



基于指纹魔方算法的云存储数据保护机制^{*}

吴昊¹, 范九伦¹, 刘建华², 张俊生¹

(1. 西安邮电大学通信与信息工程学院 西安 710121; 2. 西安邮电大学信息中心 西安 710121)

摘要:云存储有效解决了存储性能瓶颈,同时带来了数据安全和隐私保护难题。为了解决这一问题,提出了一种基于指纹魔方算法的云存储数据保护机制。该机制通过用户的指纹特征值控制魔方旋转对文件进行加密,再利用门限分割技术将文件分割成小块存储到各个服务器中。需要恢复原文件时,先对各个服务器中的文件块进行完整性验证,找到不少于门限值数量的文件块就能完整恢复文件。该机制保护了云存储用户的隐私数据,同时提高了云存储系统的抗破坏能力和灵活性。

关键词:云存储;数据安全;指纹魔方算法;门限分割

doi: 10.3969/j.issn.1000-0801.2014.11.019

Cloud Storage Data Protection Mechanism Based on a Fingerprint Cube Algorithm

Wu Hao¹, Fan Jiulun¹, Liu Jianhua², Zhang Junsheng¹

(1. School of Communication and Information Engineering, Xi'an University of

Posts and Telecommunications, Xi'an 710121, China;

2. Information Centre, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: Cloud storage is an effective solution to the storage performance bottleneck, but it also gives rise to problems of data security and privacy protection. In order to solve this problem, a cloud storage data protection mechanism based on a fingerprint cube algorithm was developed. Mechanism through the user's fingerprint characteristic value control cube rotation to encrypt files, and then use threshold segmentation technique to split the file into small blocks to each storage server. If it is expected to restore the original file, check the integrity of each block at first, and then find no less than the threshold number of file blocks, which leads to recover the full document. The mechanism protects the privacy of user data stored in the cloud, but also improve the anti-destruction capability and flexibility of cloud storage system.

Key words: cloud storage, data security, fingerprint cube algorithm, threshold segmentation

1 引言

随着计算机和互联网技术的发展,人们对计算机存储的要求越来越高,传统的集中式存储已不能满足用户需求,云存储技术作为一种新兴的基于网络的存储技术应运

而生。它旨在通过互联网为用户提供更强的存储服务,将网络中大量处于不同计算机、不同类型的存储设备通过网络和应用软件集合起来协同工作,共同对外提供数据存储和业务访问,用户可以随时随地通过网络访问“云”中的数据,从而有效地解决了存储的性能瓶颈问题。然而,在云存

^{*} 陕西省科学技术研究发展计划基金资助项目(No.2010k17-07),大学生创新创业训练计划基金资助项目(No.201211664012)

存储服务给人们带来极大便利的同时,安全问题也日益凸显。在云存储系统中,用户所属的数据外包给云服务提供商,云服务提供商就获得了该数据或应用的优先访问权,由于存在内部人员失职、黑客攻击及系统故障导致的安全机制失效等多种风险,云服务商没有充足的证据让用户确信其数据被正确地存储和使用^[1,2]。为了解决这一问题,研究人员已经提出了一些解决办法。参考文献[3]提出了一种基于数据分割与分级的云存储数据隐私保护机制,按不同的安全级别需求将数据分割为大小数据块,联合采用数据染色和不同强度的数据加密技术进行数据染色或加密。由于文件块大小的差异,黑客有可能直接从大数据块中恢复出原文件的隐私信息;参考文献[4]提出了一种基于 ID 特征码的云存储数据分片与备份策略,利用文件块 ID 和可用存储节点 ID 的对应关系随机分配有效存储节点。该方法解决了服务器的负载均衡问题,但建立多个数据副本增加了存储成本。本文提出一种基于指纹魔方算法的云存储数据保护机制,运用魔方加密算法对存储文件进行加密处理,同时将门限分解思想加入保护机制中,不仅使数据泄露问题得到了有效解决,还能有效防止数据被破坏或者丢失,系统存储效率高,数据冗余度低。

2 基于指纹魔方算法的云存储数据保护机制

2.1 文件上传流程

整个数据保护机制包括文件上传和下载两个过程。文件上传流程分为 4 个步骤,如图 1 所示,具体介绍如下。

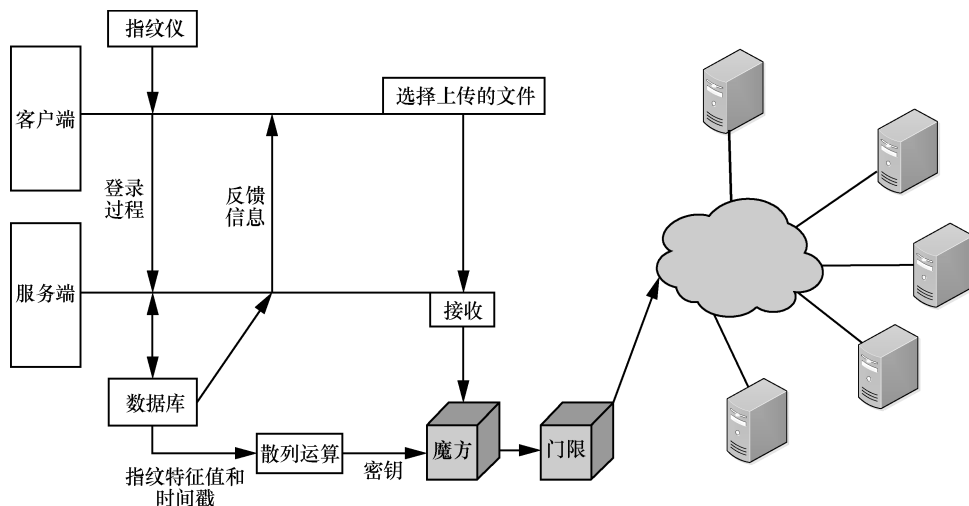


图 1 文件上传流程

- 文件传输:由客户端向服务端发送连接请求,服务端允许连接后,客户端发送文件,由服务端接收模块开辟一个线程负责接收文件流。
- 文件接收:服务端根据客户端请求,接收上传的文件。
- 魔方加密:把文件按字节附于魔方表面,运用用户指纹特征值和当前时间戳取得散列值,由该值控制魔方的旋转面、旋转方向和操作轮数。
- 门限存储:依据所设计的门限方案,将文件分块,每块增加头部信息,存储在“云”中不同的服务器主机上。

2.2 文件下载流程

文件下载主要分为 5 个步骤,如图 2 所示,具体介绍如下。

- 文件选择:选择已上传到服务器中需要下载的文件。
- 门限重组:对各文件做散列运算,检查服务器上各文件是否被破坏,根据参数做出文件分割表,组合出完整文件。
- 魔方解密:把组合出的文件附于魔方表面,同样利用用户指纹特征值和文件存储时间取得散列值,按加密时的逆方向旋转魔方解密文件。
- 文件传输:由客户端向服务端发送连接请求,服务端允许连接后,客户端接收文件。由服务端发送模块开辟一个线程负责发送文件流。
- 文件接收:服务器将客户端请求下载的文件发送至客户端,客户端接收文件。

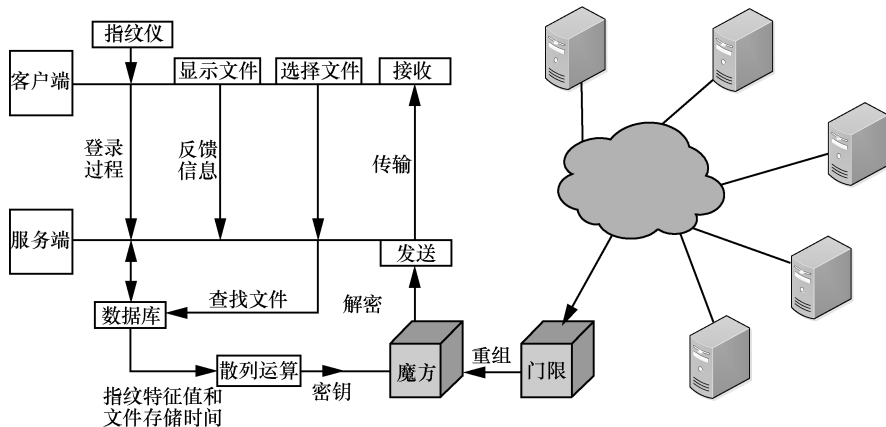


图2 文件下载流程

3 魔方算法

3.1 魔方加密原理

一个 N 阶魔方是指其长、宽、高均为 N 个单位小方格的正方体,与魔方某个表面平行的平面上的各小方格构成一个可转动的层面。魔方加密的原理就是利用魔方旋转变化的不确定性,用数学方法把魔方的旋转规则模拟出来^[5]。一个 N 维魔方旋转 M 次,会出现的旋转方式总数可以由式(1)得出,其中 $R=\lfloor M/4 \rfloor, L=M \bmod 4, Total$ 表示不同的旋转种数。

$$Total = \begin{cases} (3N)^M - (3N)^R C_{3N}^L, & M \geq 4 \\ (3N)^M, & M < 4 \end{cases} \quad (1)$$

由式(1)可知,随着 N 或者 M 的增长, $Total$ 将迅速增加。这样,只要增加魔方的维数或者增加旋转的次数就可以大幅度提高加密算法的安全性。本文设计的云存储数据保护机制,就是利用魔方加密的原理,将要存储的文件分割成 $6 \times N^2$ 份,写入 N 阶魔方的“小方格”里,再利用用户个人的指纹特征值控制魔方旋转来加密文件,这样可以保证只有用户自己能解密自己存储的文件,大大提高了用户数据的安全性。

3.2 魔方加密算法的实现

以三阶魔方为例,三阶魔方共有 6 个面,分别记作:顶、底、左、右、前、后。对于不同的面分别有两种基本的操作:顺时针旋转 90° 、逆时针旋转 90° 。其他选择方式均可由这两种组合得到,因此共有 $6 \times 2 = 12$ 种控制方式。这 12 种控制方式分别对应一个操作函数。而控制系统选择哪一个函数进行旋转操作,取决于用户注册时的指纹特征值和当前系统时间。这样既能保证加密密钥为用户独有,

又能保证用户在不同时刻密钥不同,防止黑客的窃听。具体的实现步骤如下。

(1) 用户注册时,通过客户端提取指纹特征值 $Fingerprint$ 。

(2) 将客户端上传的文件分成 $6 \times 9 = 54$ 块,使之与魔方表面的 54 个分块一一对应。

(3) 取当前系统时间 $time$, 计算 $Key = Hash(Fingerprint || time)$, $Key[i]$ 分别代表密钥的第 i 字节数据, $i = 1, 2, 3, \dots, 32$ 。

(4) 对于 $i = 1, 2, 3, \dots, 32$, 每一个 $Key[i]$ 计算控制数值 $res = Key[i] \bmod 12$ 。根据表 1 进行不同的旋转操作(解密时,对应加密的操作逆向还原)。

表1 魔方操作控制

操作名称	控制数值	还原操作控制
上表面顺时针	0	1
上表面逆时针	1	0
下表面顺时针	2	3
下表面逆时针	3	2
左面顺时针	4	5
左面逆时针	5	4
右面顺时针	6	7
右面逆时针	7	6
正面顺时针	8	9
正面逆时针	9	8
背面顺时针	10	11
背面逆时针	11	10

(5) 将加密后的文件块分别存储在不同的服务器主机上。

该加密方案在多轮加密的情况下,可以形成足够数量级的加密组合,从而能够抵挡穷举攻击的威胁。另外,魔方的维度可以任意扩展,从而提高加密强度。

4 门限方案

4.1 门限方案的意义

由于云存储系统所部署环境的复杂性,在实际应用中数据可能会受到一定程度的损坏,此外,黑客的攻击也不可避免,这就更需要在存储以及传输时对每个文件块进行完整性验证^[6,7]。因此,云存储保护机制引入门限方案,利用多个服务器中存储文件块存在的冗余,通过这些冗余补齐丢失或者破坏的文件块,获得所有源文件数据后重组文件。这样,当云存储服务器集群中的某一台或几台服务器的文件被破坏后,仍可以通过其他服务器上的文件碎片恢复完整的用户文件。

4.2 存储文件头

为了保证文件可以实现逆向恢复,需要对所有服务器上的文件块增加一些头部信息,见表 2。

4.3 门限算法

门限算法分为文件分割和文件恢复两部分,分别在用户上传和用户下载时使用。

4.3.1 分割算法

分割算法的具体步骤如下。
(1)设文件大小为 $FileSize$ byte。对于 (n, t) 门限,记

$U = \binom{n}{t}$, n 代表可供存储的服务器数, t 表示未被破坏的服务器数阈值, U 表示分割的文件块数。

(2)计算 $PartSize = FileSize / U$, 作为每台服务器存储文件的大小。若 $FileSize \neq k \times U, k_1 \in N$, 在文件末尾填充 $FileSize - \lfloor FileSize / U \rfloor \times U$ 个字节 0x00。若 $FileSize = k \times U, k \in N$, 继续。

(3)填充如表 1 所示的各文件块的头信息,其中散列值待文件组合后再计算。

(4)依据 (n, t) 值设计分割表,以 $(5, 3)$ 门限为例,文件块存储情况见表 3。每一行代表 1 台服务器,这里有 5 台,分别是 A、B、C、D、E。然后对每个文件进行分块,事实上在经过第 (2) 步的处理后,文件正好能被平均分成 U 份,这里 $U=10$ 。表 3 中 $\checkmark$ 表示该文件块在相应的服务器中有存储,空白代表不存储。

(5)根据分割表对文件进行分割,将属于不同服务器的文件块重新组合。

(6)将文件分别存储到各服务器上。

4.3.2 恢复算法

恢复算法的具体步骤如下。

(1)获取门限参数 (n, t) 、文件大小 $FileSize$, 计算 $U = \binom{n}{t}$ 。

表 2 文件头信息

参数	大小/byte	说明	默认值
start	4	开始标识	0xAA
head_size	2	文件头大小	48
N	1	门限参数 n	由管理员配置
T	1	门限参数 t	由管理员配置
org_size	4	源文件大小	单个文件不大于 4 GB
each_part_size	4	分割后各部分大小	
each_file_size	4	各服务器所存大小	
pad_size	4	填充字节	
file_type[4]	4	源文件类型	
hash[HASH_LEN]	16	本服务器文件散列值	
end	4	结束标识	0xAA

表 3 文件块存储情况

	1	2	3	4	5	6	7	8	9	10
A	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$						
B	$\checkmark$				$\checkmark$	$\checkmark$	$\checkmark$			
C		$\checkmark$			$\checkmark$			$\checkmark$	$\checkmark$	
D			$\checkmark$			$\checkmark$		$\checkmark$		$\checkmark$
E				$\checkmark$			$\checkmark$		$\checkmark$	$\checkmark$



(2)对各文件做散列运算,检查服务器上各文件是否被破坏。文件破坏的服务器 $\in A=\{SET1| \text{文件被破坏}\}$,未被破坏的服务器 $\in B=\{SET2| \text{文件完整}\}$ 。

(3)若被破坏文件数大于 $n-t$,超过门限值,无法恢复文件,结束。否则继续。

(4)对于文件块 $1\sim U$,获取集合 B 中服务器上的文件块,恢复完整文件。

4.3.3 系统冗余度

假设文件大小为 $FileSize$ byte,门限参数为 (n,t) 。将文件分成 $U=\binom{n}{t}$ 块,每块的大小为 $PartSize=FileSize/U$ byte。对于每台服务器来说,存储的文件块数为 $PartCnt=U-U\times(n-t)/n$ 。所以实际上,每台服务器存储的文件大小是源文件大小的 $PartCnt/U=t/n$ 。若使用 $(5,3)$ 门限,就只有 60% 的占用率。

5 系统实现和测试

系统测试搭建了一个小型云平台作为实验环境,选取 1 台 PC 作为客户端,5 台服务器作为云存储服务,对几个不同大小的数据文件的上传和下载速度进行了测试。

(1)文件上传和下载速度测试

实验选择了 4 个文件,大小分别为 10 MB、100 MB、500 MB、1 GB,记录每个文件上传和下载所花费的时间,测试结果如图 3 所示。测试中,文件平均下载速率为 1.9 MB/s,上传速率为 1.4 MB/s,与整个文件上传下载相比,本方案增加了数据分割和加密操作,产生了一定的额外时间开销,经计算时间开销为 $O(n)$, n 为文件块数,增加的时间开销对系统影响不大,但分割和加密操作大大增强了数据隐私的安全性。

(2)文件分块破坏恢复情况

实验采用 $(3,5)$ 的门限方案进行测试,对存储在 5 台服

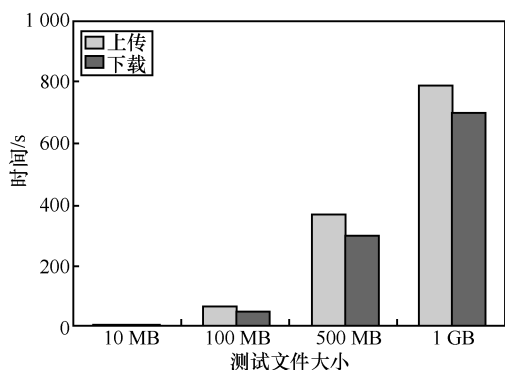


图 3 文件上传和下载速度测试结果

务器上的文件分块进行不同程度的破坏,看系统能否恢复出完整文件。测试结果见表 4。

实验结果表明:在 $(3,5)$ 门限方案下,当被破坏文件块数小于 3 时,系统仍能完整地恢复用户文件;当被破坏文件块数大于或等于 3 时,系统不能恢复用户文件,测试结果完全符合系统设计目标。

表 4 文件块破坏恢复测试

总文件块数(n)	未破坏的文件块数	破坏的文件块数	恢复结果
5	5	0	成功
5	4	1	成功
5	3	2	成功
5	2	3	失败
5	1	4	失败

6 结束语

用户数据的安全问题已经成为云存储应用发展的瓶颈。本文提出了一种基于指纹魔方算法的云存储数据保护机制,利用用户的指纹特征值控制魔方旋转对用户文件进行加密存储,具有算法效率高、安全性高的特点。另外,利用门限分割技术提出了一种安全云存储策略,使得云端各服务器分担系统负载,并且使得任意指定数目的存储服务器均能恢复出用户上传的文件,避免了个别服务器遭到破坏后无法恢复文件的情况。机制从存储策略和数据本身都对用户数据安全进行了有效防护,也在一定程度上降低了数据冗余,提高了系统的性能。

参考文献

- 冯登国,张敏,张妍等. 云计算安全研究. 软件学报, 2011, 20(1): 71~83
- 邹德清,金海,卫中等. 云计算安全挑战与实践. 中国计算机学会通讯, 2011, 7(12): 55~61
- 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. 计算机科学, 2013, 40(2): 98~102
- 钱进进,凌捷. 基于 ID 特征码的云存储数据分片与备份策略. 微电子学与计算机, 2013, 30(8): 16~22
- 杨志国. 基于魔方的混合加密算法及应用研究. 兰州大学硕士学位论文, 2010
- 于洋洋,虞慧群,范贵生. 一种云存储数据完整性验证方法. 华东理工大学学报, 2013, 39(2): 211~216
- 曹夕,许力,陈兰香. 云存储系统中数据完整性验证协议. 计算机应用, 2012, 32(1): 8~12

[作者简介]



吴昊,男,西安邮电大学讲师,主要研究方向为信息安全。



刘建华,男,西安邮电大学高级工程师,主要研究方向为信息安全。



范九伦,男,博士,西安邮电大学教授,主要研究方向为信号处理和信息安全。



张俊生,男,主要研究方向为信息安全。

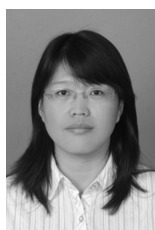
(收稿日期:2014-08-20)

(上接第 109 页)

[作者简介]



邹维福,男,国网泉州供电公司高级工程师,主要研究方向为信息安全管理,主要从事网络与信息安全研究工作。



张素香,女,国家电网公司信息通信分公司副教授、博士后,主要研究方向为通信与信息系统、智能用电等。



张翼英,男,国家电网公司信息通信分公司教授级高级工程师、博士后,主要研究方向为物联网及其安全、安全密钥、智能电网等。



杨成月,男,博士,南瑞集团公司(国网电力科学研究院)厦门亿力吉奥信息科技有限公司副总经理、高级工程师,主要从事网络安全、电力 GIS 与物联网相关信息化工作。

(收稿日期:2014-07-29)

勘误说明

本刊 2014 年第 10 期 7 页《LTE 多天线技术发展趋势》的作者简介有误。第二作者谢伟良与第三作者陈鹏的照片应互相调换。

对此,我们向作者及读者深表歉意。

《电信科学》编辑部
2014 年 11 月