doi:10.3969/j.issn.1000-436x.2014.12.020

云计算中数据隐私保护研究进展

肖人毅
（国家自然科学基金委员会，北京 100085）

摘 要：由于社会分工和资源共享的必然，公共云平台必将成为和电网、互联网等同等重要的国家基础设施。云计算面临的安全问题制约着云计算的广泛使用。数据安全在云计算中尤为重要，如何保证数据的安全性是云计算安全的核心。从数据的隐私保护计算、数据处理结果的完整性认证、数据访问权限控制以及数据的物理安全 4 个方面对已有研究工作进行了分类和总结，为后续云计算中数据的安全性研究提供参照。
关键词：云计算；数据安全；隐私保护；完整性认证；数据访问控制；数据备份
中图分类号：TP393.08 **文献标识码**：A **文章编号**：1000-436X(2014)12-0168-10

Survey of privacy preserving data queries in cloud computing

XIAO Ren-yi
(The National Natural Science Foundation of China, Beijing 100085, China)

Abstract: Driven by resource sharing, public clouds will become the national infrastructure like electricity grids and the Internet. A core issue in public cloud computing is privacy. It is crucial for public clouds to provide necessary services while protecting data privacy. Existing work on data privacy from the perspectives of privacy preserving computation was survied, integrity verification of public cloud computing results, public cloud data access control, and hardware security in cloud computing. Existing work is analyzed and future research directions for public cloud privacy is discussed.
Key words: cloud computing; data privacy; privacy preserving; integrity preserving; data access control

1 引言

云计算是继分布式计算、网格计算、对等计算后出现的一种崭新的计算模式，其核心思想是资源租用、应用托管和服务外包，希望为其他行业和个人提供便捷、经济和高可扩展性的 IT 服务平台，帮助企业和个人从繁重的 IT 基础建设、管理和维护工作中解放出来，集中精力发展自己的核心业务。

云计算模型代表了信息领域朝专业化、规模化和集约化的方向发展，是信息领域内一场新的革命，受到了各国政府和各大 IT 企业的高度重视。美国联邦首席信息委员会在 2010 年发布的《公共云计算态势》中要求所有 IT 投资必须完成基于云计算的替代方案分析，我国分别在 2010 年、2011 年颁布的《关于加快培育发展战略性新型产业的决定》和《国家“十二五”科学与技术发展规划》中都明确将云计算作为发展重点。Google、微软、IBM、Amazon 等公司都在大力推进和发展云计算。

然而，在云计算中数据拥有者、数据用户和云服务提供商分别处在不同的安全域，数据的安全问题是制约云计算发展的关键因素。2010 年，Google 解雇了 2 名入侵客户 Google Voice、Gtalk 等账户以获取隐私数据的员工，表明云计算服务提供商存在对数据拥有者敏感数据泄露的风险。2010 年 6 月，Apple 公司出现用户信息泄密事故[1]；2011 年 12 月，CSDN 网站 600 多万用户的数据库信息被盗取并公开，这一系列的安全事故加深了人们对云计算安全问题的忧虑。2011 年，国际知名非盈利研究机构——ITGI 对 21 个国家 10 个行业的 834 名首席执行官进行调查后的调查报告显示，49.6%的人对云数据的隐私性担忧，47.2%的人对云安全担忧。出于对数据安全和隐私方面的考虑，很多公司在控制云计算方面的投资或延缓云的部署[2]。

收稿日期：2014-09-19；修回日期：2014-10-20

因此，研究建立数据安全保障机制是云计算首要解决的问题之一。数据查询服务是云计算中需要提供的最重要的数据服务。在云计算环境下，数据拥有者失去了对数据存储与访问的物理控制权和直接控制权，如何建立一套完善的安全机制来保证数据的安全是一个极具挑战性的难题。其困难之处体现在以下 3 个方面：第一，为防止信息泄露，数据用户将数据存储到云服务器之前需要对数据进行加密，同样，数据用户向云服务器提交查询请求时，也需要对相应的请求条件进行加密，因此要解决云服务器在既不知数据真实值，又不知道查询条件真实值的情况下，如何进行数据的查询计算；第二，在各种利益的驱使下，云平台可能会伪造虚假的查询结果或者删除满足查询条件的一些数据，因此需要对查询结果完整性进行认证，以监测出云服务器的这种恶意行为；第三，为有效防止云平台对数据的非法使用以及其他用户对数据的非法使用，要解决数据拥有者将数据存储到云服务器后，依然能实现对数据访问权限的控制；第四，云服务器上存储了大量用户的数据，需要研究一套有效的机制来保证这些数据的物理安全。本文首先分别从云计算中数据的隐私保护、查询结果的完整性认证，数据访问权限控制以及数据的物理安全机制 4 个方面对已有工作进行了分类总结，旨在为后续研究工作提供参考。

2　数据隐私保护技术

为保护用户数据的隐私，用户在把数据交给云服务器前需要对数据进行加密，然后将密文数据提交给云服务器进行存储。随后当用户对其数据进行查询时，用户也需要将查询条件加密，否则查询条件将暴露数据信息。这就需要云服务器能够在加密的数据上根据加密的查询条件进行查询处理。下面分别从基于不经意随机访问存储器（ORAM）的隐私保护、基于对称加密的隐私保护方法、基于公钥体制的隐私保护方法以及文档的排名查询和模糊查询等方面来阐述国内外的研究现状。

2.1　ORAM 下的相关研究

加密数据的查询问题可以采用不经意随机访问内存（ORAM, oblivious RAM）来实现。所谓的不经意（oblivious）内存访问是指对存储器的访问不暴露任何查询信息，即对任意 2 个不同的输入所需要的处理时间相同，则处理过程中对内存的访问序列相同。ORAM 最开始研究是为了进行软件版权保护和防止代码反向工程[3~6]。在云计算中，不同数据的访问频度也泄漏了大量的数据信息，因此研究者基于 ORAM 上来实现可搜索对称加密问题上进行了大量研究[7~17]。文献[7]提出了一种基于二叉树结构的不经意数据存储方案，在进行数据读取时，每一次都是读取一条从根节点到叶子节点路径上的所有节点数据，利用二叉树的内部节点可以同时处在不同路径上的特点，在读取后重新对处在该路径上的数据进行重新分布，并且可以将一些数据修改到其他路径上，使服务器无从知道数据的访问频度。文献[12]将加密数据存储在 Mainpart 和 Shelter 2 部分，并且在客户端建立指向 Shelter 部分的索引结构，在进行数据读取时，分别读取索引指向的 Shelter 对应的层和 MainPart 来防止服务器了解真实读取的数据。

基于 ORAM 的可搜索加密能够达到非常高的安全保障，但这种高安全保障所需要的计算代价很高，例如文献[7]每一次要读取数据总量的对数级的数据量，当数据量很大时，这些方案很难具有实际价值。同时客户端要保存大量的相关数据，例如文献[7]和文献[12]上都要保存对数据的索引。在云计算中，数据的量往往很大，相应的数据客户也非常多，因此目前基于 ORAM 的隐私保护方案还很难在实际中应用。

2.2　基于对称加密的可搜索方案

基于关键字的隐私保护查询第一种方案就是采用可搜索对称加密技术。Song 等[18]首先明确提出了基于对称加密的可搜索密文技术（SSE, searchable symmetric encryption），并给出了一种无交互密文搜索方案。具体来说，Song 等的方案是为每一个关键词设计一个 2 层加密结构，当进行查询时，服务器通过用户提供加密的查询条件解开关键词的第一层加密来核对内层密文是否具有正确的形式来判断对应文档是否满足查询条件。该方案存的缺陷是容易遭受统计攻击，同时查询的时间复杂度是线性级。随后，Goh 给出了 2 种索引的安全模型[19]，即抵抗选择明文攻击的语义安全 IND-CKA 和 IND2-CKA。IND-CKA 和 IND2-CKA 模型保证文档内容不会被建立在其上的索引以及其他文档的索引泄漏，并提出了一种满足 IND-CKA 安全的安全索引 Z-Index。Z-Index 采用布鲁姆过滤器和伪随机

函数为每一个文档建立了一个索引，通过为每一个布鲁姆过滤器分配一个不同的 ID 以及对一些位随机置“1”使这些索引不可区分。Goh 的方案同样需要的查询时间复杂度与文档的数量呈线性级；同时，由于布鲁姆过滤器的假阳性问题使查询结果不可避免地含有假阳性数据，特别是当不同文档中出现的关键词的数量出现差异较大时，对布鲁姆过滤器的一些位随机置“1”来隐藏关键词数量将使查询结果中的假阳性比例进一步增大。

Curtmola 等[20]指出 IND2-CKA 不能够保证用户查询条件的隐私，他们构建了一个满足 IND2-CKA 安全的索引，并证明了通过该索引不能构建一个安全的可搜索的对称加密方案。在此基础上提出了 2 种攻击模型：非自适应的攻击模型和自适应的攻击模型。对于非自适应的攻击模型，攻击者构建查询条件时不考虑已有的查询历史，即已有的查询条件以及对应该查询条件的查询结果。自适应攻击模型中，攻击者的查询条件是根据已有的查询历史来进行构建，并指出在此之前的所有可搜索对称加密方案考虑的是在非自适应攻击模型下的安全，并将此前的 IND-CKA 和 IND2-CKA 称为 CKA1，称在自适应攻击模型下满足抵抗选择性明文攻击的安全为 CKA2。只有当查询条件独立于密文、加密的索引以及查询历史的情况下，满足 CKA1 安全模型的方案才真正安全[21]。文献[20]提出了一个满足 CKA2 安全的安全方案，该方案将不同文件中出现的同一个关键词采用不同的方式来表示，例如 3 个文档 D5、D8、D9 同时包含关键词“coin”，则分别采用“coin1”、“coin2”和“coin3”来表示。正如该文自身指出的一样，该方案虽然能满足 CKA2 安全，但同时增大了查询代价和存储代价。此后，研究者提出了多个满足 CKA2 安全的可搜索对称加密方案[22~24]，其中，文献[23]提出了一种更强的安全方案，该方案可以在恶意攻击模型下保证用户查询结果的正确性。

数据更新是一项基本的操作，如何在可搜索对称加密方案上实现数据更新一直是一个难点。Kamara 等在文献[21]中提出了动态可搜索对称加密的概念（dynamic searchable symmetric encryption），指出一个实际可行的可搜索对称加密方案必须满足如下 3 个条件：1）查询时间复杂度应该是亚线性；2）能抵抗适应性选择明文攻击；3）紧凑的索引结构并且支持高效的数据更新操作。并对其在文献[20]中提出的 SSE-1 方案进行改进来满足上述 3 个条件。SSE-1 方案只满足抵抗非自适应性选择明文攻击，并且不支持数据的更新操作。Kamara 通过采用非提交式加密方案实现 SSE-1 满足抵抗适应性选择明文攻击；为了实现动态性，他们通过增加一个额外的删除数组来帮助服务器找到指向被删除文件的指针；采用同态加密方案对存储在节点上的指向数据文件的指针进行加密，使服务器在不进行解密的情况下对指针内容进行修改；采用包含可利用空间表在内的额外结构来维护可用存储空间实现新节点的加入。但是该方案过于复杂，很难实现[25]，同时在更新数据的过程中泄漏了过多信息。

2013 年，Kamara 等进一步提出了一个基于红黑树结构的并行动态对称可搜索加密方案[25]。在他们的方案中，红黑树的每一个节点的数据域是 2 个长度完全相等的散列表，每一个关键词对应到 2 个散列表的同一个位置中的一个。通过这中构造来实现抵抗自适应选择明文攻击这种安全强度。通过二叉树结构的可并行性来实现查询过程的并行。该方案需要为关键字集合中所有关键词在散列表中分配一个位置，因此只适用于关键字数目较少的情况。

2.3 基于公钥体制的可搜索加密方案

与基于对称加密的可搜索方案对应的就是基于公钥体制的加密可搜索方案。

Boneh 等于 2004 年提出了一种基于双线性对的抵抗自适应选择明文攻击的可搜索的公钥加密（PEKS, public-key encryption with keyword search）方案[26]。该方案存在 2 个缺陷：由于采用公开的公钥对关键字进行加密，攻击者可以对关键字进行猜测并采用公布的公钥来验证其猜测的正确性；查询的时间复杂度与文档数量呈线性级。在此基础上，Abdalla 等进一步完善了 PEKS 的基础理论[27]。目前，科研人员基于 PEKS 提出了多种检索方法，文献[28]提出了支持范围检索和子集检索的 PEKS 方案，文献[29]提出了支持时间范围的检索方案，文献[30]提出了相似性检索方案，文献[31,32]提出了支持关键词合取的检索方案。然而，基于公钥体制的检索方法计算复杂度高，在实际应用中难以应用。

2.4 安全排名查询

安全排名查询是指系统按照一定的相关度准则（例如关键字出现的频率）将查询结果返回给用

户。安全排名查询提高了系统的适用性，符合云计算环境下的隐私数据保护的实际需求。

可搜索密码方案使用户能够安全地通过关键字搜索加密后的数据。考虑到“半可信”(semi-honest but curious)情形下，一方面，云服务器可能会主动分析存储在其本地的文件数据，窥探用户隐私，破坏数据安全性；另一方面，出于节省经济成本或带宽的目的，自私的云服务器可能只执行部分查询操作或只返回部分正确的查询结果。因此，Chai 等[33]提出了一种可验证查询结果的对称加密搜索算法。算法采用隐私保护特里树(privacy-preserving trie)对数据集合进行组织，利用对称密钥对所有文件数据进行独立加密确保其安全性。然而，可搜索密码方案只支持布尔查询，无法捕获数据的相关度信息。用户需要对查询结果中所有文件进行解密处理，并从中寻找出最佳匹配的结果，从而造成了用户需要处理大量数据文件的负担。另一方面，取回每个包含关键字的返回数据文件，势必会消耗许多不必要的网络流量，这是不适合当前“Pay-as-You-Use”的云架构。因此，可搜索密码方案在云计算环境下的查询效率并不是最佳的。

Wang 等[34]利用关键字频率和反文档频率的评分规则(TF-IDF rules)[35]对数据文档和关键字之间的相关度进行衡量，已现有可搜索密码方案[20]为基础上提出了一种 RSSE（ranked searchable symmetric encryption）算法，从而实现了对云计算环境下加密数据的安全排名查询。然而，RSSE 算法存在以下不足：1)继承可搜索加密算法的同时也引入了算法的固有缺陷，即在客户端会产生较大的计算功耗和后处理功耗；2)文件的相关度分数(relevance score)可能会被估算出来。因此，Wang 等进一步在该文中提出了一种基于保序对称加密算法[36]的 OPSE 算法，利用一对多的保序映射实现对数据文件的隐私性保护，同时通过对映射范围的优化提高了查询效率。

随后，Cao 等[37]将安全排名查询从一维关键字推广到了多维关键字，基于“协调匹配”原理[38]提出一种 MRSE（multi-keyword ranked search）算法，通过在查询条件中增加虚假关键字保护了用户的搜索行为模式和文件访问模式，同时结合内积相似度（inner product similarity）计算文件与多维关键字之间的相关度。然而 MRSE 算法存在如下问题[39]：1) MRSE 采用静态的关键字词典，一旦关键字数量增加就需要重新构建词典；2) MRSE 的令牌生成算法会导致查询结果顺序混乱的问题，查询结果可能遗漏包含较多匹配关键字的文件；3) MRSE 没有考虑关键字访问频率，降低了系统的实用性。针对上述词典重构、查询结果乱序等问题，Xu 等[39]通过引入分块矩阵构建关键字词典，使关键字词典的内容实现动态更新，同时向关键字词典引入随机因子，减少正态分布下的冗余关键字对查询结果正确性的影响。为了进一步实现精确查询，算法在相关度分数计算过程中引入了关键字频率，避免遗漏正确的查询结果。然而，该算法与 MRSE 算法均采用了矩阵保护文件隐私性，矩阵的乘法运算开销会对高维数据情形下的查询效率造成一定的影响[40]。

2.5　安全模糊查询

安全模糊查询是指系统根据查询条件和数据文档之间的相似程度将查询结果返回给用户。安全模糊查询能够容忍查询条件在文字或者格式上的细微错误，提高了系统的用户体验性，符合云计算环境下的隐私数据保护的实际需求。可搜索密码方案只支持精确查询，不能容忍少量的文字或格式错误，而这些少量错误往往在用户安全查询过程中不可避免。因此，可搜索密码方案并不适合云计算环境下的安全模糊查询。

Li 等[41]利用编辑距离（edit distance）[42]量化了关键字之间的相似度，并在通配符技术（wildcard）的基础上建立了模糊关键字集合，由此提出了一种安全模糊查询方案，实现了模糊语义下的安全查询，这是云计算环境下首次提出的安全模糊查询方案。文献[43]将各个关键字的通配符集合转换成反向索引表，通过特里树对索引进行结构化组织，提高了查询效率，同时通过采用虚假关键字提高了查询机制的安全性。然而上述工作存在以下不足：1) 算法需要将各个关键字作为索引树的叶子节点，导致存储空间开销很大；2) 不支持高效的增量式更新；3) 由于采用枚举方式构造模糊关键字集合，算法复杂度较高，当关键字长度为 l，编辑距离为 d 时，集合元素有 $2C_l^d + C_{l+1}^d$ 个。

2.6　数值型数据安全查询

与关键字查询相比，数值型数据的查询更复杂，因为对数值型数据进行查询不仅仅是进行等值查询，大多数情况还需要进行数据的大小比较。如何在保护隐私情况下进行数据大小的比较是一个极大的挑战。数值型数据的隐私查询技术同样

可以分为基于对称密钥的方案和基于非对称密钥的方案。传统基于对称密钥的数值隐私查询解决方案又可以进一步分为 3 种：基于桶划分的方案[44,45]、基于保序散列的方案[46]和基于保序加密的方案[47]。基于桶划分方案的基本思想是用户首先把数据按大小排序并划分成若干桶，然后将每个桶内的数据逐一加密，最后把桶编号并和数据一起交给云服务器；之后用户对其数据查询时，用户首先根据查询条件找出相关桶的编号并将编号交给云服务器作为查询条件。这种基于桶划分的隐私保护方案存在 3 个主要缺陷：1) 攻击者比较容易估计出数据的分布情况，这将部分暴露敏感信息；2) 当数据集中在少数几个桶中时，假阳性数据显著增加，浪费了传输带宽；3) 在多维情况时，存储空间将随数据维数增长呈指数级上升。基于保序散列方案和基于保序加密方案的基本思想类似，都是对数据进行特定的散列或者加密运算，其特点是运算后数据之间的大小关系保持不变。目前的保序加密和保序散列方法存在以下 3 个缺陷：1) 数据拥有者需要与数据使用者共享大量信息，造成胖客户端问题；2) 当攻击者已知数据域范围以及数据分布时，容易对数据真实值进行估计；3) 保序方法部分暴露了数据用户的兴趣信息。这 2 种方案的一个共同弱点是安全性不高，不能抵抗选择明文攻击。

2.7 国内隐私保护计算现状

在国内，研究人员对云计算下的隐私保护计算也进行了一系列的研究，并取得了一定的成绩。黄汝维等[48]设计了一个基于矩阵和向量运算的可计算加密方案，该方案通过运用向量和矩阵的各种运算实现对数据的加密，并支持对加密字符串的模糊检索和对加密数值型数据的加、减、乘、除 4 种算术运算。张逢喆等[49]从系统的角度设计了一个数据隐私保护与自我销毁的安全系统 Dissolver。该系统采用可信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。通过虚拟监控机来监控、保护数据的隐私性以及负责对数据的销毁处理。黄勤龙等[50]提出了一种云环境中支持隐私保护的数字版权保护方案。该方案采用基于属性基加密算法和同态加密算法保护内容加密秘钥的安全性。该协议允许用户匿名向云服务器订购内容和申请授权，在保护用户隐私的前提下防止云服务器收集用户使用习惯等敏感信息。朱旭东等[51]针对云计算环境下加密图像检索问题，提出了一种基于安全相似度运算的隐私保护检索方案。Li 等[52]针对目前云计算环境中隐私保护的范围查询协议存在的问题，提出了一个满足 IND-CKA 安全模型的快速范围查询协议。这些研究工作对云计算下的隐私保护研究具有积极的意义。

3 数据完整性认证

由于腐败雇员等问题，云平台可能并不可信，当用户向云服务器提交数据查询时，云服务器在返回给用户的查询结果中有可能包括虚假伪造的数据或者删除部分满足查询条件的数据。因此在云计算环境下查询结果的完整性认证非常重要，可以及时发现云的恶意行为。查询结果的完整性认证与传统意义上数据的完整性和所有权认证有一定的联系，但存在较大区别。传统意义上的数据完整性和所有权认证通常采用基于散列函数的消息认证码和数字签名来实现。例如，数据拥有者为了让数据使用者验证数据 D 的完整性和所有权，数据拥有者分别采用散列函数和签名函数计算数据的散列结果 H=Hash(D)和签名结果 S=Signpk(D)，其中签名采用数据拥有者的私钥进行。数据用户收到密文数据后进行解密得到明文，通过计算明文的散列结果与收到的散列结果 H 进行比对来确定数据的完整性，通过数据拥有者的公钥来验证签名 S 的真实性确定数据是否被伪造。但在云计算环境下，用户的查询结果只是整个数据文档中的一部分，而且查询结果随查询条件的改变而改变，数据用户很难实现计算某一个散列值，同样，也不能为所有的数据组合来计算签名值。

2007 年，Ateniese 等在 ACM CCS 会议上从云计算的角度提出了数据拥有的可证明性问题（PDP, provable data possession），即如何在不可信的服务器上存储文件，并且服务器能够向用户证明数据是完整的。在此基础上，Shacham 等[53]在 2008 年提出了一种基于椭圆曲线双线性对的紧凑查询结果证明（CPOR, compact proof of retrievability），该方案能支持所有用户的私有验证与授权用户的公开验证，但该方案建立在独立算法的基础上，缺乏统一的安全框架。2009 年，Dodis 等[54]研究了困难更大的 POR 方案。但所有这些方案都针对静态数据且基于公钥体制，因此都存在认证开销大的问题。

为解决动态数据操作下的数据验证问题，

Ateniese 等[55]提出了一种可扩展的数据拥有证明方案，该方案建立在密码学的 Merkele 散列树和对称密钥加密的基础上。但该方案缺乏随机性，攻击者可以通过窃听获得的信息欺骗数据用户；该方案另外一个缺陷是数据的修改次数是预先设定，存在容易失效的问题。2009 年，Erway 等[56]为了支持动态数据修改，提出了一种动态数据拥有证明方案（DPDP, dynamic provable data possession）。这些方案依然建立在散列树的基础上。对于具有 n 个数据块的数据文件，其基本方案在完整性认证方面需要 $O(\log n)$的通信与计算复杂度，在其改进的方案中存在信息泄露的问题。2009 年，Wang 等[57]采用 CPOR 和 Merkle 散列树相结合的办法提出了一种动态方案，该方案中散列树的变化依然非常复杂，与此相关的工作还有文献[58~60]。这些工作都不适合对多维数据的完整性认证问题。

在数据库方面，Narasimha 等[61]提出基于聚集签名与邻居链的关系数据库的查询认证方案。Chen 等[62]提出了一种规则区间树（CRT, canonical range trees）用于存储多维数据的统计信息，在不缺失界信息的情况下可利用统计信息对查询结果的完整性进行验证。但签名技术和数据链技术需要额外的验证对象，空间数据结构建立复杂并且不支持隐私保护下的查询。

在国内，咸鹤群等[63]提出了一种带掩码验证树的完整性认证方案。刘媛等[64]提出了一种基于 RSA 数据完整性机制来确保数据的完整性。李睿等[65]提出了一种水印链技术对查询结果进行验证。其思路是将数据组成呈线性结构，对每一个数据计算其散列值并将该散列值采用水印的方式嵌入到其前驱数据中，形成一条水印链。通过验证查询结果中水印链的完整性来验证查询结果的完整性。周恩光等[66]研究了在本地无副本的情况下对远程数据进行完整性认证问题，并利用基于等级的认证跳表提出了一种支持完全数据更新和公开审计的数据完整性认证方案。

4　数据访问权限控制

数据的访问控制是指数据拥有者管理数据用户访问数据的权限，包括对用户进行数据访问权限的授权以及对访问权限的回收。访问控制技术是保护数据不被非法使用的基础。在云计算环境中，数据拥有者和数据存储服务提供商不在同一个安全域内。一方面，为了保护数据的隐私性，云服务器不会被授权对数据内容的访问；另外一方面，数据存储到云服务器上后，数据拥有者已经无法从物理上对数据进行控制。因此采用传统的访问控制技术，即便数据拥有者制定了非常完善的访问控制策略，这些策略都可能不被执行。如何将数据存储到不可信的云服务器中还能对数据访问进行控制是一个极大的挑战。

目前，研究者主要从密码学角度出发来设计实现对数据的访问权限进行控制，即通过控制对数据加密密钥的分配来实现数据访问权限的控制。采用密钥的方式进行数据访问权限控制需要考虑的 2 个最基本问题：在实现数据访问权限的前提下如何降低密钥管理以及数据加密所需要的代价。

2001 年，Boneh 等[67]和 Cocks[68]分别基于双线性对和二次剩余假设提出了不同的身份加密方案，这些方案为后续构造灵活实用的属性和身份加密方案奠定了基础。

2005 年，文献[69]给出了第一个属性加密（ABE, attribute based encryption）方案，其基本思想是将数据与一系列的属性相关，并且用户也具有一定量的属性，属性系统根据用户的属性为其颁发密钥。数据拥有者在对数据进行加密时，设置解密者需要满足的属性条件；数据用户只有在满足所设置的属性条件才能对数据进行解密。此后出现了大量的 ABE 方案，根据访问控制策略通过密文或者密钥来实现的不同，可以分为基于密钥的属性加密（key policy ABE）方案[70,71]和基于密文的属性加密(cipher text policy ABE)方案[72~74]。在 KP-ABE 中，属性用于标记密文，属性的策略表达采用用户的私钥形式进行表达。在 CP-ABE 中，密钥与用户的属性相关，属性的策略在密文中表达。2007 年，Bethencour 等[72]采用密钥共享的方式提出了一种支持逻辑“与”、“或”运算的基于密文的 CP-ABE 方案，该方案可以产生多种访问控制策略。

2010 年，文献[75,76]分别提出了与广泛使用的脚色访问模型（RBAC, role based access control）一致的密码系统模型。Shahandashi 提出了基于门限的属性签名方案(t-ABS)，同年，Khader 提出了基于属性的群签名方案[74]。

文献[77]观察到在实际应用中每一个文件都可以赋予一组与该文件内容相关的属性，将每一

个用户的访问权限采用定义在这些属性之上的一个逻辑表达式来表达。通过逻辑表达式的强表达能力来实现用户细粒度数据访问权限的控制。该方案采用基于双线性对的公钥密码体制，为每一个属性分配一个公钥分量，数据文件采用与其相关的属性的公钥分量进行加密。分配给用户的密钥直接反应该用户的访问结构，一个文件的属性如果满足该用户的访问结构，则该用户的密钥能对该文件进行解密。该方案的加密复杂性只与数据文件的属性相关，与用户数目没有关系，因此适合用户数量大的应用场合。

在国内，洪澄等[78]提出了一种称之为 AB－ACCS 的访问控制方法，其核心思想是采用基于密文属性的加密算法为用户私钥设置属性，为数据密文设置属性条件，通过匹配用户私钥属性和密文数据属性来确定用户解密能力，从而达到对用户权限的控制。

5 数据物理安全

数据拥有者将数据存储在云服务器上，如何保证这些数据不被丢失是云端需要解决的问题。数据备份是一种常用的方法，Chen 等[79]提出了采用随机可扩展码技术对大规模云存储系统中的数据进行备份，对一个含有 n 个硬盘的阵列使用 t 个硬盘进行冗余备份，系统能够容忍 $t/2$ 硬盘的数据失效。Sun 等[80]首先对云环境中动态数据备份策略进行建模，分析出副本数和系统可靠性的关系，在此基础上设计出了一种动态备份算法，以提高系统数据可用性、容错能力，并能够优化系统带宽开销。Bonvin 等[81]针对云环境数据失效问题，提出一种可靠、高效的自管理密钥存储管理机制，动态地为多个应用分配资源，并对每个数据分区进行单独优化，根据最大化分区、存储和维护开销的净收益来选择是否进行数据迁移、备份和移除。Chen 等[82]提出了一种称为 ECT(ensure codes token)的容错机制保证云存储数据的正确性和错误的快速定位、恢复。数据用户事先计算一系列校验标记，每个标记对应一组随机的数据块，用户通过检查云服务器上随机生成字符串的块索引是否与校验标记是否一致来检查数据存储是否发生异常。Bicer 等[83]设计开发了一个应用程序接口，为云平台下数据的计算提供可靠性保障，该接口可由编程人员显式定义归约对象，系统周期性地将每个节点上的这些对象缓存在其他地方，并用其支持失效还原。当一个节点失效后，不在该节点处理的数据可以分发给其他节点，从该失效节点拷走的归约对象可以和其他节点上的归约对象合并处理以获得最终结果，其实验结果证明当失效发生时，能够快速地进行恢复。

目前，国内针对云计算中数据物理安全方面鲜有研究。

6 结束语

数据安全是安全云计算的核心，本文从数据的隐私保护、计算结果的完整性认证、数据访问权限控制以及数据的物理安全等 4 个方面对已有工作进行了总结，希望对后续云计算中的数据安全研究提供帮助。

参考文献：

[1] TechCrunch. iPad breach update: more personal data was potentially at risk[EB/OL]. http://techcrunch.com/2010/06/15/iPad-breach-personal-data/,2010.

[2] ITGI. Global status report on the gorvance of enterprise IT (GETIT)-2011[EB/OL].http://www.isaca.org/ITGI-Global-Survey-Results, 2011.

[3] GOLDREICH O. Towards a theory of software protection and simulation by oblivious RAMs[A]. STOC[C]. 1987.

[4] OSTROVSKY R, SHOUP V. Private information storage (extended abstract)[A]. STOC[C]. 1997.294-303.

[5] GOLDREICH O, OSTROVSKY R. Software protection and simulation on oblivious RAMs[J]. J ACM, 1996, 43(3): 431-473.

[6] OSTROVSKY R. Efficient computation on oblivious RAMs[A]. ACM Symposium on Theory of Computing(STOC)[C]. 1990.

[7] STEFANOV E, DIJK M, SHI E, *et al*. Path oram: an extremely simple oblivious ram protocol[A]. CCS[C]. 2013.

[8] GOODRICH M T, MITZENMACHER M, OHRIMENKO O, *et al*. Privacy-preserving group data access via stateless oblivious RAM simulation[A]. SODA[C]. 2012.

[9] KUSHILEVITZ E, LU S, OSTROVSKY R. On the (in)security of hash-based oblivious RAM and a new balancing scheme[A]. SODA[C]. 2012.

[10] WILLIAMS P, SION R. Round-optimal access privacy on outsourced storage[A]. CCS[C]. 2012.

[11] SHI E, CHAN T H H, STEFANOV E, *et al*. Oblivious RAM with O $((\log N)^3)$ worst-case cost[A]. ASIACRYPT[C]. 2011.197-214.

[12] BONEH D, MAZIERES D, POPA R A. Remote oblivious storage: making oblivious RAM practical manuscript[EB/OL]. http://dspace.

mit.edu/bitstream/handle/1721.1/62006/MIT-CSAIL-TR-2011-018.pdf, 2011.

[13] DAMGARD I, MELDGAARD S, NIELSEN J B. Perfectly secure oblivious RAM without random oracles[A]. TCC[C]. 2011.

[14] GOODRICH M T, MITZENMACHER M. Privacy-preserving access of outsourced data via oblivious RAM simulation[A]. ICALP[C]. 2011.

[15] GOODRICH M T, MITZENMACHER M, OHRIMENKO O, *et al*. Oblivious RAM simulation with efficient worst-case access overhead[A]. ACM Cloud Computing Security Workshop (CCSW)[C]. 2011.

[16] PINKAS B, REINMAN T. Oblivious RAM revisited[A]. CRYPTO[C]. 2010.

[17] WILLIAMS P, SION R, CARBUNAR B. Building castles out of mud: practical access pattern privacy and correctness on untrusted storage[A]. CCS[C]. 2008.

[18] SONG D, WAGNER D, PERRIG A. Practical techniques for searching on encrypted data[A]. Proc Symposium on Research in Security and Privacy (S&P)[C]. 2000.44-55.

[19] GOH E J. Technical report 2003/216, IACR ePrint Cryptography Archive[EB/OL]. http://eprint.iacr.org/2003/216.

[20] CURTMOLA R, GARAY J, KAMARA S, *et al*. Searchable symmetric encryption: improved definition and effcient constructions[A]. Proc ACM Conference on Computer and Communications Security (CCS)[C]. 2006. 79-88.

[21] KAMARA S, PAPAMANTHOU C, ROEDER T. Dynamic searchable symmetric encryption[A]. ACM CCSI[C]. 2012. 965-976.

[22] CHASE M, KAMARA S. Structured encryption and controlled disclosure[A]. Proc Int Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT)[C]. 2010. 577-594.

[23] KUROSAWA K, OHTAKI Y. UC-secure searchable symmetric encryption[A]. Proc Financial Cryptography and Data Security (FC)[C]. 2012.

[24] LIESDONK P, SEDGHI S, DOUMEN J, *et al*. Computationally efficient searchable symmetric encryption[A]. Proc Workshop on Secure Data Management (SDM)[C]. 2010. 87-100.

[25] KAMARA S, PAPAMANTHOU C. Parallel and dynamic searchable symmetric encryption[A]. Financial Cryptography and Data Security (FC'13)[C]. 2013.

[26] BONEH D, CRESCENZO G, OSTROVSKY R, *et al*. Public key encryption with keyword search[A]. Proceedings of Eurocrypt 2004, LNCS 3027[C]. 2004.506-522.

[27] ABDALLA M, BELLARE M, CATALANO D, *et al*. Searchable encryption revisited: consistency properties, relation to anonymous IBE, and extensions[J]. J Cryptology, 2008, 21(3): 350-391.

[28] BONEH D, WATERS B. Conjunctive, subset, and range queries on encrypted data[A]. The Theory of Cryptography Conference (TCC)[C]. 2006.535-554.

[29] DAVIS D, MONROSE Fn, MICHAEL K. Reiter: Time-scoped searching of encrypted audit logs[A]. ICICS 2004[C]. 2004. 532-545.

[30] CHEUNG D W, MAMOULIS N, WONG W K, *et al*. Anonymous fuzzy identity-based encryption for similarity search[A]. ISAAC[C]. 2010.61-72.

[31] PARK D J, KIM K, LEE P J. Public key encryption with conjunctive field keyword search[A]. WISA 2004[C]. Springer, Heidelbeg, 2004.73-86.

[32] GOLLE P, STADDON J, WATERS B. Secure conjunctive keyword search over encrypted data[A]. ACNS 04: 2nd International Conference on Applied Cryptography and Network Security[C]. 2004.31-45.

[33] WANG J, CHEN X, MA H, *et al*. A verifiable fuzzy keyword search over encrypted data[J]. Journal of Internet Services and Information Security, 2012:49-58.

[34] WANG C, CAO N, REN K, *et al*. Enabling secure and efficient ranked keyword search over outsourced cloud data[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(8): 1467-1479.

[35] ZOBEL J, MOFFAT A. Exploring the similarity space[J]. SIGIR Forum, 1998, 32(1):18-34.

[36] BOLDYREVA A, CHENETTE N, LEE Y, *et al*. Order-preserving symmetric encryption[A]. Proc of Eurocrypt[C]. 2009.

[37] CAO N, WANG C, LI M, *et al*. Privacy-preserving multi-keyword ranked search over encrypted cloud data[A]. Proc of INFOCOM[C]. 2011. 829-837.

[38] WITTEN I H, MOFFAT A, BELL T C. Managing Gigabytes: Compressing and Indexing Documents and Images[M]. Morgan Kaufmann Publishing, San Francisco, 1999.

[39] XU Z, KANG W, LI R, *et al*. Efficient multi-keyword ranked query on encrypted data in the cloud[A]. Proc of ICPADS[C]. 2012.

[40] SAVAŞ E, ÖRENCIK C. Efficient and secure ranked multi-keyword search on encrypted cloud data[A]. Proc of Joint EDBT/ICDT Workshops[C]. 2012.186-195.

[41] LI J, WANG Q, WANG C, *et al*. Fuzzy keyword search over encrypted data in cloud computing[A]. Proc of IEEE INFOCOM[C]. 2010. 441-445.

[42] LI J, WANG Q, WANG C, *et al*. Fuzzy keyword search over encrypted data in cloud computing[A]. Proc of IEEE INFOCOM[C]. 2010. 441-445.

[43] WANG C, REN K, YU S, *et al*. Achieving usable and privacy- assuredsimilarity search over outsourced cloud data[A]. Proc of INFOCOM[C]. 2012.

[44] AGRAWAL R, KIERNAN J, SRIKANT R, *et al*. Order preserving encryption for numeric data[A]. Proc SIGMOD[C]. 2004.563-574.

[45] BONEH D, CRESCENZO G D, OSTROVSKY R, *et al*. Public key encryption with keyword search[A]. Proceedings of Eurocrypt 2004[C]. 2004. 506-522.

[46] MICHEL A, MIHIR B, DARIO C, *et al*. Searchable encryption revisited: consistency properties, relation to anonymous IBE, and extensions[J]. Cryptology,2008, 21(3):350-391.

[47] DAN B, BRENT W. Conjunctive, subset, and range queries on encrypted data[A]. The Theory of Cryptography Conference (TCC)[C]. 2006.535-554

[48] 黄汝维，桂小林，余思等. 云计算环境中支持隐私保护的可计算加密方法[J]. 计算机学报, 2011, 34(12): 2391-2402.
HUANG R W, GUI X L, YU S, *et al*. Privacy-preserving computable encryption scheme of cloud computing[J]. Chinese Journal of Computers, 2011, 34(12): 2391-2402.

[49] 张逢喆, 陈进, 陈海波等. 云计算中的数据隐私性保护与自我销毁[J]. 计算机研究与发展, 2011, 48(7):1155-1167.
ZHANG F Z, CHEN J, CHEN H B, *et al*. Life time privacy and self-destruction of data in the cloud[J]. Journal of Computer Research and Development, 2011, 48(7):1155-1167.

[50] 黄勤龙, 马兆丰, 傅镜艺等. 云计算环境中支持隐私保护的数字版权保护方案[J]. 通信学报, 2014, 35(2):95-103.
HUANG Q L, MA Z F, FU J Y, *et al*. Privacy-preserving digital rights management scheme in cloud computing[J]. Journal on Communications, 2014, 35(2):95-103.

[51] 朱旭东, 李晖, 郭祯. 云计算环境下加密图像检索[J]. 西安电子科技大学学报(自然科学版), 2014, 41(2): 151-158.
ZHU X D, LI H, GOU Z. Privacy-preserving query over the encrypted image in cloud computing[J]. Journal of Xidian University, 2014, 41(2): 151-158.

[52] LI R, LIU A X, WANG L Y, *et al*. Fast range query processing with strong privacy protection for cloud computing[A]. The 40th International Conference on Very Large Data Bases[C]. 2014.1953-1964.

[53] SHACHAM H, WATERS B. Compact proofs of retrievability[A]. Proceedings of Asia Crypt[C]. Melbourne, Australia, 2008.90-107.

[54] DODIS Y, SALIL P, VADHAN D. Wichs: proofs of retrievability via hardness amplification[A]. IACR Cryptology ePrint Archive[C]. 2009.41.

[55] ATENIESE G, PIETRO R D, LUIGI V, *et al*. Scalable and efficient provable data possession[A]. IACR Cryptology ePrint Archive[C]. 2008.114.

[56] ERWAY C, KUPCU A, PAPAMANTHOU C, *et al*. Dynamic provable data possession[A]. Proceedings of the 16th ACM conference on Computer and communications security[C].2009. 213-222.

[57] WANG Q, WANG C, LI J, *et al*. Enabling public verifiability and data dynamics for storage security in cloud computing[A]. ESORICS 2009[C]. Saint Malo, France, 2009.21-25.

[58] BOWERS K D, JUELS A, OPREA A. HAIL: a high-availability and integrity layer for cloud storage[A]. Proceedings of the 16th ACM Conference on Computer and Communications Security[C]. 2009. 187-198.

[59] CURTMOLA R, KHAN O, BURNS R, *et al*. MR-PDP: multiple-replica provable data possession[A]. Proceeding ICDCS '08 Proceedings of the 2008 The 28th International Conference on Distributed Computing Systems[C]. 2008. 411-420.

[60] ATENIESE G, KAMARA S, KATZ J. Proofs of storage from homomorphic identification protocols[A]. Cryptology-ASIACRYPT'09[C]. 2009.319-333.

[61] NARASIMHA M, TSUDIK G. Authentication of outsourced databases using signature aggregation and chaining[A]. Proc Inte Conf on Database Systems for Advanced Applications[C]. Springer Berlin, 2006. 420-436.

[62] CHEN H, MAN X, HSU W, *et al*. Access control friendly query verification for outsourced data publishing[A]. Proc 13th European Symposium on Research in Computer Security[C]. Springer-Verlag, 2008. 177-191.

[63] 咸鹤群, 冯登国. 外包数据库模型中的完整性检测方案[J]. 计算机研究与发展, 2010,47(6):1107-1115.
XIAN H Q, FENG D G. An integrity checking scheme in outsourced database model[J]. Journal of Computer Research and Development, 2010,47(6):1107-1115.

[64] 刘媛, 涂晓东, 张兵. 关于外包数据库完整性验证的研究[J]. 计算机技术与发展, 2010, 20(5):150-153, 157.
LIU Y, TU X D, ZHANG B. Research on integrity verification of outsourcing database[J]. Journal of Computer Technology and Development, 2010, 20(5):150-153, 157.

[65] 李睿, 林亚平, 易叶青等. 两层传感器网络中隐私与完整性保护的范围查询协议[J]. 计算机学报, 2013, 36(6): 1194-1209.
LI R, LIN Y P, YI Y Q, *et al*. A privacy and integrity preserving range query protocol two-tiered sensor networks[J]. Chinese Journal of Computers, 2013, 36(6): 1194-1209.

[66] 周恩光, 李舟军, 郭华等. 一个改进的云存储数据完整性验证方案[J]. 电子学报, 2014, 42(1): 150-154.
ZHOU E G, LI Z J, GUO H, *et al*. An improved data integrity verfication scheme in cloud storage system[J]. Acta Electronica Sinica, , 2014, 42(1): 150-154.

[67] BONEH D, FRANKLIN M. Identity based encryption from the Weil pairing[A]. Crypto[C]. 2001.213-229.

[68] COCKS C. An identity based encryption scheme based on quadratic residues[A]. Proceedings of the 8th IMA International Conference on Cryptography and Coding[C]. 2001.360-363.

[69] SAHAI A, WATERS B. Fuzzy identity-based encryption[A]. EUROCRYPT 2005[C]. 2005. 457-473.

[70] GOYAL V, PANDEY O, SAHAI A, *et al*. Attribute-based encryption

for fine-grained access control of encrypted data[A]. ACM Conference on Computer and Communications Security 2006[C]. 2006.89-98.

[71] OSTROVSKY R, SAHAI A, WATERS B. Attribute-based encryption with non-monotonic access structures[A]. ACM Conference on Computer and Communications Security 2007[C]. 2007.195-203.

[72] BETHENCOURT J, WATERS B, SAHAI A. Ciphertext-policy attribute-based encryption[A]. SP '07 Proceedings of the 2007 IEEE Symposium on Security and Privacy[C]. 2007.321-334.

[73] GOYAL V, JAIN A, PANDEY O, *et al*. Bounded ciphertext policy attribute based encryption[A]. ICALP '08 Proceedings of the 35th international colloquium on Automata, Languages and Programming[C]. 2008.579-591.

[74] KHADER D. Attribute Based Group Signatures[R]. Cryptology ePrint Archive, Report 2007/159.

[75] ZHU Y, AHN G J, HU H X, *et al*. Cryptographic role-based security mechanisms based on role-key hierarchy[A]. Proceedings of 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS 2010)[C]. Beijing, China, 2010.

[76] ZHU Y, AHN G J, HU H G, *et al*. Cryptographic role-based security mechanisms based on role-key hierarchy[A]. ASIACCS 2010[C]. 2010. 314-319

[77] YU S C, WANG C, REN K, *et al*. Achieving secure, scalable, and fine-grained data access control in cloud computing[A]. Proc of INFOCOM 2010[C]. 2010.

[78] 洪澄, 张敏, 冯登国. AB-ACCS: 一种云存储密文访问控制方法[J]. 计算机研究与发展, 2010, 47(Z1): 259-265.
HONG C, ZHANG M, PENG D G. AB-ACCS: a cryptographic access control scheme for cloud storage[J]. Journal of Computer Research and Development, 2010, 47(Z1): 259-265.

[79] CHEN Z, XU Y, WANG X, *et al*. A new fault tolerance system for cloud storage[J]. Journal of Convergence Information Technology, 2011, 6(4): 34-41.

[80] SUN D, CHANG G, GAO S, *et al*. Modeling a dynamic data replication strategy to increase system availability in cloud computing environments[J]. Journal of Computer Science and Technology, 2012, 27(2):256-272.

[81] BONVIN N, PAPAIOANNOU T G, ABERER K. A self-organized, fault-tolerant and scalable replication scheme for cloud storage[A]. Proceedings of the 1st ACM Symposium on Cloud Computing (SoCC)[C]. New York, NY, 2010.

[82] CHEN D, PING X. Research on data fault tolerance mechanism based on ECT in cloud storage[J]. Communications in Computer and Information Science, 2013, 334:14-25.

[83] BICER T, JIANG W, AGRAWAL G. Supporting fault tolerance in a data-intensive computing middleware[A]. IEEE International Symposium on Parallel & Distributed Processing (IPDPS)[C]. 2010.1-12.

作者简介：

肖人毅，2010 毕业于大连理工大学管理学院，获得管理学博士学位；1993 年至 2000 年在国家自然科学基金委员信息中心工作；2000 年至今在国家自然科学基金委员会信息学部计算机处工作。