为一种力图从根本上解决计算机和网络安全问题的重要技术,理所当然地也可以用来确保云存储的安全性,然而,可信技术本身并不成熟,还有很多问题亟待解决.

云数据的安全存储面临很多问题,这些问题已成为当今的研究热点.从研究的发展趋势看,在云数据的安全存储上,未来的研究关注主要包括:

(1)隐私数据的安全存储.一直以来,隐私性和和机密性被混为一谈,因此,确保隐私性和确保机密性都采用了一样的方法——加密,而事实上,隐私是一种属于私人的排他性的不愿为他人知晓或干涉的信息,包括身份信息,行为模式等,这就决定了隐私性和机密性并不一样.加密并不是确保隐私性的唯一方法,机密性数据和隐私数据应该有着不同的安全存储技术.

(2)海量数据的安全存储.能解决企业存储能力不足是云计算的重要优势之一,这意味着云数据中心应支持海量数据的安全存储.显然,对于海量数据,加密存储并不是一种高效的方式.

(3)同态加密技术.作为一种堪称最能彻底解决云存储安全悖论的技术,同态加密技术的数据处理速度和效率亟待提高.

(4)支持大量用户的加密存储技术.当前最可行最有效的云数据安全存储技术是基于加解密技术的,该技术面临的紧迫任务是降低系统的复杂性,将系统复杂性和用户数量解耦,使系统具有可扩展性和规模性.

(5)公开审计时数据安全性和隐私性的确保.数据在公开审计时,不可避免地需要提供部分与数据有关的信息,这可能威胁到数据本身的安全性和隐私性.另外,用户存储在云中的数据往往是动态变化的,这种动态性会大大增加数据公开审计的难度.

(6)高效的代理重加密技术.某个用户权限撤销时,为安全起见,存储在云中的数据需要重新加密.如果采用将密文数据下载到数据所有者处加密再上传方式,效率会非常低,也没有充分发挥云计算的优势.解决这个问题的有效方式是代理重加密,但目前的代理重加密技术加密效率还很低,无法处理

大量数据.

5　结束语

如何确保存储在云中的数据的安全性和隐私性是云计算面临的难题，它已严重阻碍了云计算的推广和发展.该文首先列举了云计算在数据安全上面临的主要挑战，包括加密存储、数据隔离、数据迁移、安全审计和数据残留，指出云计算的租用商业模式和其采用的两种关键技术——虚拟化技术和多租户技术是云存储存在诸多安全问题甚至安全悖论的根本原因.鉴于已有的云数据安全存储相关研究主要围绕加密存储、完整性安全审计和密文的访问控制3个技术领域展开，对这3个领域的最新研究进展分别进行了深入分析和评述.在加密存储上，介绍了云数据安全存储框架和主要的安全存储技术；在安全审计上，分析了外包数据安全审计特别是公开审计面临的主要难题，介绍了包括云数据在内的外包数据完整性公开证明的主要模型和方法，并指出了它们的优势和不足；在云密文的访问控制上，详述了基于属性的云密文访问控制方法，并指出了这些方法的优劣.最后归纳了云数据安全存储研究面临的主要问题并预测了相关研究的未来发展趋势.

参　考　文　献

[1] Mell P, Grance T. The NIST definition of cloud computing. National Institute of Standards and Technology (NIST), Washington, USA: Technical Report Special Publication 800-145, 2011

[2] Feng Deng-Guo, Zhang Min, Zhang Yan. Study on cloud computing security. Journal of Software, 2011, 22(1): 71-83(in Chinese)
(冯登国，张敏，张妍. 云计算安全研究. 软件学报，2011，22(1)：71-83)

[3] Kaufman L M. Data Security in the world of cloud computing. IEEE Security & Privacy, 2009, 7(4): 61-64

[4] Takabi H, Joshi J B D, Ahn G. Security and privacy challenges in cloud computing environments. IEEE Security & Privacy, 2010, 8(6): 24-31

[5] Ren Kui, Wang Cong, Wang Qian. Security challenges for the public cloud. IEEE Internet Computing, 2012, 16(1): 69-73

[6] IBM. Virtualization and Cloud Computing. Beijing: Publishing House of Electronics Industry, 2009(in Chinese)
(IBM. 虚拟化与云计算. 北京：电子工业出版社，2009)

[7] Kamara S, Lauter K. Cryptographic cloud storage//Proceedings of the 14th International Conference on Financial Cryptography and Data Security. Berlin, Germany, 2010: 136-149

[8] Bian Gen-Qing, Gao Song, Shao Bi-Lin. Security structure of cloud storage based on dispersal. Journal of Xi'an Jiaotong University, 2011, 45(4): 41-45(in Chinese)
(边根庆，高松，邵必林. 面向分散式存储的云存储安全架构. 西安交通大学学报，2011，45(4)：41-45)

[9] Gentry C. Fully homomorphic encryption using ideal lattices //Proceedings of the 41st ACM Symposium on Theory of Computing. New York, USA, 2009: 169-178

[10] Huang Ru-Wei, Gui Xiao-Lin, Yu Si, Zhuang Wei. Privacy-preserving computable encryption scheme of cloud computing. Chinese Journal of Computers, 2011, 34(12): 2391-2402(in Chinese)
(黄汝维，桂小林，余思，庄威. 云环境中支持隐私保护的可计算加密方法. 计算机学报，2011，34(12)：2391-2402)

[11] Hou Qing-Hua, Wu Yong-Wei, Zheng Wei-Min. A method on protection of user data privacy in cloud storage platform. Journal of Computer Research and Development, 2011, 48(7): 1146-1154(in Chinese)
(侯清铧，武永卫，郑纬民. 一种保护云存储平台上用户数据私密性的方法. 计算机研究与发展，2011，48(7)：1146-1154)

[12] Ateniese G, Fu K, Green M, Hohenberger S. Improved proxy re-encryption schemes with applications to secure distributed storage. ACM Transactions on Information and System Security(TISSEC), 2006, 9(1): 1-30

[13] Vimercati S, Foresti S, Jajodia S, et al. Over-encryption: Management of access control evolution on outsourced data// Proceedings of the 33rd International Conference on Very Large Data Bases. Vienna, Austria, 2007: 23-134

[14] Kallahalla M, Riedel E, Swaminathan R, et al. Plutus: Scalable secure file sharing on untrusted storage//Proceedings of the 2nd Conference on File and Storage Technologies. San Francisco, USA, 2003: 29-42

[15] Wang W, Li Z, Owens R, et al. Secure and efficient access to outsourced data//Proceedings of the ACM Workshop on Cloud Computing Security. Chicago, USA, 2009: 55-66

[16] Goh E, Shacham H, Modadugu N, Boneh D. Sirius: Securing remote untrusted storage//Proceedings of the Internet Society Network and Distributed Systems Security. San Diego, USA, 2003: 131-145

[17] Song D, Wagner D, Perrig A. Practical techniques for searches on encrypted data//Proceedings of the 2000 IEEE Symposium on Security and Privacy. Berkeley, USA, 2000: 44-55

[18] Li Jin, Wang Qian, Wang Cong, et al. Fuzzy keyword search over encrypted data in cloud computing//Proceedings of the INFOCOM 2010 Mini-Conference. San Diego, USA, 2010: 1-5

[19] Wang Cong, Cao Ning, Li Jin, et al. Secure ranked keyword search over encrypted cloud data//Proceedings of the 30th International Conference on Distributed Computing Systems. Genoa, Italy, 2010: 253-262

[20] Wang Cong, Cao Ning, Ren Kui, et al. Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(8): 1467-1479

[21] Cao Ning, Wang Cong, Li Ming, et al. Privacy-preserving multi-keyword ranked search over encrypted cloud data// Proceedings of the IEEE INFOCOM 2011. Shanghai, China, 2011: 829-837

[22] Zhang Feng-Zhe, Chen Jin, Chen Hai-Bo. Lifetime privacy and self-destruction of data in the cloud. Journal of Computer Research and Development, 2011, 48(7): 1155-1167 (in Chinese)
(张逢喆, 陈进, 陈海波. 云计算中的数据隐私性保护与自我销毁. 计算机研究与发展, 2011, 48(7): 1155-1167)

[23] Mao Jian, Li Kun, Xu Xian-Dong. Privacy protection scheme for cloud computing. Journal of Tsinghua University (Science & Technology), 2011, 51(10): 1357-1362(in Chinese)
(毛剑, 李坤, 徐先栋. 云计算环境下隐私保护方案. 清华大学学报(自然科学版), 2011, 51(10): 1357-1362)

[24] Ateniese G, Burns R, Curtmola R, et al. Provable data possession at untrusted stores//Proceedings of the 14th ACM Conference on Computer and Communications Security. Alexandria, USA, 2007: 28-31

[25] Ateniese G, Pietro R D, Mancini L V, et al. Scalable and efficient provable data possession//Proceedings of the 4th ACM Conference on Security and Privacy in Communication Networks. Istanbul, Turkey, 2008: 1-10

[26] Erway C, Kupcu A, Papamanthou C, et al. Dynamic provable data possession//Proceedings of the 16th ACM Conference on Computer and Communications Security. Chicago, USA, 2009: 213-222

[27] Juels A, Burton J, Kaliski S. PORs: Proofs of retrievability for large files//Proceedings of the 14th ACM Conference on Computer and Communications Security. Alexandria, USA, 2007: 584-597

[28] Shacham H, Waters B. Compact proofs of retrievability. Journal of Cryptology, 2008, 26(3): 90-107

[29] Bowers K D, Juels A, Oprea A. HAIL: A high-availability and integrity layer for cloud storage//Proceedings of the 16th ACM Conference on Computer and Communications Security. Chicago, USA, 2009: 187-198

[30] Wang Cong, Wang Qian, Ren Kui, et al. Towards secure and dependable storage services in cloud computing. IEEE Transactions on Service Computing, 2012, 5(2): 220-232

[31] Wang Cong, Ren Kui, Lou Wen-Jing, et al. Toward publicly auditable secure cloud data storage services. IEEE Network, 2010, 24(4): 19-24

[32] Wang Cong, Chow S, Wang Qian, et al. Privacy-preserving public auditing for cloud storage security in cloud computing //Proceedings of the IEEE INFORCOM 2010. San Diego, USA, 2010: 1-9

[33] Wang Cong, Chow S, Wang Qian, et al. Privacy-preserving public auditing for secure cloud storage. IEEE Transactions on Computers, 2013, 62(2): 362-375

[34] Wang Qian, Wang Cong, Li Jin, et al. Enabling public verifiability and data dynamics for storage security in cloud computing. IEEE Transactions on Parallel and Distributed Systems, 2011, 22(5): 847-859

[35] Xu Jian, Zhou Fu-Cai, Chen Xu. Data outsourcing authentication model based on authenticated data structures for cloud computing. Journal of Communications, 2011, 32(7): 153-160(in Chinese)
(徐剑, 周福才, 陈旭. 云计算中基于认证数据结构的数据外包认证模型. 通信学报, 2011, 32(7): 153-160)

[36] Zhao Chun-Hong, Liu Guo-Hua, Wang Ning, He Ling-Ling. Text data Integrity detection scheme in outsourced database model. Journal of Chinese Computer System, 2010, 31(9): 1790-1796(in Chinese)
(赵春红, 刘国华, 王柠, 何玲玲. 外包数据库模型中文本数据的完整性检测方案. 小型微型计算机系统, 2010, 31(9): 1790-1796)

[37] Zhang Min, Hong Cheng, Chen Chi. Server transparent query authentication of outsourced database. Journal of Computer Research and Development, 2010, 47(1): 182-190 (in Chinese)
(张敏, 洪澄, 陈驰. 一种服务器透明的外包数据库查询验证方法. 计算机研究与发展, 2010, 47(1): 182-190)

[38] Xian He-Qun, Feng Deng-Guo. An integrity checking scheme in outsourced database model. Journal of Computer Research and Development, 2010, 47(6): 1107-1115 (in Chinese)
(咸鹤群, 冯登国. 外包数据库模型中的完整性检测方案. 计算机研究与发展, 2010, 47(6): 1107-1115)

[39] Sahai A, Waters B. Fuzzy identity-based encryption// Proceedings of the EUROCRYPT 2005. Berlin, Germany, 2005: 457-473

[40] Goyal V, Pandey O, Sahai A, et al. Attribute based encryption for fine-grained access control of encryption security data//Proceedings of the ACM Conference on Computer and Communications Security. Alexandria, USA, 2006: 89-98

[41] Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption//Proceedings of the 2007 IEEE Symposium on Security and Privacy. Berkeley, USA, 2007: 321-334

[42] Yu Shu-Cheng, Wang Cong, Ren Kui, Lou Wen-Jing. Achieving secure, scalable, and fine-grained data access control in cloud computing//Proceedings of the IEEE INFORCOM 2010. San Diego, USA, 2010: 1-9

[43] Luan I, Muhammad A. An encryption scheme for a secure policy updating//Proceedings of the International Conference on Security and Cryptography (SECRYPT 2010). Athens, Greece, 2010: 1-10

[44] Wang Guo-Jun, Liu Qin, Wu Jie. Achieving fine-grained access control for secure data sharing on cloud servers. Concurrency and Computation: Practice and Experience, 2011, 23(12):

1443-1464

[45] Hong Cheng, Zhang Min, Feng Deng-Guo. Achieving efficient dynamic cryptographic access control in cloud storage. Journal of Communications, 2011, 32(7): 125-132(in Chinese)
(洪澄，张敏，冯登国. 面向云存储的高效动态密文访问控制方法. 通信学报，2011，32(7)：125-132)

[46] Lv Zhi-Quan, Zhang Min, Feng Deng-Guo. Cryptographic access control scheme for cloud storage. Journal of Frontiers of Computer Science and Technology, 2011, 5(9): 835-844(in Chinese)
(吕志泉，张敏，冯登国. 云存储密文访问控制方案. 计算机科学与探索，2011，5(9)：835-844)

[47] Sun Guo-Zi, Dong Yu, Li Yun. CP-ABE based data access control for cloud storage. Journal of Communications, 2011, 32(7): 146-152(in Chinese)
(孙国梓，董宇，李云. 基于 CP-ABE 算法的云存储数据访问控制. 通信学报，2011，32(7)：146-152)

FENG Chao-Sheng, born in 1971, Ph. D., professor, M. S. supervisor. His current research interests include cloud computing, privacy protection and data security.

QIN Zhi-Guang, born in 1956, Ph. D., professor, Ph. D. supervisor. His research interests include distributed computing and information security

YUAN Ding, born in 1967, Ph. D., professor. His research interests include data security and cryptography.

Background

Cloud computing is a model for providing convenient, on-demand network access to a shared centralized pool of configurable computing resources that can be rapidly deployed with great efficiency and minimal management overhead. The benefits brought by cloud computing model include but are not limited to: relief of the burden for IT management, universal data access with independent geographical locations, and reduction of capital expenditure on hardware, software, and maintenances, etc. Although the benefits are tremendous, security and privacy in 14 domains are the primary concern hindering its adoption. The 14 domains consist of architecture, risk management, contracts and electronic discovery, compliance and audit management, information management and data security, interoperability and portability, business continuity and disaster recovery, data center operations, incident response, application security, encryption and key management, access management, virtualization, and security as a service. If no appropriate security and privacy solutions are proposed, the potentially revolutionary computing paradigm could become a huge failure. In this paper, the security of data storage in cloud computing, which is involved in the fifth domain, i. e. information management and data security, is concerned and surveyed. The challenges that cloud computing is confronted with in terms of data storage, are listed. Further, the reasons resulting in these challenges are analyzed and identified. At last, the problems that the study of secure cloud storage is confronted with currently are identified and the study trend of secure cloud data storage is predicted.

This work is supported by National Science and Technology Major Project of the Ministry of Science and Technology of China under Grant No. 2011ZX03002-002-03, and the National Key Technology Support Program of China under Grant Nos. 2014BAH11F01 and 2014BAH11F02, and the National Natural Science Foundation of China under Grant No. 61373163. This work is also supported by the project of Visual Computing & Virtual Reality Key Laboratory of Sichuan Province under Grant No. PJ2012002.