云环境下一种基于数据分割的CP-ABE隐私保护方案*

施荣华，刘 鑫，董 健，胡炳浩，李西柯
（中南大学 信息科学与工程学院，长沙 410075）

摘 要：针对云计算隐私安全保护，提出了一种基于数据分割的CP-ABE（密文策略的基于属性的加密方案）隐私保护方案，克服了云环境下不可信第三方、安全性和性能开销的三大难题。本方案利用数据分割思想将数据分为大数据块和小数据块，通过分割策略对大数据块再进行分块，并用CP-ABE算法对小数据块进行加密。经理论分析及实验仿真表明，在云环境下，此方案在安全问题、开销问题及扩展问题上都有很大优势。

关键词：云计算；隐私保护；数据分割；基于属性加密算法；CP-ABE算法
中图分类号：TP309.2 **文献标志码**：A **文章编号**：1001-3695(2015)02-0521-03
doi:10.3969/j.issn.1001-3695.2015.02.044

Privacy protection scheme in cloud computing using CP-ABE based on data partition

SHI Rong-hua, LIU Xin, DONG Jian, HU Bing-hao, LI Xi-ke
(*School of Information Science & Engineering, Central South University, Changsha* 410075, *China*)

Abstract: According to cloud computing privacy protection, this paper proposed a CP-ABE(cipertext policy-attribute based encryption) scheme based on data partition which improved security, it reduced the performance overhead, overcome untrustful three sides. This scheme used the data ideological to divide date into big and small block of data, then divided the big data block into small pieces and encrypting the small data block with CP-ABE algorithms. Manager deals with experimental analysis, in a cloud environment, this scheme has advantages on the safety and performance overhead and extension.

Key words: cloud computing; privacy protection; data partition; attribute-based encryption algorithm; CP-ABE algorithms

0 引言

云计算被喻为信息领域工业化革命的一种新型的网络计算模型[1]。云计算构建了一个聚合被虚拟化的计算资源数据中心，为用户提供了动态的、高效的、高性价比的、高扩展性的信息服务[2]。在云环境中，用户不需要了解云中基础设施的具体细节，也不要掌握相关专业知识，不用直接进行控制。一般而言，由业务供应商提供通用的网络应用业务，用户直接可以通过Web服务进行访问，其中相关的数据和软件都放在云服务器上。起初云计算只在企业内部网络运行，所涉及的隐私安全问题并不深刻，随着云计算的发展，隐私安全问题已经成为制约云计算推广的重要绊脚石[3]。目前云服务提供商给用户提供的隐私和安全保护很有限，带来了一系列的安全问题[4,5]。云计算的隐私保护有如下主要难点：

a)在云环境下，数据存储在云中，数据管理和数据拥有是相分离的[6]。云是由第三方作为服务器，第三方的可信度会受到用户的质疑。

b)由于云环境下数据量相当大，使用传统的数据加密技术会导致服务器开销太大。这就限制了很多安全性高但算法复杂的一些加密算法。

c)在云环境下，用户的服务系统升级及更新大部分都是用户远程运行，这样使得每次升级和更新时存在一些潜在的威胁，其中密钥的发放也存在很大难度。

为了解决云计算的安全问题，近年来不少学者提出了许多加密技术和方法，其中有学者提出数据分割加密方案[7]，即云存储的用户端系统首先自动将待托管的数据切割为小数据块和大数据快，小数据块存储在本地，大数据块则按照用户指定的安全级别需求进行加密后由云端文件系统分块存储[8]。分割数据虽然增加了安全性，但加大了时间开销，而且已经不能适应高速发展的云计算的应用。也有学者提出了基于属性加密算法(ABE)[9]，基于属性加密机制自2005年开始研究，Sahai等人[10]首次提出了基于模糊身份加密，将生物特性作为身份信息应用于加密方案中；Goyal等人[11]于2006年在基于身份加密方案的基础上提出了基于数据加密方案(ABE)；Bethencount等人[12]于2007年提出了密文策略的基于属性的加密方案(CP-ABE)。其中CP-ABE[12,13]是一种有效的云储存访问控制机制，基于CP-ABE的方案在更容易管理密钥的同时也

收稿日期：2014-01-17；**修回日期**：2014-03-01 **基金项目**：国家自然科学基金资助项目(61272495,60773013)；长沙市科技计划基金资助项目(K1104009-11)

作者简介：施荣华(1963-)，男，湖南常德人，教授，博士，主要研究方向为密码学、信息安全(shirh@csu.edu.cn)；刘鑫(1989-)，男，湖南邵阳人，硕士研究生，主要研究方向为信息安全、云计算；董健(1980-)，男，湖南常德人，副教授，博士，主要研究方向为网络与信息安全、移动通信；胡炳浩(1990-)，女，河北邢台人，本科生，主要研究方向为移动通信；李西柯(1975-)，男，湖南长沙人，博士，主要研究方向为云计算.

对用户更加透明，但随着数据量以及属性值的增大，CP-ABE 算法将会给系统带来巨大的开销。

本文提出了基于数据分割策略的 CP-ABE 方案，它是利用定长的数据分割思路并采用 CP-ABE 加密算法对数据块加密的隐私保护方案，能确保安全性的同时减少了服务器的开销。

1 方案设计

1.1 数据分割

LPCA(linear partition-combination algorithm)[14] 提出了线性分割数据的方法，而且还提供了安全恢复数据策略，非常适用于大规模数据存储的分布式系统，但其安全性并不高。随后提出了另一种数据分割机制，将数据在客户端进行分块并加密传输到网络上的存储服务器，并将目录信息存在本地，这种新的机制将数据管理和数据拥有进行了分离，应用中可以防止第三方的信息泄露。

数据分割机制的基本思路是将数据分割为 m 块，当完全具备 $n(n\leqslant m)$ 个分块时才能恢复数据。也就是说，任意 $m-n$ 个数据块丢失或者损坏时都能恢复原数据，增强机制的可靠性和可用性，被盗任意少于 n 个子块时，不能恢复原数据，从而提高安全性。

分割数据根据原数据大小来决定其流程，如图 1 所示。

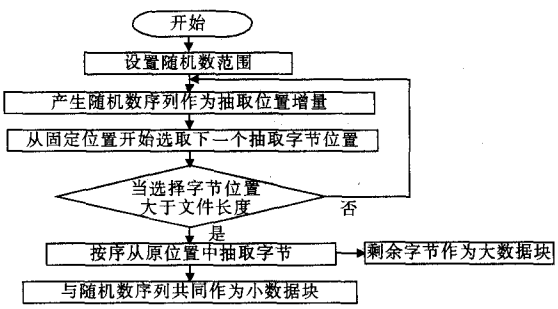

图 1 数据分割机制流程

该机制的安全性比较高，可以让数据管理者也无法直接查看数据拥有者的数据，而且可以减少服务器端的数据量，给服务器减少了不小的压力，但分割时间的开销为 $O(n)$，其中 n 是小数据块的大小，n 的合理选取也有相当的难度。最关键的是这种数据分割策略只能用于云计算的个人数据上传与下载，并不能适应现在云计算的发展。

1.2 基于属性加密算法(ABE)

ABE 属于公钥加密机制，传统的公钥加密机制必须获取用户公钥证书才能加密，并且还有密钥发送的开销大、占用宽带多等缺点。ABE 引入了属性的概念，这使得 ABE 面向的解密对象是一个群体，而不是单个用户，这使得 ABE 算法在云计算中有广阔的用途。基于属性加密算法在云计算法中最大的优势是基于 CP-ABE 的方法将生成访问控制的权利交给数据拥有者，这样数据所有者可以自主方便地选择用户访问文件。

CP-ABE 算法：假设 $I=\{I_1,I_2,\cdots,I_n\}$ 为所有属性的集合，则 $M\subseteq\{I_1,I_2,\cdots,I_n\}$，其中 M 为每个用户的属性，那么 N 个属性可以鉴别 2^N 个用户，访问结构 $T\subseteq 2^{\{I_1,I_2,\cdots,I_n\}}\setminus\{\varphi\}$，在 T 中属性集合表示授权集合，不在 T 中的表示非授权集合。该算法主要包括四个步骤：

a) 生成主密钥 MK 和公开参数 PK。

b) $CT=\text{Encrypt}(PK,W,T)$。使用 PK 和 T 以及加密明文 W 加密后得到密文 CT。

c) $SK=\text{KeyGen}(MK,M)$。使用 MK 和 M 生成用户的私密 SK。

d) $W=\text{Decrypt}(CT,SK)$。使用 SK 解密 CT 得到 W。

本文选取 CBHAC 算法与 CP-ABE 算法进行对比。CBHAC 算法是利用对称加密方法来进行数据加密的，是一种简单有效的加密方法。图 2 是通过仿真来比较两种算法的结果，选取属性数为 10，文件数为 500，实验结果如图 2 所示。

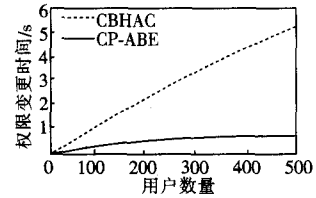

图 2 权限变更时间

由图 2 不难看出，权限变更时间 CP-ABE 比 CBHAC 小得多，并且所占用的存储空间 CP-ABE 比 CBHAC 算法要小 10 倍左右。这样一种加密算法在云计算中相对于其他加密算法而言具有其特别的优势[15]。尽管这样，随着数据量以及属性值的增大，CP-ABE 算法将会给系统带来相当大的开销。

1.3 基于数据分割的 CP-ABE 隐私保护方案

上文已经提到了数据分割策略和 CP-ABE 算法。基于两者的优势提出了基于数据分割的 CP-ABE 隐私保护方案。

基于数据分割的 CP-ABE 加密方案是根据数据分割机制衍生而来，数据分割的简单原理如图 3 所示。

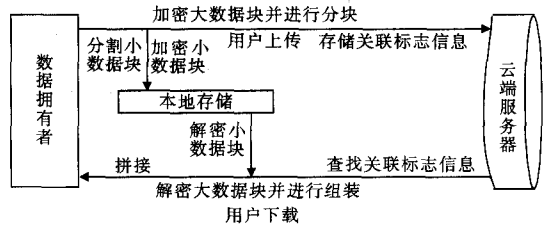

图 3 数据分割机制原理

由图 3 不难看出，数据分割机制在云存储方面体现了其巨大的优势，这种分割方法在安全性上可以与公钥加密算法相提并论，而且对大数据块的加密中，低强度的加密方式大大地减小了系统开销。但是这种分割机制只能利用在云存储上，这种本地存储方式只能给数据拥有者上传和下载自己的东西。随着云服务的发展，这种模式使用范围相当狭小。比如在医疗云范畴内，病人远程(远程医生或移动医疗设备等)医疗，需要看到医生上传到云服务端的诊断结果。可见，这种分割机制只能完成数据拥有者的数据保存，让数据拥有者的数据不被他人所使用，而云的发展是让数据拥有者上传数据并可以控制其数据只能被他授权的人读取和使用。针对上述情况，本文提出一种改进的隐私保护方案，设计框架如图 4 所示。上传数据流程为：

a) 数据拥有者选择上传数据，数据分割后分为大数据块和小数据块。本方案采用抽取固定大小的小数据块方案，这样能减少存储和读写的压力。流程如图 5 所示。

b) 用 CP-ABE 算法给小数据块进行加密。

c) 用中、低强度加密算法给大数据块进行加密并进行分块。

d) 存储大数据块和小数据块以及相关标志信息。

下载数据流程为：

a) 服务器验证用户，根据相关标志查看用户是否属于小

数据块的 CP-ABE 授权集合。

b）如果属于授权集合，那么解密小数据块及组装大数据块。

c）大数据块和小数据块进行拼接。

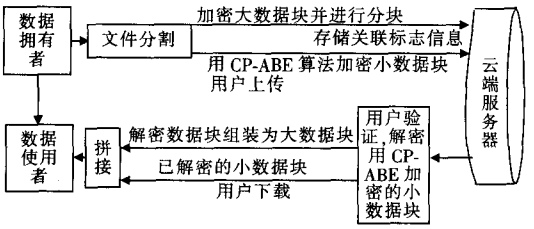

图 4　基于数据分割的 CP-ABE 加密方案设计

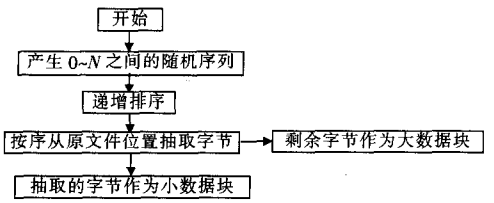

图 5　抽取固定大小的小数据块流程

以上为本方案设计流程，在方案设计中本文将数据分割中的小数据块也存储在云端并用 CP-ABE 算法进行加密，这使得数据拥有者可以指定数据使用者。本文之前提到的 CP-ABE 算法是一种基于属性加密算法，只有属于 CP-ABE 的授权集合才能解密小数据块，这样就连数据管理者也不能解密数据，这使得用户无须担心第三方可信程度，而且由于大数据块的分块使得本方案具有非常高的安全性，在这种情况下甚至可以用低强度的加密算法对大数据块进行加密来减小系统开销。随着云的发展，用户属性将越来越多，数据量也将越来越大，如果用 CP-ABE 算法来加密所有数据，那将是系统的一个大开销。本方案只用 CP-ABE 算法来加密小数据块，有效地降低了开销。

1.4　方案扩展性

本方案的扩展性是设计的重点之一。本文以电子健康档案（EHR）为例，将所有用户的电子健康档案存储在某一云服务商上，这里假设为“总云”。随着云的普及，企业有企业的私有云，行业也将会出现所谓行业的“私有云”，那么医疗行业也有了它自己的云，此时出现了另一个云，这里假设它为“医疗行业云”。这样，本文设计的方案中在上传数据时，用户拥有者可将 CP-ABE 加密的小数据块放在自己所处的“医疗行业云”上，能解密小数据块的用户使用者就能进入“总云”获取他所需要的信息。这样，在减少了“总云”的开销的同时也能更好地划分用户属性。如果医疗行业有所谓的“行业云”，进一步如果每个医院有所谓的私有云，这里假设为“医院云”，那么本方案中的小数据块就可以放在“医院云”中，这种情况下只要提供云服务的服务商之间做好协调就可以了。如果用户资料泄露，责任归属将更加明确。所以随着云的发展，该方案将会有更广泛的应用范围。

2　安全性和性能分析

2.1　安全性分析

基于数据分割的 CP-ABE 加密方案采用数据分割和 CP-ABE 算法双重加密，具有相当高的安全性。下面从几个方面来分析本方案的安全性。

a）访问权限的安全管理。本方案无须考虑第三方服务供应商是否可靠，因为服务供应商不参与数据加密的密钥产生与管理，完全由数据拥有者对其他用户进行访问授权。

b）数据的高安全性。由于数据是分块存储的，所以攻击者并不能获取完整的元数据信息来得到所有数据块以恢复大数据块。数据分割后进一步保证了云端数据的隐私安全。

c）数据完整性。由于分割了小数据块并用 CP-ABE 算法加密，这使得就算攻击者获取了大数据，也不能通过完整性验证。大数据的分块模式组装策略就算有部分分块丢失或者被攻击者窜改，也能用其他分块进行恢复，确保了数据的完整性。

2.2　性能分析及仿真

基于数据分割的 CP-ABE 加密方案用了数据分割思路并结合了 CP-ABE 算法，在目前的云环境下看似加大了系统开销，但本方案使用的数据分割法是固定大小分割方案，这使得时间复杂度为 $O(1)$，比起数据大小块分割的时间复杂度 $O(n)$ 减少了时间开销，并且本方案采用的 CP-ABE 算法只加密小数据块，在一定程度上减少了系统开销。

本方案基于 MATLAB 进行了数据仿真，实验环境如下：CPU 为 Intel Core i5-3210M 2.50 GHz，内存为 4.00 GB，操作系统为 Win 7 64 位操作系统，仿真软件为 MATLAB 7.0。数据分割小块选择为 10，数据文件大小为 1 MB，大数据块加密选择了简单的染色加密。通过仿真来比较本方案和 CP-ABE 完全加密方案在属性个数增大下加密时间的变化。文件大小 1 MB 时性能比较如图 6 所示，文件大小为 10 MB 时性能比较如图 7 所示。

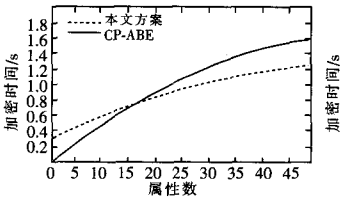

图 6　基于数据分割的 CP-ABE 加密方案和 CP-ABE 完全加密方案性能比较（文件大小为 1 MB）

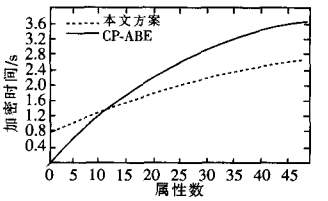

图 7　基于数据分割的 CP-ABE 加密方案和 CP-ABE 完全加密方案性能比较（文件大小为 10 MB）

根据图 6 和 7 所示，当属性数量比较少时，本方案比起 CP-ABE 全加密算法所花费时间稍高，但随着属性数的增大，本方案在用时上大大小于 CP-ABE 算法所耗时间，在文件较大时，本方案优势更加明显。当下，百度云盘、360 云盘等大型云盘都已逾 1 000 万用户，表 1 为在 100 万用户（属性个数最小取值为 20）的情况下本方案与 CP-ABE 全加密方案的比较。

表 1　属性个数为 20 的情况下本方案与 CP-ABE 全加密方案加密所耗时间比较

文件大小/MB	本文方案/s	CP-ABE/s
1	0.81	0.92
10	1.94	2.27

在云环境中，随着用户数量的增加，CP-ABE 算法中的属性个数将越来越多，所以基于数据分割的 CP-ABE 加密方案非常适用于云环境中。本方案还具有强大扩展性的特点，如果将小数据块存放在低级云中，开销将会有质的下降。

3　结束语

本文提出了基于数据分割的 CP-ABE 加密方案将数据分割与CP-ABE 算法结合，在考虑其性能的前提下，增强了安全性。其采用数据分割方式增加安全性以外并用 CP-ABE 算法加密小数据块减小系统开销，可以在第三方服务 （下转第 527 页）

n^2+1 轮交互通信；协议 3 最坏情况只需 1 轮交互通信；协议 4 只需 1 轮交互通信。计算复杂性和通信复杂性的比较见表 1。

表 1　计算复杂性和通信复杂性(通信轮数)比较

比较项	协议 2	协议 3	协议 4
计算复杂性	$2n^2+2$	6	1
通信复杂性	n^2+1	1	1

5　结束语

数据服务外包是目前企业重要的商业行为，而在外包时希望不被承接方获取机密信息，需要利用多方保密计算协议来实现数据处理和整合。矩阵的保密计算在控制科学、系统优化等外包服务项目中具有重要作用。本文提出了将哥德尔编码应用到矩阵判等问题的多方保密计算中，以及保密计算矩阵特征值的解决方案。两协议在原有朴素计算方法的基础上降低了计算复杂性和通信复杂性，并利用模拟范例证明了协议对半诚实者是保密的。该协议在数据服务外包领域具有实用价值。

参考文献：

[1] YAO A C. Protocols for secure computations[C]//Proc of the 23th IEEE Symposium on Foundations of Computer Science. [S. l.]: IEEE Press, 1982: 160-164.

[2] GOLDREICH O, MICALI S, WIGDERSON A. How to play any mental game[C]//Proc of the 19th Annual ACM Conference on Theory of Computing. [S. l.]: IEEE Press, 1987: 218-229.

[3] GOLDREICH O. The fundamental of cryptography: basic applications [M]. London: Cambridge University Press, 2004.

[4] GOLDWASSER S. Multiparty computations: past and present[C]// Proc of the 16th Annual ACM Symposium on Principles of Distributed Computing. New York: ACM Press, 1997: 1-6.

[5] CHOI S G, HWANGY K W, KATZ J, *et al*. Secure multi-party computation of Boolean circuits with applications to privacy in on-line market places [C] //Lecture Notes in Computer Science, vol 7178. Berlin: Springer, 2012: 416-432.

[6] TOFT T. Secure data structures based on multiparty computation [C]//Proc of the 30th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing. New York: ACM Press, 2011: 291-292.

[7] LOFTUS J, SMART N P. Secure outsourced computation[C]//Lecture Notes in Computer Science, vol 6737. Berlin: Springer, 2010: 1-20.

[8] LIU Wen, WANG Yong-bin. Secure multiparty comparing protocol and its applications[J]. Acta Electronica Sinica, 2012, 40(5): 871-876.

[9] SHEIKHA R, MISHRA D K, KUMAR B. Secure multiparty computation: from millionaires problem to anonymizet[J]. Information Security Journal: A Global Perspective, 2011, 20(1): 25-33.

[10] CRAMER R, DAMGAARD I. Introduction to secure multi-party computations[EB/OL]. [2005-06-11]. http://homepages. cwi. nl/cramer/.

[11] 徐仲，张凯院，陆全，等. 矩阵论简明教程[M]. 北京：科学出版社，2005.

[12] 李顺东，戴一奇，尤启友. 姚氏百万富翁问题的高效解决方案[J]. 电子学报，2005，33(5)：770-773.

[13] 李顺东，王道顺. 现代密码学：理论、方法与研究前沿[M]. 北京：科学出版社，2009.

[14] 李顺东，王道顺，戴一奇，等. 两个集合相等的多方保密计算[J]. 中国科学，F 辑：信息科学，2009，39(3)：305-310.

[15] DIFFIE W, HELLMAN M E. New direction in cryptography[J]. IEEE Trans on Information Theory, 1976, 22(6): 644-654.

[16] LI Shun-dong, WANG Dao-shun, DAI Yi-qi. Secure signature protocol [J]. Intelligent Information Management, 2009, 1(3): 174-179.

[17] LIN H Y, TZENG W G. An efficient solution to the millionaires problem based on homomorphic encryption [C]//Applied Cryptography and Network Security (LNCS 3531). Berlin: Springer, 2005: 456-466.

[18] LI Shun-dong, WANG Dao-shun, DAI Yi-qi, *et al*. Symmetric cryptographic solution to Yao's millionaires' problem and an evaluation of secure multiparty computations [J]. Information Sciences, 2008, 178(1): 244-255.

[19] 李顺东，王道顺. 基于同态加密的高效多方保密计算[J]. 电子学报，2013，41(4)：798-803.

(上接第 523 页)商不可信的情况下确保数据安全及完整。随着云计算的发展，本方案有着其强大的扩展性，本文中小数据块的存放地点与大数据块加密算法是下一步研究的重点。

参考文献：

[1] VAQUERO L M, RODERO-MERINO L, CACERES J, *et al*. A break in the clouds: towards a cloud definition[J]. ACM SIGCOMM Computer Communication Review, 2009, 39(1): 50-55.

[2] 金海，吴松，廖小飞，等. 云计算的发展与挑战，180734[R]. 北京：机械工业出版社，2010.

[3] WANG Qian, WANG Cong, REN Kui, *et al*. Enabling public auditability and data dynamics for storage security in cloud computing[J]. IEEE Trans on Parallel and Distributed Systems, 2011, 22(5): 847-859.

[4] 冯登国，张敏，张妍，等. 云计算安全研究[J]. 软件学报，2011，22(1)：71-83.

[5] 邹德清，金海，羌卫中，等. 云计算安全挑战与实践[J]. 中国计算机学会通讯，2011，7(12)：55-61.

[6] HACIGUMUS H, IYER B, MEHROTRA S. Providing database as a service[C]//Proc of the 18th International Conference on Data Engineering. 2002: 29-40.

[7] 徐小龙，周静岚，杨庚. 一种基于数据分割与分级的云存储数据隐私保护机制[J]. 计算机科学，2013，40(2)：98-102.

[8] 王保军. 电子数据分离存储于安全恢复系统的研究及实现[D]. 南京：南京邮电大学，2009.

[9] 苏金树，曹丹，王小峰，等. 属性基加密机制[J]. 软件学报，2011，22(6)：1299-1315.

[10] SAHAI A, WATERS B. Fuzzy identity-based encryption[C]//Proc of the 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques. Berlin: Springer-Verlag, 2005: 457-473.

[11] GOYAL V, PANDEY O, SAHAI A, *et al*. Attribute based encryption for fine-grained access control of encrypted data[C]//Proc of ACM Conference on Computer and Communications Security. New York: ACM Press, 2006: 89-98.

[12] BETHENCOUNT J, SAHAI A, WATERS B. Ciphertext-policy attribute-based encryption[C]//Proc of IEEE Symposium on Security and Privacy. [S. l.]: IEEE Press, 2007: 321-334.

[13] 孙国梓，董宇，李云. 基于 CP-ABE 算法的云存储数据访问控制[J]. 通信学报，2011，32(7)：146-152.

[14] 张薇，马建峰. LPCA—分布式存储中的数据分离算法[J]. 系统工程与电子技术，2007，29(3)：453-458.

[15] LI Ming, YU Shu-cheng, ZHENG Yao, *et al*. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption[J]. Trans on Parallel and Distributed Systems, 2013, 24(1): 131-143.