

云环境下数据隐私保护模型设计

邵必林¹, 赵华龙¹, 边根庆², 王 辉¹

(1 西安建筑科技大学 管理学院, 陕西 西安 710055; 2 西安建筑科技大学 信控学院, 陕西 西安 710055)

摘 要: 提出利用文件存储结构与属性加密的方法, 来解决云存储数据的隐私安全问题, 并给出了一个可以支持用户数据隐私保护的安全模型. 该方法的实质是通过运用属性加密与文件存储结构相结合的方式对用户隐私数据进行加密, 并提供细粒度的访问控制. 该模型可以有效地对用户隐私数据实施保护, 并能提高数据安全保护的等级和可靠性.

关键词: 云存储; 隐私安全; 属性加密; 访问控制

中图分类号: TP393.08

文献标识码: A

文章编号: 1000-7180(2014)06-0036-04

Design on Data Privacy Protection Model for Cloud Environment

SHAO Bi-lin¹, ZHAO Hua-long¹, BIAN Gen-qing², WANG Hui¹

(1 School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China;

2 School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: Considering the issue of the existing cloud storage privacy security, the dissertation proposes the approach of using the file storage structure and attribute-based encryption, to resolve the problem of cloud storage data privacy security, and explore a security model that can support the privacy protection of user data. The essence of the method is to encrypt user privacy data by the way of combining the attribute encryption and file storage structure, and provides Fine-grained access control. The result shows that the model can effectively protect user privacy data, and improve the level of data security protection and reliability through the security model demonstration.

Key words: cloud storage; privacy security; attribute-based encryption; access control

1 引言

随着云计算的不断发展, 云存储服务自身的安全隐患已经随着应用的不断深入逐渐暴露出来, 存储在云环境下的数据安全问题成为亟待解决的重要问题.

在云存储环境下, 针对用户以及企业的隐私信息泄漏、非法篡改等恶意行为, Parakh^[1]提出了一种基于 K-阈值计算分配一个秘密的隐式数据安全方式, 而且这种隐式安全方式不需要使用加密机制. 文献[2]提出了基于云模型进化算法的数据挖掘模型, 该模型可以保证 web 日志挖掘过程中不会泄露用户隐私. 文献[3]等采用数字水印技术和加密协

议, 通过加强对数据使用信息的控制来保障数据的隐私安全问题.

通过对已有结果的分析, 主要存在两方面的问题, 一是云存储密钥管理繁琐; 二是没有充分考虑系统用户解密能力与密文保护的细粒度的访问控制问题. 据于此, 本文提出并采用一个新的加密方案来解决云存储隐私安全保护问题, 其实质是采用多叉树结构构建数据索引, 采用基于属性的加密实现密钥的分发管理.

2 建立系统模型

云存储的体系架构模型, 如图 1 所示. 该模型包括四个实体: 数据所有者(OWNER)、可信的第三方

收稿日期: 2013-08-23; 修回日期: 2013-10-12

基金项目: 国家自然科学基金项目(61272458)

(Trusted Third Party, TTP)、服务提供者(Cloud Service Provider, CSP)和数据使用者(USER)。

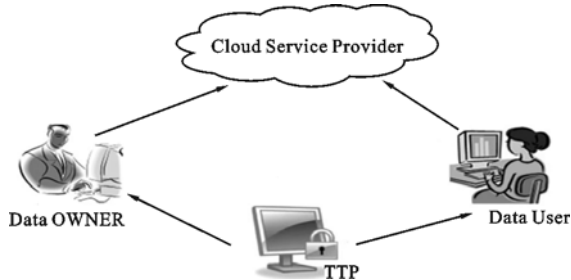


图1 云存储体系架构模型

四个实体的主要功能描述：

Owner, 用加密算法对数据进行加密, 得到的加密文件存储到服务提供者。

User, 只有满足指定的访问策略, 有数据所有者的合法权利才可以访问加密的内容, 并恢复它的原始消息。

服务提供者(CSP), 向用户提供数据存储和检索服务。CSP 验证 User 的权限, 并将其操作限制在允许的范围之内, 然后, 根据允许的 User 访问操作, 提供相应的服务对其进行权限范围内的操作。如果 CSP 是不被信任的, 那么, 寄存于服务器中的数据必须经过加密处理才能保证隐私数据的安全性。

可信的第三方(TTP), 是系统所有参与者的密钥生成中心。它通过产生的公共参数和主密钥, 以及生成特定于用户的私人密钥, 对数据访问、密文解密、匿名密钥数据所有者的属性集实施控制与管理^[4]。

3 云存储模型中涉及的关键技术

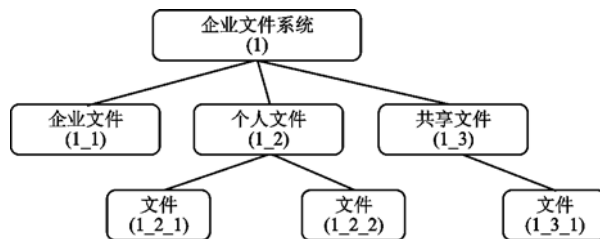
3.1 多叉树型文件存储结构

采用文件的逻辑存储结构存储数据, 用多叉树型表示。如图2所示。采用这种逻辑, 具有真实反映现实、操作简便及效率高等特点。当用户上传文件时, 服务器端(CSP)会自动生成文件的存储结构。

在该文件系统中, 为了更好的保护用户的隐私安全问题, 采用基于属性的非对称加密算法 ABE 将文件加密, 再把文件名改为 $1_3_1 \$ K$ (文件名, $k_{1_3_1}$), 服务提供商 CSP 不知道文件的具体内容和文件名, 只知道这个文件的编号 1_3_1 , 因此可以达到有效地保护用户隐私数据的目的^[5]。

3.2 基于属性加密及密钥管理

基于属性的加密是 Sahai 和 Waters 两位密码学家在 2005 年的欧洲密码年会上发表的《模糊基于身份加密方案》中首次提出来的。是为了提高基于生



物信息身份加密体系的容错能力, 而采用属性集合表示用户的身份, 避免了在基于身份加密体系中仅能用唯一的标示符来表示用户身份的缺陷。基于属性加密的体制也属于公钥加密的体制, 因为采用的都是公钥加密, 私钥解密的加密策略。

访问结构(Access Structure)是访问控制策略的逻辑结构。本文采用了文献[6]中 (t, n) 门限访问结构, n 表示属性的数量。当用户有属性集合与密文集合至少有 t 个属性相同, 用户才有权限解密密文。例如, 用 $1, 2, 3, \dots, 7$ 表示系统属性, 系统中倘若有三个用户分别是: $A(1, 2, 3, 4)$, $B(2, 3, 4, 5)$, $C(1, 4, 5, 6)$, 从认证中心获得各自的私钥, 现一个密文, 以属性集合 $(2, 3, 4, 7)$ 进行加密。访问门限设为 $(3, 4)$, 即是把他们的属性集与密文的属性集做交集, 大于 3 的能够解密。上述例子中, 可以看出 A 与 B 满足解密条件, 能够解密, 用户 C 不能解密密文。同时, 当某一用户少数属性份额丢失或者损毁不会影响他的访问控制, 如用户 A 将属性 1 丢失并不影响其解密密文。为了便于研究, 本文所指访问结构都是单调的属性集合。

基于身份加密体系中公钥是基本确定的, 用户属性集合作为用户的身份, 由可信第三方私钥生成中心(Private Key Generator, PKG)根据用户的属性集生成相应的私钥。该模型的密文不是加密给特定的用户, 而是加密给所有满足一定条件下的用户。如下公式表示用户私钥的生成:

用户私钥 = $f(\text{可信第三方私钥}, \text{用户属性集})$, 式中, f 表示一个函数。在该系统中, 可信的第三方负责管理用户的密钥。密钥的管理方案与文件的存储逻辑结构类似, 如图3所示。根据用户具有的属性集, 当满足限定条件时, 密钥生成中心便根据属性集为用户生成相应的密钥 $K(x)$ 。一般情况下, $t_1 \geq t_{1_2} \geq t_{1_2_1}$ 。

在图3中, 当用户需要访问 1_2_1 文件时, 系统首先会判断该用户需要访问的文件是否在文件系统中, 再判断属性是否满足根门限 $(t_{1_2_1}, n)$, 如果满足的数目达到了结点门限值, 则系统将依据用户的属

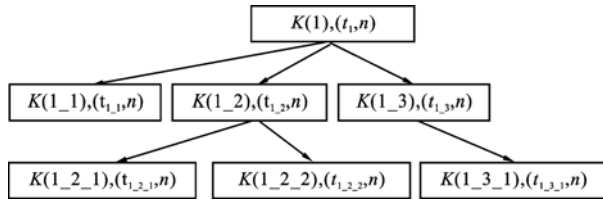


图3 访问树索引结构

性集合计算出私钥 $K(1_2_1)$, 并返回私钥。

4 基于属性的加密算法的实现

本文采取基于属性的动态多重加密算法^[6-8], 适合于大规模属性全集加密方案, 并允许同时加密多个消息, 这样就解决了计算与通信方面的开销, 能有效提高系统的运行效率。

4.1 加密算法描述

在基于属性的动态多重加密方案中, 系统中的一个用户将用一个对 (id, ω') 来标识, 其中 id 指示该用户的身份, ω' 表示用户 id 的属性集合。在设计的模型当中允许不同的用户具有相同的属性集合, 并且一个用户仅能被赋值一个属性集合。一个基于属性的动态多重加密方案由以下几个算法构成:

Setup: 该算法输入安全参数 s , 输出系统参数, 其中包括系统主公钥 PK 和系统主密钥 MK , 它是一个随机化算法。

Key Extraction: 该算法输入系统主密钥 MK 和一个对 (id, ω') , 输出用户的解密密钥 $d(id, \omega')$, 算法是一个概率算法。

Encryption: 该算法输入系统主公钥 PK , 用户属性集合 ω' 和一个需要加密的消息 m , 输出消息的密文 CT 。

Decryption: 该算法输入用户的解密密钥 $d(id, \omega')$ 和密文 CT , 当 $|\omega \cap \omega'| \geq d$ 时输出明文消息 M , 其中 d 为系统的门限参数, ω 是用于产生密文 CT 的属性集合^[9]。

4.2 加密算法的实现过程

设 G_1 与 G_2 是两个以素数 q 为阶的循环群, g 是其生成元, 双线性对 $e: G_1 \times G_1 \rightarrow G_2$ 。对于所有的 $\alpha, \beta \in Z_p$ 都满足 $e(g^\alpha, g^\beta) = e(g, g)^{\alpha\beta}$, 其中 Z_p 为除了零元之外所有的元素集合:

$$\Delta_{i,S}(x) = \prod_{j \in S, i \neq j} \frac{x-i}{i-j} \pmod{q},$$

式中, $\Delta_{i,S}(x)$ 为拉格朗日参数, $i \in Z_p$, S 属于 Z_p 中任意 L 个属性的集合。

初始设置: 设门限值为 d , 所有属性集合 ω 的大小为 k , 允许同时加密消息的最大规模为 n , Λ 表示

系统松弛属性的集合, $|\Lambda| = n$, h 表示群 G_1 中的随机元素, 同时设集合 $N = \{1, 2, \dots, n+k+1\}$, 在群 G_1 中随机选取一组 $t_1, t_2, \dots, t_{n+k+1}$ 。定义函数 $T(x)$ 如下:

$$T(x) = h^{x^n} \prod_{i=1}^{n+k+1} t_i^{\Delta_{i,N}(x)}$$

(1) 在 Z_p 中随机选取 $\alpha, \beta \in Z_p$, 其中 $\alpha = \{u_0, u_1, \dots, u_{n-1}\}$ 和 $\beta = \{\omega_0, \omega_1, \dots, \omega_{n-1}\}$ 。

(2) 设 $g_1 = g^{\omega_0}, g_2 = g^{\omega_1}, \dots, g_n = g^{\omega_{n-1}}$ 。

系统公共参数和系统主密钥组成如下:

$PK = \{g_1, g_2, \dots, g_n, h, t_1, t_2, \dots, t_{n+k+1}\}$, $MK = \{\omega_0, \omega_1, \dots, \omega_{n-1}\}, t_1, t_2, \dots, t_{n+k+1}$ 是在群 G_1 中随机选取的。

密钥生成: 算法根据用户拥有的属性集合 ω' , 密钥生成中心通过拉格朗日插值公式构造一个 $n+k-1$ 次多项式 $q(x)$, 生成对于属性集合 ω' 的私钥:

当 $i = 0, 1, \dots, |\omega'| - 1$ 时, $D_i = h^{q(i)} T(i)^{r_i}$, $d_i = g^{r_i}$ 。

当 $j = 1, 2, \dots, n$ 时, $D_j' = h^{q(\omega_j)} T(j)^{r_j}$, $d_j' = g^{r_j}$ 。

加密: CT 为属性集合为 ω 加密后的密文, M_i 为需要加密的信息, W 为 β 中的值, 下面式子为密文的格式:

$$CT = \{(E_i = T(i)^s)_{i \in \omega},$$

$$(E_j' = T'(j)_j^s)_{j \in \Lambda'},$$

$$(C_v = M_v W_v^s)_{v=0}^{n-1},$$

$$E^s = g^s, \omega\}.$$

解密算法: 系统的门限值为 d , 密文的属性集合为 ω , 用户的属性集合为 ω' 。若满足 $|\omega \cap \omega'| \geq d$, 则用户可以解密。用户可以通过下面公式解密密文:

$$M_v = C_v \prod_{i \in S} \left[\frac{e(d_i, E_i)}{e(D_i, E_i)} \right]^{\Delta_{i,S}(v)} \prod_{j \in \Lambda'} \left[\frac{e(d_j', E_j')}{e(D_j', E_j')} \right]^{\Delta_{j,S}(v)}.$$

5 模型安全与效率分析

5.1 模型安全论证

为了证明的该方案的安全性, 首先输入一个安全参数 s 对系统进行初始化, 然后假设对一份机密文件 F 进行加密, 其在安全查看的形式应为 $(E_i = T(i)^s)_{i \in \omega}, (E_j' = T'(j)_j^s)_{j \in \Lambda'}$, 现在需要构建一个非法用户 A , 非法用户 A 对加密的文件进行攻击解密。假如非法用户 A 在任何多项式时间内中获得成功的优势是不可忽略的, 则系统不安全, 反

之则安全. 非法用户 A 根据协议的制约, 应该拥有足够的属性才可以破解密钥来解密加密的文件 F. 非法用户 A 提交两个等长的 M_1, M_2 发送至模拟器, 定义 λ' 为模拟器根据非法用户的猜想值正确与否, 假如猜想正确模拟器输出 $\lambda' = 0$, 表明给出一个既定的元组, 否则模拟器输出 $k' = 1$, 表明给出一个随机元组. 在这个模拟的过程中, 非法用户 A 可以获得成功解密的概率为

$$\begin{aligned} \text{adv}(s) &= \left| \frac{1}{2} \Pr[\lambda' = \lambda \mid \lambda = 1] + \right. \\ &\quad \left. \frac{1}{2} \Pr[\lambda' = \lambda \mid \lambda = 0] - \frac{1}{2} \right| \\ &= \frac{1}{2} \left| \frac{1}{2} + \left\{ \frac{1}{2} + \epsilon \right\} - 1 \right| \\ &= \frac{\epsilon}{2} \end{aligned}$$

因此, 非法用户 A 的优势概率是可以忽略的, 证明模型是安全可行的.

5.2 模型效率分析

下面将所提出的方案与 Muller、Bethencourt 和 Sahai、Waters 的方案在性能方面进行对比分析. 假设需要对 m 个文件同时进行加密, 分析各方案的初始算法和密钥生成算法, 再运行加密与解密算法. 四个不同方案的性能比较如表 1 所示.

表 1 模型效率分析

	Muller	Bethencourt	ABE (Sahai, Waters)	Our scheme
初始化	$O(n)$	$O(2n)$	$O(n)$	$O(n+m)$
密钥生成	$O(3n)$	$O(2n)$	$O(n)$	$O(n+m)$
加密	$O(3n)$	$O(2n)$	$O(n \cdot m)$	$O(n+m)$
解密	$O(1)$	$O(2n)$	$O(1)$	$O(1)$

本方案在初始化与密钥生成性能显示略低于 Sahai 和 Waters 两人提出的基于属性加密算法, 但是在加密方面有着优势; 与 Muller 等与 Bethencourt 等的方案相比, 文中提出的方案在初始化、密钥生成、加密和解密各个方面均有着明显的优势.

6 结束语

本文所建立的云环境下的数据隐私保护模型, 是以可信的第三方完全可信为条件的, 其实质是将

可信第三方、文件索引结构和基于属性加密体系进行有机融合. 研究表明, 所建立的模型对解决云存储中数据隐私的安全保护问题, 具有合理性与有效性.

参考文献:

- [1] Parakh A, Kak S. Space efficient secret sharing for implicit data security[J]. Information Sciences, 2011, 181(2):335-341.
- [2] Jiang Changbin, Chen Li. Research on privacy preserving data in web log mining[C]//2010 2nd International Symposium on Information Engineering and Electronic Commerce (IEEC). Cardiff City Hall, 2010:1-4.
- [3] Tharaud J. Privacy by data provenance with digital watermarking[C]//Proceedings of the 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Germany, Darmstadt, 2010:510-513.
- [4] Dimitrios Zissis, Dimitrios Lakkas. Addressing cloud computing security issues[J]. Future Generation Computer Systems, 2012, 28(3):583-592.
- [5] 黄汝维, 桂小林. 支持隐私保护的云存储框架设计[J]. 西安交通大学学报, 2011, 45(10):59-63.
- [6] 郭振洲. 基于属性的加密方案的研究[D]. 大连:大连理工大学, 2012.
- [7] 王永涛. 基于属性密码体制的相关研究[D]. 上海:上海交通大学, 2011.
- [8] 杨晓峰. 基于 (t, n) 门限的数字水印协议研究[D]. 曲阜:曲阜师范大学, 2011.
- [9] Sahai A, Waters B. Fuzzy identity-based encryption[C]//Advances in Cryptology(EUROCRYPT 2005). Lecture Notes in Computer Science(3494). Berlin:Springer-Verlag, 2005:457-473.

作者简介:

邵必林 男, (1965-), 教授, 博士生导师. 研究方向为信息管理技术、存储安全技术等.

赵华龙 男, (1988-), 硕士. 研究方向为云计算技术和信息安全技术.

边根庆 男, (1968-), 副教授, 硕士生导师. 研究方向为信息管理技术、存储安全技术等.

王 辉 男, (1988-), 硕士. 研究方向为云计算技术和信息安全技术.