第45卷 第4期 西 安 交 通 大 学 学 报 Vol. 45 No. 4
2011年4月 JOURNAL OF XI′AN JIAOTONG UNIVERSITY Apr. 2011

面向分散式存储的云存储安全架构

边根庆[1]，高松[1]，邵必林[2]
（1.西安建筑科技大学信息与控制工程学院，710055，西安；
2.西安建筑科技大学管理工程学院，710055，西安）

摘要：针对云存储系统和应用过程中的数据安全性问题，提出了一种面向分散式存储的云存储安全架构.该架构采用信息扩散法、分散存储管理、数据自举恢复等技术，分层实现存储数据在云存储中完成应用系统的数据安全存储管理和传输.该方法通过检查可用片重新计算数据片中所有的数据，若存在受损数据，则根据互为冗余的存储设备数据加以恢复，从而提高数据的可用性，从数据传输到存储，都建立了相应的保护措施进行云存储层与其他层间的安全防范，实现了数据的有效防护.仿真测试表明，该架构在保证数据高安全性的同时提高了系统的整体性能.

关键词：云存储；安全架构；分散式存储管理
中图分类号：TP393.08 **文献标志码**：A **文章编号**：0253-987X(2011)04-0041-05

Security Structure of Cloud Storage Based on Dispersal

BIAN Genqing[1], GAO Song[1], SHAO Bilin[2]
(1. School of Information and Control Engineering, Xi′an University of Architecture and Technology, Xi′an 710055, China;
2. School of Management, Xi′an University of Architecture and Technology, Xi′an 710055, China)

Abstract: A storage security architecture based on distributed storage is presented to deal with the data security in the cloud storage systems and applications. The mechanism of safety management and transmission of storage data are realized by layer through the use of information dispersal algorithms (IDA), distributed storage management, and data restore methods. The method improves the high availability of data by examining the usable slices, recalculating all data in data segment, and recovering damaged data according to mutual redundant storage equipment data. Protection measures from the data transmission to storage are taken to ensure safety between cloud storage layer and other layers such that effective data protection is realized. Simulation tests show that the architecture improves the overall performance of the system while ensures the high security of data.

Keywords: cloud storage; security structure; dispersal storage management

随着信息科技的现代化发展，人们对计算能力的要求也不断提高，作为与计算密不可分的存储技术，也伴随着计算模型的升级，从最初的单机存储、网络存储、分布式存储发展到现在的云存储.云存储是在云计算概念上延伸发展出来的一个新概念，是实现云计算的系统架构中重要组成部分之一.与云计算类似，云存储是指通过集群应用、网格技术或分布式文件系统等功能，将网络中大量不同类型的存储设备通过虚拟化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功能[1]，是对虚拟化存储资源的管理和使用.

尽管很多研究机构认为云计算提供了可靠安全的数据存储中心，但安全问题依然是云存储中存在的主要问题之一.从用户角度考虑，数据的保存都交

收稿日期：2010-09-21. 作者简介：边根庆(1968－)，男，副教授. 基金项目：国家自然科学基金资助项目(61073196).
网络出版时间：2011-02-21 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20110221.1605.008.html

给云存储供应商,因此数据的可用性及安全性也成为云存储系统突出的问题.在2010年3月召开的云计算中国峰会(The Cloud Computing China Congress-CCCC 2010)中指出[2],随着云计算技术的逐步成熟,云安全问题将日益突出,云计算的数据安全正在成为人们关注的重要问题.

目前,在国内外的研究中,对云存储安全方面的研究还比较少.Bowers等[3]提出了分布式加密系统,Cachin等[4]通过使用加密工具来解决数据完整性和一致性问题,研究数据可恢复机制,典型的包括Weatherspoon的Antiquity[5]与Kotla的SafeStore[6],Antiquity是OceanStore的最新改进版本,它被设计用于文件系统和备份的应用程序存储服务系统.国内清华大学、华中科技大学、国防科技大学等科研院校也开始在云存储技术相关领域进行基础性研究工作.如何在复杂的网络环境中保障数据发布及存储服务中的隐私,实现云存储对用户数据的安全性与可信性,是目前亟需解决的问题.

1 云存储系统安全威胁分析

1.1 云存储系统结构

从实际应用和服务的角度考虑,云存储首先利用了网络,其次它可以按需分配,此外它的虚拟化主要用于存储和数据管理.与传统的存储相比,云存储不仅是一个硬件,而且是一个由网络设备、存储设备、服务器、应用软件、公用访问接口、接入网和客户端程序等多个部分组成的复杂系统.各部分以存储设备为核心,通过应用软件来对外提供数据存储和业务访问服务.云存储系统的体系结构有以下4层[7],如图1所示.

(1)存储层是云存储最基础的部分,它由各种各样的存储设备和网络设备组成.同时,还有一个存储管理系统,负责对硬件设备的集中管理、状态监控以及维护升级等.

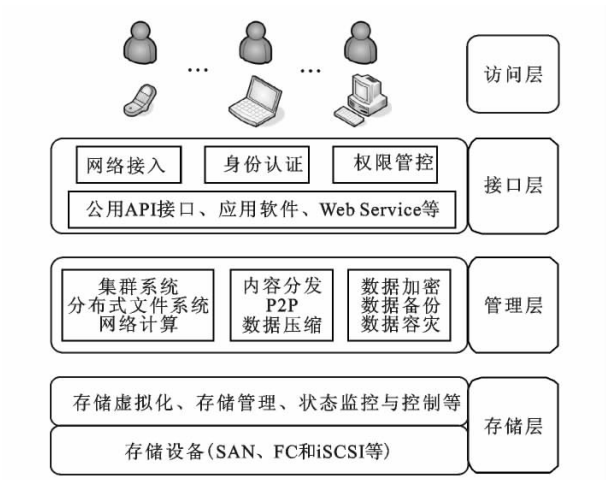

图1 云存储系统结构

(2)基础管理层是云存储最为核心的部分,也是最复杂的部分.基础管理层大量采用了集群管理技术和分布式存储系统的成熟方法,在实现良好的可扩展性的同时,也满足了可用性及性能的需求,它还负责数据加密、备份及容灾等任务.

(3)应用接口层是利用云存储资源进行应用开发的关键部分.云存储供应商通过应用接口层对客户提供统一的协议和编程接口,以进行应用程序的开发.通常这种协议都是基于网络的跨平台协议.

(4)访问层是基于云存储开发的应用程序的入口.任何一个授权用户都可以通过标准的公用应用接口来登录云存储系统,共享云存储所提供的服务.

1.2 云存储系统安全分析

灵活性、易于使用的服务和易于共享基础设施是云计算的优势,然而数据通过互联网在各层之间进行传输并存储,用户对于敏感数据存取时,无法对风险进行直接控制.可以说,云存储自身的特点决定了它在现有的技术方面存在一些安全问题[8],具体表现如下.

(1)传统的安全域划分无效.由于云存储中服务必须是可伸缩的,对外部来讲并不是透明的,因此云存储中无法清晰地定义安全边界及保护设备,为具体保护措施的实施增加了一定的难度.

(2)云存储是通过网络来传输数据的,其中包括网络中的恶意攻击等造成的服务中断、数据破坏、信息被窃取和篡改等,对实现数据的安全存储造成一定的影响,数据的安全通信、访问认证与保密性也是有待解决的问题.

(3)数据存储的安全性防护包括最终存储数据的存放位置、数据完整性、数据间分散存放等.此外,即使数据采用加密技术,也只是在网络上加密传输,数据在处理和存储时也需要保护.

(4)数据的可靠性、可用性.数据在存储系统的容错性、可恢复性和完整性面临一些问题,如何避免在灾难(停电、地震、水灾、火灾等)发生时带来的服务中断乃至数据介质被直接破坏等问题.

(5)如何实现数据之间的逻辑卷管理、存储虚拟化管理和多链路冗余管理将会是一个巨大的难题,也将是整个云存储架构的性能瓶颈,而且还会带来后期容量和性能扩展难等一系列问题.

由此可见,数据的安全性问题贯穿于整个云架

构的各个层次，单独讨论云存储在某一层中的安全性是毫无意义的. 总体而言，对该方面的研究存在两种思路：①借鉴信息安全的 C. I. A 特性（机密性、完整性、可用性），为某一特定应用提出专门的实现思路（如增强文件服务器的安全性、客户端加密文件系统、对磁盘磁带全盘静态加密、客户端直接访问磁盘的认证机制等），即将适用于信息安全的措施（如加密技术、完整性技术）移植到存储系统中；②从存储系统的体系结构入手，寻找安全高效的网络存储与安全管理模式. 为此，本文设计了一种面向分散式分片存储管理来解决云存储和应用过程中的数据安全性问题.

2　云存储安全架构设计

根据云存储中数据的安全性分析，从数据传输到存储，都需要建立相应的保护措施进行层与层之间的防范. 按照云存储的层次结构，通过不同的保护策略逐层对需要存储的数据进行保护，从而实现从数据传输到存储位置的全面防护. 该安全架构采用信息扩散法、分散式存储管理、数据自举恢复等技术，分层实现存储数据在云存储中完成应用系统的数据安全存储管理和传输，其系统总体安全框架结构如图 2 所示.

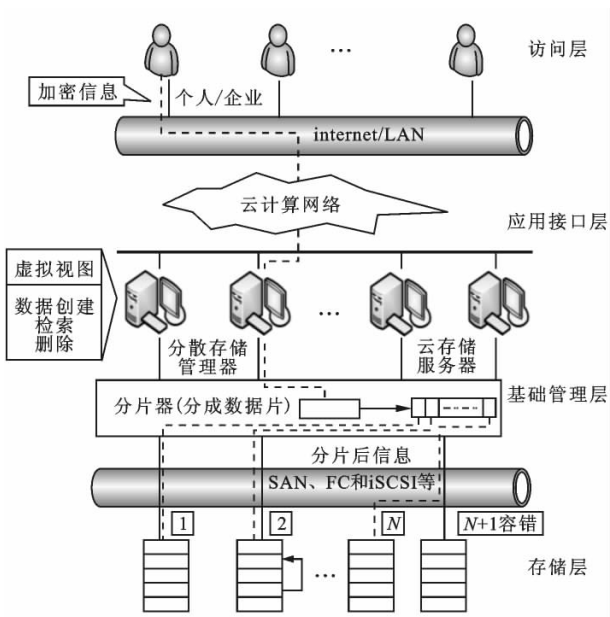

图 2　面向分散式存储的云存储安全架构

2.1　访问层到应用接口层的设计

通过访问控制与身份认证，采用加密技术 SSL 对用户存储数据进行保护，使数据在网络传输中得到较为安全的保障. 用户与云存储服务器相互认证，对双方安全证书和身份进行鉴别，成功后用户代理通过安全 API 和云通信连接进行数据存储服务.

2.2　基础管理层设计

Rabin 提出的信息扩散法[9]（Information Dispersal Algorithms，IDA）将一个长 $L=|\boldsymbol{F}|$ 的文件 $\boldsymbol{F}$ 分成 n 片，每片的长度 $|\boldsymbol{F}_i|=L/m$，$1\leqslant i\leqslant n$，使得 m 片文件片段可以重构文件 $\boldsymbol{F}$，且分块文件的长度总和为 $(m/n)L$[10].

算法具体步骤如下.

步骤 1　设文件 $\boldsymbol{F}$ 由字符串组成，$\boldsymbol{F}=b_1,b_2,\cdots,b_N$，其中字符 $b_i\in[0,\cdots,B]$ 且 $b\in Z(B=2^l-1,l$ 为一个字节拥有的位数）；取最小的大于 B 的素数 p 作为该文件中字符运算的模，该字符串中所有元素上限为 Z^p.

步骤 2　选择一个合适的整数 m，令 $n=m+k$，其中 $m/n<1+\varepsilon,\varepsilon>0$.

步骤 3　选择 n 个向量 $\boldsymbol{a}_i=(a_{i1},a_{i2},\cdots,a_{in})\in Z_p^m$，$1\leqslant i\leqslant n$，使 n 个向量中任意不同的 m 个向量非线性相关. 从 $(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$ 中随机抽取 m 个向量，这 m 个向量是非线性的可能性非常大.

步骤 4　利用选择的向量组对 $\boldsymbol{F}$ 进行拆分. 把文件 $\boldsymbol{F}$ 分成长度为 m 的序列，则

$$\boldsymbol{F}=(b_1,b_2,\cdots,b_m),(b_{m+1},b_{m+2},\cdots,b_{2m}),\cdots$$

令 $\boldsymbol{s}_1=(b_1,b_2,\cdots,b_m),\boldsymbol{s}_2=(b_{m+1},b_{m+2},\cdots,b_{2m}),\cdots$，则 $\boldsymbol{F}_i=c_{i1},c_{i2},\cdots,c_{iN/m}$，其中 $i=1,\cdots,n$ 且

$$c_{ik}=\boldsymbol{a}_i\boldsymbol{s}_k=a_{i1}b_{(k-1)m+1}+\cdots+a_{im}b_{km}$$

$$|\boldsymbol{F}_i|=|\boldsymbol{F}|/m$$

步骤 5　利用选择的向量组对矩阵 $\boldsymbol{F}$ 进行重组. 如果文件 $\boldsymbol{F}$ 的 m 个分块 $\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_m$ 已经给出，那么可以依照以下步骤重建文件 $\boldsymbol{F}$. 令 $\boldsymbol{A}=(a_{ij})$，$1\leqslant i,j\leqslant m$ 是一个 $m\times m$ 的矩阵，其中第 i 行为 $\boldsymbol{a}_i$，可以看出 $\boldsymbol{A}[b_1,\cdots,b_m]^{\mathrm{T}}=[c_{11},\cdots,c_{m1}]^{\mathrm{T}}$，则 $[b_1,\cdots,b_m]^{\mathrm{T}}=\boldsymbol{A}^{-1}[c_{11},\cdots,c_{m1}]^{\mathrm{T}}$，其中 $b_j=a_{i1}c_{ik}+\cdots+a_{im}c_{mk}$，$1\leqslant j\leqslant N$.

通过信息扩散法，数据分片后在网络传输和数据存储时具有相对的保密性和安全性. 在云存储的基础管理层中利用 IDA 思想，通过分片器把存储信息分片，使数据变成无法被其他非认证系统所识别的数据片段. 对于每一个单独的数据片段来说，这些数据片段是不具有任何意义的，如果数据在网络传输过程中被他人截获，被植入的木马病毒扫描获取或在存储设备上被意外窃取，由于截取方只是获得信息的部分数据片段，截取的信息并不具有任何实际含义，这样就能够保证数据分片后不会产生保密

信息泄露或扩散. 此外,当这些分片后的数据放入地理位置不同的存储器中,即便被其他用户误操作提取时,也能保证需要保护的信息不会被分析出来. 数据的分散式存储机制也使存储系统具备一定的容错、容灾能力,提高了信息的可用性.

分片器[11]根据 IDA 算法将数据分片后,云存储服务器将每个分片数据用一个固定不变的 64 位句柄对其进行标识,这些句柄也是唯一的,读取数据时根据存储服务器中虚拟视图中句柄和字节范围来进行,在主服务器中建立一个数据列表,用来存储系统中的元数据,其中包括用户存储的文件名、对应的句柄号和文件大小等信息.

数据的存储无论在内容上还是在存储设备中都是分散的,当用户需要对云存储中的数据进行访问或者操作时,分散存储管理器(DSM)需要把分散的数据整合起来,提供给用户一个虚拟的视图,这些分散数据对于用户来说是透明的. 用户可以根据提供的视图,对存储的数据进行管理. 此外,DSM 还支持对元数据的管理,便于用户对数据进行创建、检索和删除等操作.

2.3 存储层设计

在存储层设计中,为了实现数据存储的安全策略,云存储中需要满足用户存储海量数据的需求,存储系统规模及存储容量都在不断增长,与存储相关的出错率将越来越高. 为了确保云存储安全系统中数据存储的高可用性和可靠性,系统存储层中的设备都必须异地存放,并且互为冗余,这样能够提高设备容错能力和存储利用率. 系统使用 Reed-Solomon 码[12]提供任意高错误恢复技术,保证系统在发现问题后能够被迅速检测到. 如果设备上的数据损坏、丢失,存储系统中自动化检测过程会发现这个问题,通过检查可用片重新计算数据片中所有的数据,根据其他存储设备中完好的数据恢复被破坏的数据. 通过这样的数据自举恢复,提高了云存储系统的平均无故障时间.

2.4 面向分散式存储的云存储数据存取过程

面向分散式存储的云存储用户进行数据存取的过程如图 3 所示,当用户(个人或企业)存取数据时,在客户端通过系统安全认证后将信息进行 SSL 加密,传送到 Internet 中,通过云存储服务器管理,用分片器将数据分片,然后再把数据片传给分散在各地物理位置不同的存储介质中去;当用户读取数据或者查询数据时,在云存储服务器通过身份验证后,用户通过分散存储管理器提供的虚拟视图,完成数据的检索或者删除等操作,此后分散存储管理器再通过设备上的记录表,对存储层中的存储设备进行相应的操作,最后将操作结果返回给用户.

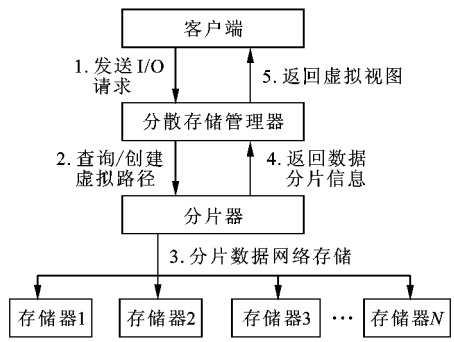

图 3 用户 I/O 存取过程

DSM 主要负责将数据转发给分片器,记录路径并为用户建立虚拟视图,具体工作过程如下:

(1)接收用户数据的 I/O 请求;

(2)转发给分片器并记录数据存取路径,等待分片器返还数据状态信息;

(3)创建用户虚拟视图表.

分片器根据 IDA 算法进行数据分片,分片规则为每个数据片中包含的信息内容不会被泄露,其具体工作过程如下:

(1)从分散存储管理器中提取数据;

(2)根据 IDA 算法进行数据分片;

(3)将分片后的信息传输到各地的存储设备中;

(4)将存储完成后的状态信息(成功或者错误)、存储设备号、存储位置等返回给分散存储管理器.

存储器主要实现数据最终存取,并带有定期检测数据、自动发现存储数据错误功能,并根据其他互为冗余的存储设备数据修复受损数据,以提高信息的可用性,其工作过程如下:

(1)存储数据后同时在其他存储设备中写入校验码;

(2)向分片器返回数据存储信息;

(3)定期自动检查数据的完整性.

3 系统分析与验证

用户保存数据时,系统首先在访问层实现用户身份认证、授权、防止非法访问和越权访问,并将保存数据进行加密,当数据在访问层和应用接口层被拦截时,拦截者不会得到有效的数据信息.

数据在基础管理层被分片后,每个信息中单独的片段也是无效的,在数据从基础管理层到存储层

传输的过程中，假设此时数据被拦截（木马病毒扫描获取），截获者得到的只是数据片段，无法对数据的有效性加以分析.

在数据最终存放的存储层中，分片的数据也因为其单独被获取后，获取者由于没有完整的分片数据信息，并且即使得到所有数据分片，也因其没有数据合成的方法，最终不能得到完整的数据. 此外，当数据被破坏时，存储层中互为冗余的数据也可将其恢复，保证了数据的安全性和可用性.

可以看出，本系统根据云存储的层次结构，通过相应的保护策略逐层对数据进行保护，从数据传输到存储，都建立了相应的保护措施进行云存储层以及层与层之间的防范，实现了数据的全面防护，防止了需要保密的数据泄漏后造成不必要的损失. 该方法在 Windows 环境下通过系统仿真加以了认证.

4　结　论

由于云计算的复杂性以及用户的动态性，云存储主要存在数据间分开存放、数据恢复、数据加密、数据完整性保护等问题. 本文提出的云存储安全架构对于数据保密性要求高、基于内容存储的用户，在安全性、可靠性和可用性方面具有较大的优势，具备一定的容灾、数据恢复及容错能力. 然而，这些需要通过空间和时间上的代价来满足安全性需求. 因此，根据本文提出的安全架构特点，通过增强动态分析明确最佳的性能所在，优化数据存取路径，提高数据整体读取性能，以实现更具有存储空间效率的存储策略，增强云存储服务效率等，将是我们下一步要研究的重点问题.

参考文献：

[1] 冯丹. 网络存储关键技术的研究及进展[J]. 移动通信，2009，33(11)：35-39.
FENG Dan. Network storage key technology of research and progress [J]. Mobile Communications，2009，33(11)：35-39.

[2] 中国计算机用户协会. 云计算中国峰会 2010[EB/OL]. (2010-03-25)[2010-08-10]. http：//www.cloudcomputingchina.org/cn/.

[3] BOWERS K D，JUELS A，OPREA A. HAIL：a high-availability and integrity layer for cloud storage[C]// Proceedings of the 16th ACM Conference on Computer and Communications Security. New York，USA：ACM，2009：489-501.

[4] CACHIN C，KEIDAR I，SHRAER A. Trusting the cloud [J]. ACM SIGACT News，2009，40(2)：455-461.

[5] WEATHERSPOOS H，EATON P，UBIATOWIZ J. Antiquity：exploiting a secure log for wide-area distributed storage [C]// Proceedings of the Second ACM European Conference on Computer Systems. New York，USA：ACM，2007：75-89.

[6] KOTLA R，ALVISI L，DAHLIN M. SafeStore：a durable and practical storage system [C]// Proceedings of the 12th USENIX Annual Technical Conference. Berkeley，CA，USA：USENIX Association，2007：129-142.

[7] 存储部落. 深度剖析云存储[EB/OL]. (2008-09-17)[2010-08-15]. http：// www. sansky. net/article/2008-09-17-depth-analysis-of-storage. html.

[8] STEVE M. Danger in the clouds [J]. Network Security，2008，2008(12)：9-11.

[9] RABIN M O. Efficient dispersal of information for security，load balancing，and fault tolerance [J]. Journal of the ACM，1989，1(38)：335-348.

[10] 屈志毅，苏文洲，赵玲. 一种基于信息分散算法的分布式数据存储方案[J]. 计算机应用，2006，26(5)：1102-1105.
QU Zhiyi，SU Wenzhou，ZHAO Ling. Distributed data storage method based on information decentralization algorithm [J]. Computer Application，2006，26(5)：1102-1105.

[11] HAMAN M，HIEROUS R M. An overview of program slicing [J]. Software Focus，2001，2(3)：85-92.

[12] PlANK J S，XU Lihao. Optimizing Cauchy Reed-Solomon codes for fault-tolerant storage applications [C]// Proceedings of the 5th IEEE International Symposium on Network Computing and Applications. Piscataway，NJ，USA：IEEE，2006：173-180.

（编辑　武红江）