



面向“互联网+”的公有云数据安全

王永建¹, 宋爱波¹, 叶亚伟², 姜海波³

(1. 东南大学计算机科学与工程学院, 江苏 南京 211189;

2. 中国通信建设集团设计院有限公司, 北京 100079;

3. 国网河南省电力公司检修公司, 河南 郑州 450052)

摘要: “互联网+”催生了许多新的经济形态与商业模式, 公有云面临着严峻的安全挑战。研究了公有云数据安全问题, 并提出了研究思路。首先, 分析了同态加密的概念、加法同态加密与乘法同态加密的特点以及当前的研究成果和需要解决的难题。然后, 根据乘法同态加密算法、散列表和相似性理论, 提出了一种数据安全保护方案, 并阐述了具体实现流程, 采用欧氏距离检验公有云中加密数据的相似性与完整性。最后, 理论分析了该方案的正确性与安全性。仿真实验验证了该方案的可行性与有效性。

关键词: 公有云; 同态加密; 散列表; 相似性; 欧氏距离

中图分类号: TN919.8

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017241

Data security for internet plus in public cloud

WANG Yongjian¹, SONG Aibo¹, YE Yawei², JIANG Haibo³

1. School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

2. China International Telecommunication Construction Group Design Institute Co., Ltd., Beijing 100079, China

3. State Grid Henan Electric Power Company Maintenance Company, Zhengzhou 450052, China

Abstract: “Internet plus” has given birth to many new economic forms and business models, the public cloud faces a serious security challenges. The security of public cloud data was studied and the research ideas were put forward. Firstly, the concept of homomorphic encryption, the characteristics of additive homomorphic encryption and multiplicative homomorphic encryption, and the current research results and difficult problems to be solved were analyzed. Then, based on the multiplicative homomorphic encryption algorithm, hash table and similarity theory, a data security protection scheme was proposed, and the concrete realization process was described. The similarity and completeness of the encrypted data in the public cloud were tested by Euclidean distance. Finally, the correctness and security of the scheme were analyzed by theory. The simulation results verify the feasibility and validity of the scheme.

Key words: public cloud, homomorphic encryption, hash table, similarity, Euclidean distance

收稿日期: 2017-07-29; 修回日期: 2017-10-11

基金项目: 国家自然科学基金资助项目 (No.61173138)

Foundation Item: The National Natural Science Foundation of China (No. 61173138)



1 引言

随着云计算、大数据、移动互联网等信息技术的发展,2012年11月,诞生了互联网发展的新理念——“互联网+”。“互联网+”是基于信息传播的传统互联网发展的新业态,催生了大量新的产业经济形态与商业服务模式。

云计算作为“互联网+”的关键技术,提供了新的交付模式和服务模式,如快速便捷的交付模式、动态高效的资源配置、强大可扩展的处理能力、灵活方便的订购服务、较低可控的投资风险等,受到云服务提供商与租户的青睐。

然而,从云计算诞生之日起,关于其安全性问题一直争议不断。思科 CEO 钱伯斯曾说:云计算是安全的噩梦^[1]。此言虽有夸张之嫌,但确实反映了云计算在产生巨大经济效益与提供良好服务的同时,面临诸多挑战,其中安全问题被认为是最大的挑战,是云服务亟待解决的最大难题^[2,3]。

根据服务对象,云计算可分为私有云、公有云、混合云。面向“互联网+”的云服务属于公有云,也是云环境中安全问题最为突出的一种,其数据安全主要体现在以下方面。

(2) 不信任问题

云服务中源数据主要由租户(或者用户)提供,虽然云服务商向租户提供一系列信任与安全措施,效果却不理想。租户担心敏感数据被云服务商泄露、传播、篡改,或者被其他租户、互联网用户等窃取、篡改、传播、销毁等。

(3) 虚拟化问题

在虚拟化环境中,每台物理机上有多个虚拟机,虚拟机之间的部分流量不可视,且缺乏流量行为审计^[4]。存在虚拟机被非法访问、虚拟机计算数据泄露、虚拟机迁移过程中未彻底清除缓存或者数据未被完全覆盖等隐患^[5]。

(4) 多方保密计算问题

实际的公有云往往提供分布式云服务,多个

用户提供数据,用户之间分享数据,用户之间是对等的,在没有可信第三方的情况下,用户之间协作使用私有数据,同时又不能泄露、篡改各自私有数据是个难题。典型的应用为 P2P、电子商务^[6]、数据挖掘^[7]、计算外包^[8]、保密存储^[9]、密钥共享等。

(5) 差分隐私问题

“互联网+”催生了多种多样的经济形态与服务模式,用户的初衷是分享非敏感信息、隐匿敏感信息,而现实情况是这些隐私信息被泄露或者出现“分享扩大”问题。

公有云中数据安全是个比较大的概念,本文认为数据有两种状态:静态时的存储数据和动态时的运算数据。静态数据是指用户的文档、报表、资料等不参与计算的用户数据;动态数据是指需要动态验证或参与计算的用户数据^[10]。因此,数据的安全要保证静态与动态时都要安全,有如下两种解决思路。

- 思路 1: 数据以密文形式存储,直接使用密文计算。该思路不仅能保障数据存储时的安全性,并且能保证计算过程中数据的安全性。目前仅有同态加密(homomorphic encryption)可实现存储的密文直接计算,计算后仍以密文形式存储。同态加密能够从根本上解决信息委托给第三方时的保密问题,非常适合于云计算环境^[11]。
- 思路 2: 数据以明文形式存储,计算时采用加密计算。该思路无法保证数据存储时的安全性。虽然理论上也可采用同态加密的方法,先对明文进行同态加密,然后进行密文计算;但是采用该方法,加密的运算量可能比用户对数据进行运算更复杂,增加了不必要的开销,实际意义不大^[12]。反之,数据以密文形式存储,计算时采用明文,计算后再以密文形式存储,这个过程中密文与明文两次转换,运算量更大,更不具备实际意义。

因此, 本文以思路 1 进行研究。从云环境中数据以密文形式存储与检索出发, 基于同态加密思想设计了一种数据安全保护方案, 以解决公有云中明文数据在存储与检索过程中的安全保密问题。

2 理论分析

2.1 同态加密

同态加密概念于 1978 年由 Rivest 等人^[13]提出。同态是指对密文进行操作的同时也对明文执行相应的操作^[14], 即对密文进行的特定运算与对明文进行的另一种(可能与前者不同)运算是等价的^[15]。

本文对同态加密定义如下: 设加密操作为 E , 密文为 e , 解密操作为 D , 明文为 m , 对明文操作为 h , 则:

$$e = E(m), \quad m = D(e) \quad (1)$$

若对于 E , 存在操作 F , 使:

$$F(e) = E(h(m)) \quad (2)$$

则称 E 是 h 的同态加密算法。

同态加密分为加法同态加密与乘法同态加密: 设 m 与 n 为明文, 如果存在有效操作 “ $\oplus$ ”, 使 $E(m+n) = E(m) \oplus E(n)$, 则称 E 为加法同态加密; 设 m 与 n 为明文, 如果存在有效操作 “ $\otimes$ ”, 使 $E(m \times n) = E(m) \otimes E(n)$, 则称 E 为乘法同态加密; 同时满足加法同态与乘法同态的称为全同态加密, 仅满足加法同态或者乘法同态的称为部分同态加密。

全同态加密是解决本文研究思路的理想方法, 不过同态加密概念自提出后即成为密码学界一个开放的难题。虽然研究者先后提出了多种同态加密方法, 但基本都属于部分同态加密。具有代表性的如 Elgamal^[16]提出的算法、RSA 算法(以发明者的名字命名: Rivest、Shamir 和 Adleman)^[13]属于乘法同态加密。Paillier^[17]、Benaloh^[18]、

Okamoto^[19]、Naccache^[20]等提出的算法属于加法同态加密。2005 年, Boneh^[21]利用代数环的结构特性, 首次提出了同时支持加法同态与乘法同态的 BGN(以发明者的名字命名: Dan、Goh、Nissim)加密算法, 支持任意次数加法同态加密运算, 但是只能做一次乘法同态加密运算, 该方案是距离全同态加密最接近的研究成果。

由于理论上电路与函数可以互相转化, 于是产生了通过电路实现全同态加密的方法, 目前全同态加密算法基本上是电路全同态加密, 如 Fellows^[22]提出的 Polly Cracker 同态加密方案。不过函数转化为电路计算非常复杂, 设计难度巨大, 很难实际应用, 所以, 全同态加密的应用在于代数全同态加密算法的重大突破。

直到 2009 年, 全同态加密算法在理论上有了重大突破。IBM 公司的 Gentry^[23,24]基于理想格(ideal lattice)的方法提出了第一个全同态加密算法, 设计了一个具有自主性的同态加密方案, 能够转换为全同态加密方案, 可以支持任意深度的电路计算。这在学术界产生了巨大影响, 不过 Gentry 在算法设计过程中引入了很多数学困难问题假设, 这些假设问题并未经过充分论证, 还有待进一步的研究验证, 因此目前还无法实际应用。

Gentry 奠定了第一代全同态加密的基础。在此基础上, 学术界又做了新的研究。如 Dijk 等^[25]在 Gentry 研究的基础上提出了一种全同态加密方案, 不过该方案不需要理想格。Stehle 等^[26]对 Bootstrapping 电路全同态加密进行了改进。Coron 等^[27]提出了一种短密钥提高计算效率的方法。

Brakerski 和 Gentry^[28-31]等基于差错学习(learning with error)提出了第二代全同态加密方案, 这些方案具有较短的公钥长度、更小的解密复杂度、更高的效率。Wang 等^[32]提出了一种基于整多项式环上差错学习(ring-learning with error,



R-LWE) 问题的有效线性同态加密方案 (linear homomorphic encryption scheme, LHES), 不过只局限于对云端密文数据进行线性操作。汤殿华等^[33]提出了带密钥转换的重线性化技术, 结合模转换, 设计了一个基于 R-LWE 的非自举的层次化全同态加密方案。Liang^[34]基于通用量子电路, 设计了一种全同态加密方法。Dulek Y 等^[35]提出了一种量子同态加密设计方案, 适用于任意多项式大小的量子电路。陈智罡等^[36]从离散高斯分布上选取 LWE (learning with error) 样例, 提出一个 LWE 上的短公钥多位全同态加密方案。

虽然国内外学者对同态加密做了大量的研究与改进, 但基本上是电路全同态加密, 运行效率仍很低, 并且有许多假设问题尚未解决或者验证, 实际应用难度仍然较大。

鉴于此, 本文采用部分同态性加密算法进行研究。

2.2 数据相似性分析

实际的公有云往往提供分布式云服务, 通过完全分散的机制来发现和分配数据资源, 每个租户可以既是数据提供者又是数据使用者, 数据呈分布式存储, 因此数据安全保护与检索难度加大。本文借鉴散列表 (hash table) 建立检索表, 通过将关键码值 (key value) 映射到表中某个位置来查询数据, 以加快检索的速度。

为提高数据检索精度与安全性, 本文引入相似性理论, 它能提供数据隐私保护和非常高的数据挖掘精度^[37]。常见的相似性度量方法有欧氏距离、闵可夫斯基距离 (Minkowski distance)、切比雪夫距离 (Chebyshev distance)、曼哈顿距离 (Manhattan distance)、余弦相似度、马氏距离 (Mahalanobis distance) 等。

(1) 欧氏距离

欧氏距离源自欧氏空间中两点间的距离, 是一种最简单、直接的相似性度量方法, 能衡量多个维度上属性的差异性。一维、二维、三维或高

维空间中两个点 x 和 y 之间的欧氏距离可表示为:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (3)$$

其中, n 是维数, 而 x_k 和 y_k 分别是 x 和 y 的第 k 个属性值 (分量)。

(2) 闵可夫斯基距离

闵可夫斯基距离不是一种距离, 而是一组距离的定义。两个 n 维变量 $a(x_{11}, x_{12}, \dots, x_{1n})$ 与 $b(y_{21}, y_{22}, \dots, y_{2n})$ 之间的闵可夫斯基距离可表示为:

$$d(x, y) = \sqrt[p]{\sum_{k=1}^n |x_{1k} - y_{2k}|^p} \quad (4)$$

其中, p 是一个变参数: $p=1$ 时, 是曼哈顿距离; $p=2$ 时, 是欧氏距离; $p \rightarrow \infty$ 时, 是切比雪夫距离。根据变参数 p 的不同, 闵可夫斯基距离可以表示一类距离。

闵可夫斯基距离 (包括欧氏距离、曼哈顿距离和切比雪夫距离) 都存在明显的缺点, 它将对象的不同属性 (分量) 之间的差别等同看待, 没有考虑各个分量 (期望、方差等) 的分布可能是不同的, 这一点有时不能满足实际要求。

(3) 余弦距离

余弦距离也称为余弦相似度, 是用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小的度量。相比欧氏距离, 余弦距离更加注重两个向量在方向上的差异, 更多的是从方向上区分差异, 而对绝对的数值不敏感, 因此没法衡量每个维度上数值的差异。

(4) 马氏距离

马氏距离也可以定义为两个服从同一分布并且其协方差矩阵为 Σ 的随机变量之间的差异程度, 如果协方差矩阵为单位矩阵, 那么马氏距离就简化为欧氏距离。马氏距离的优点是与量纲无关, 排除了变量之间的相关性的干扰; 缺点为夸大了变化微小的变量的作用, 测量结果不稳定。

综上分析, 欧氏距离虽然有明显的缺点, 但是能够体现个体数值特征的绝对差异, 比较适用

于需要从多维度的数值大小中体现差异的分析。尤其是随着近些年对欧氏距离的不断改进研究,其缺点得到了显著改善。为保证数据检索的准确性,需要从多维度属性值进行判断。

因此,本文选择欧氏距离,在能保持隐私的前提下,又能度量两个向量的差异^[38]。Hiroaki 等^[39]提出了一种改进的欧氏距离的方案,该方案基于零知识证明和 Fujisaki-Okamoto 承诺,具有同态加密的性质,效率比较高,并可以抵御恶意攻击。为保证安全隐私, Barni 等^[40]提出了一种改进的欧氏距离的方案,该方案中引入了安全多方计算协议。

参考文献[38]基于分布式 El Gamal 乘法同态加密算法,提出了一个新的、加密的欧氏距离改进方案,该协议的计算和通信复杂度都较低。因此,本文采用参考文献[38]的方法检测加密数据的相似性,验证加密数据的完整性,结合第 2.1 节的分析,基于 El Gamal 乘法同态加密进行设计。

3 所提数据安全保护方案

3.1 研究思路

利用 El Gamal 算法对明文加密与解密,并利用 El Gamal 算法的乘法同态性,保证对计算后的密文解密后得出正确的明文。方案分为两部分:建立检索表,首先对数据进行切片,然后对文件名称与数据片段进行乘法同态加密,并建立相应的检索表;检索数据,首先验证文件名称,然后检索数据并进行比对,采用欧氏距离检验数据是否安全。

3.2 El Gamal 算法

El Gamal 是一种基于离散对数问题的非对称公钥加密算法,可用于加密、解密、数字签名,实现简单、安全性高、应用广泛。

3.2.1 El Gamal 算法实现

(1) 密钥对生成

设 p 是一个大素数; g 和 x 是两个随机数, $g < p$, $x < p$; $y = g^x \pmod{p}$, 则其公钥为 y 、

g 和 p ; 私钥是 x 、 g 和 p , 可由一组用户共享^[41]。

(2) El Gamal 用于数字签名

设被签名信息为 M , 首先选择一个随机数 k , k 与 $p-1$ 互质, 计算:

$$a = g^k \pmod{p} \quad (5)$$

再用 Euclidean (欧几里德) 扩展算法对下面方程求解 b :

$$M = xa + kb \pmod{(p-1)} \quad (6)$$

签名就是 (a, b) , 随机数 k 需丢弃。

验证时要验证式 (7):

$$y^a a^* b \pmod{p} = g^M \pmod{p} \quad (7)$$

其中, $*$ 代表类似“乘”的逻辑操作。同时一定要检验是否满足 $1 \leq a < p$, 否则签名容易伪造。

(3) El Gamal 加密解密

设被加密信息为 M , 首先选择一个随机数 k , k 与 $p-1$ 互质, 计算:

$$a = g^k \pmod{p} \quad (8)$$

$$b = y^k M \pmod{p} \quad (9)$$

其中, (a, b) 为密文, 是明文的两倍长。

解密时计算:

$$M = b/a^x \pmod{p} \quad (10)$$

El Gamal 签名的安全性依赖于乘法群上的离散对数计算, 要求素数 p 必须足够大, 且 $p-1$ 至少包含一个大素数^[42]。

3.2.2 El Gamal 算法同态性

在群 G 里, 公钥为 (G, q, g, h) , 其中 $h = g^x$, x 是私钥。加密函数 $\Phi(m) = (g^r, m * h^r)$, 其中, $r \in \{0, 1, 2, \dots, q-1\}$, $*$ 代表类似“乘”的逻辑操作, 则有式 (11):

$$\begin{aligned} & \Phi(x_1) * \Phi(x_2) \\ &= (g^{r_1}, x_1 * h^{r_1}) * (g^{r_2}, x_2 * h^{r_2}) \\ &= (g^{r_1+r_2}, (x_1 * x_2) h^{r_1+r_2}) \\ &= \Phi(x_1 * x_2) \end{aligned} \quad (11)$$

所以, El Gamal 满足乘法同态加密。



3.3 建立检索表

定义：设文件名称为 f ，数据为 d ；乘法同态加密为 E_m ；公钥为 p ，私钥为 k ，公钥与私钥对为 (p, k) ；数据切为 n 个片段 $S(d, n)$ ；随机数为 r ， $r \in (1, n)$ ；欧氏距离为 v 。

步骤 1 利用乘法同态加密算法，加密文件名称为 $E_m(f)$ 。

步骤 2 数据切片，将数据 d 切分为 n 个片段， $S(d, n): d_1, d_2, d_3, \dots, d_n$ ；对所有的数据片段 d_i 进行乘法同态加密， $E_m(d_i): E_m(d_1), E_m(d_2), E_m(d_3), \dots, E_m(d_n)$ ， $i \in [1, n]$ 。

步骤 3 计算欧氏距离，计算所有 $E_m(d_i)$ 随机对 r 之间的欧氏距离 v ，同理计算 $E_m(f)$ 随机对 r 之间的欧氏距离 v' 。

步骤 4 建立检索表，设 $E_m(d_i)$ 与 $E_m(d_j)$ 之间欧氏距离为 v_x ，建立加密数据片段检索表 $R_d: <v_x, E_m(d_i), E_m(d_j)>$ ， $j \in [1, n]$ ， $x \in [1, r]$ ；同理建立加密文件名称检索表 $R_f: <v'_x, E_m(f_i), E_m(f_j)>$ 。

步骤 5 存储信息，将加密数据片段 $E_m(d_i)$ 和加密文件名称 $E_m(f)$ 分别独立存储在公有云中。

建立检索表流程如图 1 所示。

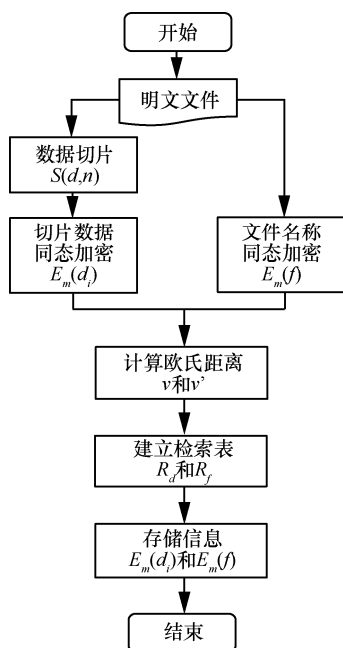


图 1 建立检索表流程

3.4 数据检索流程

定义：解密为 D ，对数据 d 解密为 $D(d, k)$ 。

步骤 1 验证文件名称，设要查询的数据的加密后的文件名称为 $E_m(f_i)$ ，在加密文件名称检索表 R_f 中进行检索比对；设 $E_m(f_i)$ 随机对 r 之间的欧氏距离为 v' ，如果 $E_m(f)$ 与 $E_m(f_i)$ 之间距离 $v'_i=0$ ，则执行步骤 2；否则，继续检索。

步骤 2 检索数据，设要查询的数据片段为 d_i ，在加密数据检索表 R_d 中检索结果为 $E_m(d_i)$ ，随机对 r 之间的欧氏距离为 v_i ；如果 $v_i=v'_i$ ，说明 $E_m(d_i)$ 没被篡改过，则执行步骤 3；否则，继续检索。

步骤 3 解密文件，解密文件 $D(E_m(f_i), k)$ 和 $D(E_m(d_i), k)$ 。

数据检索流程如图 2 所示。

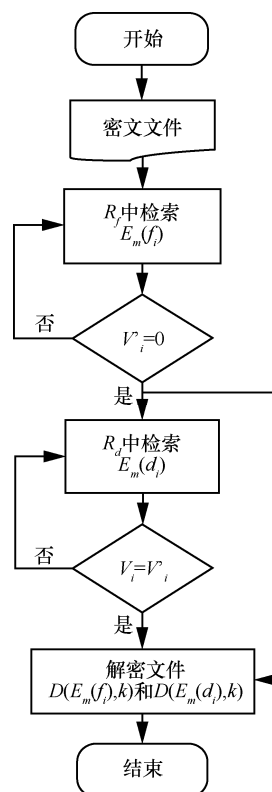


图 2 数据检索流程

4 理论分析与仿真验证

4.1 正确性

由于采用了乘法同态加密，两个数据的明文

与加密后的密文的欧氏距离应该是相等的, 可由理论推导证明。

设文件名称为 f_i 、 f_j , 则:

$$\begin{aligned} & D(E_m(f_i) \otimes E_m(f_i) + E_m(f_j) \otimes E_m(f_j) - \\ & 2 \cdot E_m(f_i) \otimes E_m(f_j)) = D(E_m(f_i - f_j)^2) = (f_i - f_j)^2 \end{aligned} \quad (12)$$

设数据片段为 d_i 、 d_j , 则:

$$\begin{aligned} & D(E_m(d_i) \otimes E_m(d_i) + E_m(d_j) \otimes E_m(d_j) - \\ & 2 \cdot E_m(d_i) \otimes E_m(d_j)) = D(E_m(d_i - d_j)^2) = (d_i - d_j)^2 \end{aligned} \quad (13)$$

以上推导可证明该方案的正确性。

4.2 安全性

主要分析方案对数据的隐私保护。建立检索表时对不同用户分配各自的公钥与私钥对 (p, k) , 云服务商与其他用户是不知道的, 并对文件名称与数据片段进行了加密, 因此云服务商与其他用户无法知道原始数据。检索数据时, 只有当文件名称验证通过后, 方可进行下一步——检索数据片段, 检索数据片段时再次通过欧氏距离验证数据是否被篡改过。

上述过程的核心是保证明文的安全, 由 El Gamal 同态加密算法来保障, 因此主要分析文中同态加密算法的安全性。

El Gamal 同态加密算法的困难问题可归约到部分近似最大公因子问题 (partially approximate common division problem, PACDP)^[43]。PACDP 的证明可见参考文献[44]中定理 2 的证明方法, 本文简要阐述。

定理 1 假设攻击者可以用一个优势为 ε 的算法 A 攻破同态加密算法, 则存在一个算法 B, 通过调用 A, 能够解决 PACDP, 其优势为 $\varepsilon/2$ 。

证明: 黑客 H 产生一个 PACDP, 假设 α 比特位大素数为 p , β 比特位大素数为 q , 随机整数 $r \in (-2^p, 2^p)$, $n = pqr$ 。

(1) 将 N 发送给算法 B, B 使用 N 作为公钥。

(2) B 首先访问加密算法, 产生许多 0 的密文, 记为 $z_1, z_2, \dots, z_n$, 则存在某两个数 r_i 和 k_i , 使得 $z_i = 2r_i + k_i p$ 。

(3) 从步骤 (2) 中 0 的密文中随机选择两个数 z_1^* 和 z_2^* , 调用算法二进制 GCD。

(4) 输出元素 $z = p + r$, 用 $z_1 = z_1^*$ 和 $z_2 = z$ 调用算法二进制 GCD, 输出 $q_p(z_1^*)$ 的二进制比特。

(5) 计算 $p = \lceil z_1^* / q_p(z_1^*) \rceil$, 还原得出 p 。

(6) 根据参考文献[45], 得到 B 的优势为 $\varepsilon/2$ 。

对本文方案的任何攻击方法都可以转换为解决 PACDP, 目前解决 PACDP 的方法主要是格归约求解^[46]。根据参考文献[47]对格归约求解 PACDP 的详细深入研究得出的结论, 用格归约求解 PACDP 是困难的, 目前几乎不可能在实际中应用。

因此, 本方案能对数据提供有效的安全隐私保护。

4.3 仿真实验

构建仿真实验环境: Intel Core i7, 16 GB DDR-IV, Windows 10, JDK v1.8, CloudSim v4.0。由于在 Java 7 的 API 中没有 El Gamal 算法组件, 本实验采用 BouncyCastle 中的 El Gamal 算法组件, 密钥的长度可选范围为 160~16 384 bit (密钥长度必须是 8 的整数倍), 为了更好地满足安全性兼顾运行效率, 选用默认密钥长度 1 024 bit, 使用 ECB 和 none 的工作模式。

测试数据集选用笔者集团公司项目管理系统的文件, 筛选出 300 个文件, 大小为 10 KB~79 MB, 每个文件分为两种情况进行实验: 明文检索、利用本文设计方案的密文检索。

因此, 利用 CloudSim v4.0 中 examples 目录中的 CloudSimExample3.java, 创建一个两台主机、两个任务的数据中心, 给每个任务提供相同的资源。



先将测试数据集的明文文档上传至分布式文件系统中, 再将这些明文数据切片, 并利用 El Gamal 算法组件加密。经过处理后得到的密文文档数量远远大于明文文档数量, 所占空间也变大, 需要增加二进制文件来存储密文文件中多出的字节。为了更好地体现实验结果, 每个明文文档处理后产生的多个密文文档之间的顺序与明文文档之间的顺序相对应。另外, 为了验证本方案的优越性, 将本方案与参考文献[48]提出的检索算法做对比。对比结果如图 3 所示。

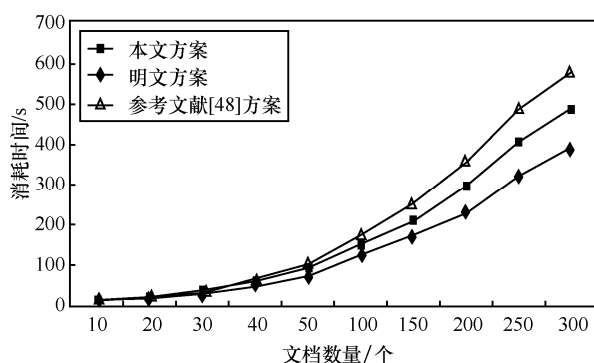


图 3 仿真实验结果

根据图 3 的实验结果, 由于每个明文文档的大小不同, 并且经过处理后得到的密文文档数量所占空间变大, 3 种方案消耗的时间不随着文档数量的增加呈线性关系。在文件数量较少时, 3 种方案的消耗时间基本是相等的。文档数量不超过 30 个时, 参考文献[48]的方案比本文的方案消耗时间少; 当文档数量超过 30 个后, 本文的方案优势开始显现, 效率更高。

5 结束语

本文利用数据相似性、乘法同态加密算法设计了一种公有云中的数据安全保护方案。经过分析, 该方案在理论上是正确的; 通过仿真实验, 验证了该方案的有效性与优越性, 能够为公有云中的数据提供有效安全性保障。利用全同态加密保障数据安全的研究存在许多待解决的难题, 例如当前的全同态加密方案运行效率低, 仍无法实

际应用。未来同态加密的研究重点是: 简化复杂度, 提高云环境中密文数据的检索效率; 解决全同态加密中的噪声问题、运算效率低问题; 在电路实现全同态加密的思路之外, 期望在代数全同态加密算法领域有所突破。另外, 公有云中的数据安全是个比较大的概念, 是个系统工程, 除了许多关键技术外, 良好的管理制度、操作规范等也很重要。该领域的研究任重道远, 还需要做大量的研究工作。

参考文献:

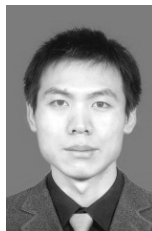
- [1] MCMILLAN R. Cloud computing a 'security nightmare', says Cisco CEO[EB/OL]. (2014-07-02)[2016-10-09]. <http://www.networkworld.com/news/2009/042309-cloud-computing-a-security-nightmare.html>.
- [2] HARAUIZ J, KAUFMAN L M, POTTER B. Data security in the world of cloud computing[J]. IEEE Security & Privacy, 2009, 7(4): 61-64.
- [3] GUO F D, ZHANG M, ZHANG Y, et al. Study on cloud computing security[J]. Journal of Software, 2011, 22(1): 71-83.
- [4] 王笑帝, 张云勇, 刘楠, 等. 云计算虚拟化安全技术研究[J]. 电信科学, 2015, 31(6): 8-12, 24.
WANG X D, ZHANG Y Y, LIU D, et al. Research on security of virtualization on cloud computing[J]. Telecommunications Science, 2015, 31(6): 8-12, 24.
- [5] 张逢喆. 公共云计算环境下用户数据的隐私性与安全性保护[D]. 上海: 复旦大学, 2010.
ZHANG F Z. Privacy and security protection of user data in public cloud computing environment[D]. Shanghai: Fudan University, 2010.
- [6] CHOI S G, HWANGY K W, KATZ J, et al. Secure multi-party computation of boolean circuits with applications to privacy in online marketplaces[M]. Berlin: Springer Berlin Heidelberg, 2012.
- [7] AGRAWAL R, SRIKANT R. Privacy-preserving data mining[J]. Application Research of Computers, 2000, 29(2): 439-450.
- [8] LOFTUS J, SMART N P. Secure outsourced computation[M]. Berlin: Springer Berlin Heidelberg, 2010.
- [9] TOFT T. Secure data structures based on multi-party computation[C]//The 30th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing, June 6-8, 2011, San Jose, California, USA. New York: ACM Press, 2011: 291-292.
- [10] 刘明辉, 张尼, 张云勇, 等. 云环境下的敏感数据保护技术研究[J]. 电信科学, 2014, 30(11): 2-8.
LIU M H, ZHANG N, ZHANG Y Y, et al. Research on sensitive data protection technology on cloud computing[J]. Tele-

- communications Science, 2014, 30(11): 2-8.
- [11] 李浪, 余孝忠, 杨娅琼, 等. 同态加密研究进展综述[J]. 计算机应用研究, 2015(11): 3209-3214.
- LI L, YU X Z, YANG Y Q, et al. Survey on homomorphic encryption technology[J]. Application Research of Computers, 2015(11): 3209-3214.
- [12] 李顺东, 窦家维, 王道顺. 同态加密算法及其在云安全中的应用[J]. 计算机研究与发展, 2015, 52(6): 1378-1388.
- LI S D, DOU J W, WANG D S. Survey on homomorphic encryption and its applications to cloud security[J]. Journal of Computer Research and Development, 2015, 52(6): 1378-1388.
- [13] RIVEST R, ADLEMAN L, DERTOUZOS M. On data banks and privacy homomorphisms[J]. Foundations of Secure Computation, 1978, 4(11): 169-180.
- [14] 任福乐, 朱志祥, 王雄. 基于全同态加密的云计算数据安全方案[J]. 西安邮电大学学报, 2013, 18(3): 92-95.
- REN F L, ZHU Z X, WANG X. A cloud computing security solution based on fully homomorphic encryption[J]. Journal of Xi'an University of Post and Telecom, 2013, 18(3): 92-95.
- [15] 王辉. 同态加密的分析与优化[D]. 杭州: 杭州电子科技大学, 2012.
- WANG H. Homomorphic encryption: analysis and optimization[D]. Hangzhou: Hangzhou Dianzi University, 2012.
- [16] ELGAMAL T. A public-key cryptosystem and a signature scheme based on discrete logarithms[J]. IEEE Transactions on Information Theory, 1985, 31(4): 469-472.
- [17] PAILLIER P. Public-key cryptosystems based on composite degree residuosity classes[C]//The 17th international conference on Theory and application of cryptographic techniques, May 2-6, 1999, Prague, Czech Republic. New York: ACM Press, 1999: 223-238.
- [18] BENALOH J. Verifiable secret-ballot elections[D]. New Haven: Yale University, 1988.
- [19] OKAMOTO T, UCHIYAMA S. A new public-key cryptosystem as secure as factoring[C]//International Conference on the Theory & Applications of Cryptographic Techniques, May 31-June 4, 1998, Espoo, Finland. Berlin: Springer, 1998: 308-318.
- [20] NACCACHE D, STERN J. A new public key cryptosystem based on higher residues[C]//The 5th ACM Conference on Computer and Communications Security, November 2-5, 1998, San Francisco, California, USA. New York: ACM Press, 1998: 59-66.
- [21] DAN B, GOH E J, NISSIM K. Evaluating 2-DNF formulas on cipher texts[C]//The 2nd International Conference on Theory of Cryptography, February 10-12, 2005, Cambridge, England. Berlin: Springer, 2005: 325-341.
- [22] FELLOWS M, KOBLITZ N. Combinatorial cryptosystems galore[J]. Contemporary Mathematics, 1994(1): 51-61.
- [23] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//The 41st Annual ACM Symposium on Theory of Computing, May 31-June 2, 2009, Bethesda, MD, USA. New York: ACM Press, 2009: 169-178.
- [24] GENTRY C. A fully homomorphic encryption scheme[D]. Stanford: Stanford University, 2009.
- [25] DIJK M V, GENTRY C, HALEVI S, et al. Fully homomorphic encryption over the integers[C]//The 29th Annual international conference on Theory and Applications of Cryptographic Techniques, May 30-June 3, 2010, French Riviera, France. New York: ACM Press, 2010: 24-43.
- [26] STEHLE D, STEINFELD R. Faster fully homomorphic encryption[C]//The 16th Int Conf on the Theory and Application of Cryptology and Information Security, December 5-9, 2010, Singapore. Berlin: Springer, 2010: 377-394.
- [27] CORON J S, MANDAL A, NACCACHE D, et al. Fully homomorphic encryption over the integers with shorter public keys[C]//Conference on Advances in Cryptology, August 14-18, 2011, Santa Barbara, CA, USA. New York: ACM Press, 2011: 487-504.
- [28] BRAKERSKI Z, GENTRY C. Fully homomorphic encryption without bootstrapping[J]. ACM Transactions on Computation Theory, 2015, 6(3): 1-36.
- [29] BRAKERSKI Z. Fully homomorphic encryption without modulus switching from classical GapSVP[M]. Berlin: Springer Berlin Heidelberg, 2012.
- [30] GENTRY C, SAHAI A, WATERS B. Homomorphic encryption from learning with errors conceptually-simpler, asymptotically-faster, attribute-based[M]. Berlin: Springer Berlin Heidelberg, 2013.
- [31] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient fully homomorphic encryption from (standard) LWE[J]. SIAM Journal on Computing, 2014, 43(2): 831-871.
- [32] WANG T, YU J P, YANG Y J. Linear homomorphic encryption scheme for privacy protection of cloud storage[J]. Journal of Signal Processing, 2013, 29(11): 1463-1469.
- [33] 汤殿华, 祝世雄, 王林, 等. 基于 RLWE 的全同态加密方案[J]. 通信学报, 2014, 35(1): 173-182.
- TANG D H, ZHU S X, WANG L, et al. Fully homomorphic encryption scheme from RLWE[J]. Journal on Communications, 2014, 35(1): 173-182.
- [34] LIANG M. Quantum fully homomorphic encryption scheme based on universal quantum circuit[J]. Quantum Information Processing, 2014, 14(8): 1-11.
- [35] DULEK Y, SCHAFFNER C, SPEELMAN F. Quantum homomorphic encryption for polynomial-sized circuits[M]. Berlin: Springer Berlin Heidelberg, 2016.
- [36] 陈智翌, 宋新霞, 赵秀凤. 一个 LWE 上的短公钥多位全同态加密方案[J]. 计算机研究与发展, 2016, 53(10): 2216-2223.
- CHEN Z G, SONG X X, ZHAO X F. A multi-bit fully homomorphic encryption with better key size from LWE[J]. Journal of Computer Research and Development, 2016, 53(10):



- 2216-2223.
- [37] 刘立坤. 基于噪音添加和欧几里德距离的隐私保护数据挖掘方法[D]. 长春: 吉林大学, 2011.
- LIU L K. Privacy-preserving data mining method based on noise additive and Euclidean distance[D]. Changchun: Jilin University, 2011.
- [38] YANG D X, LIU L M, YANG B. Secure Euclidean distance computation in the presence of malicious adversaries [J]. Computer Engineering and Applications, 2015, 51(24): 109-113.
- [39] KIKUCHI H, NAGAI K, OGATA W, et al. Privacy-preserving similarity evaluation and application to remote biometrics authentication[J]. Soft Computing, 2014(5): 529-536.
- [40] BARNI M, BIANCHI T, CATALANO D, et al. Privacy-preserving fingercode authentication[C]//The 12th ACM workshop on Multimedia and security, September 9-10, 2010, Roma, Italy. New York: ACM Press, 231-240.
- [41] YANG B, LI S D, ZHOU S F. Privacy-preserving online transaction scheme based on El Gamal signature[J]. Application Research of Computers, 2014, 31(12): 3707-3710.
- [42] ZHANG C. Realization of information hiding using picture as carrier[J]. Journal of Mudanjiang Normal University (Sci.&Tech.), 2008(1): 20-21.
- [43] HOWGRAVE-GRAHAM N. Approximate integer common divisors[M]. Berlin: Springer Berlin Heidelberg, 2001: 51-66.
- [44] DIJK M V, GENTRY C, HALEVI S, et al. Fully homomorphic encryption over the integers[C]//International Conference on Theory & Applications of Cryptographic Techniques, May 30- June 3, 2010, French Riviera, France. New York: ACM Press, 2010: 24-43.
- [45] 何文才, 杜敏, 刘培鹤, 等. 基于 Paillier 同态的无线自组网组密钥管理方案[J]. 计算机科学, 2013, 40(10): 114-118.
- HE W C, DU M, LIU P H, et al. Wireless Ad-hoc network group key management scheme based on Paillier homomorphic[J]. Computer Science, 2013, 40(10): 114-118.
- [46] 程帅, 姚寒冰. 基于同态加密的密文全文检索技术的研究[J]. 计算机科学, 2015, 42(6A): 413-416.
- CHENG S, YAO H B. Study of cipher text retrieval based on homomorphic encryption[J]. Computer Science, 2015, 42(6A): 413-416.
- [47] HOWGRAVE-GRAHAM N. Approximate inter common divisors[M]. Berlin: Springer Berlin Heidelberg, 2001.
- [48] LI J, CHEN S C, SONG D J. Security structure of cloud storage based on homomorphic encryption scheme[C]//IEEE International Conference on Cloud Computing & Intelligent Systems, Oct 30-Nov 1, 2012, Hangzhou, China. New Jersey: IEEE Press, 2012: 224-227.

[作者简介]



王永建 (1981-), 男, 东南大学计算机科学与工程学院博士生, 主要研究方向为云安全、大数据处理。



宋爱波 (1970-), 男, 博士, 东南大学计算机科学与工程学院教授、博士生导师, 主要研究方向为云计算、大数据处理。



叶亚伟 (1981-), 男, 中国通信建设集团设计院有限公司工程师, 主要研究方向为云计算。



姜海波 (1981-), 男, 国网河南省电力公司检修公司高级工程师, 主要研究方向为电网系统数据安全。