基于云计算的安全数据存储研究

施　珺,李　慧,周立东

(淮海工学院计算机工程学院,江苏 连云港 222005)

[摘要]　基于互联网的超级计算模式(云计算)引起了人们极大的关注,同时也面临着越来越多的安全问题.针对云存储系统在应用过程中面临的数据安全性问题,本文提出了一种基于云计算的安全数据存储架构.该架构从数据存储和数据安全两方面研究了云计算的安全问题,分别提出了基于 Cache 的数据存储模式与基于第三方认证的数据安全模式,从而提高了数据的可用性,从数据存储到传输都建立了相应的保护措施,实现了云数据的有效防护.

[关键词]　云计算,数据安全,架构,研究

[中图分类号] TP313　[文献标志码] A　[文章编号] 1001-4616(2012)03-0138-05

Research of Security Data Storage Based on Cloud Computing

Shi Jun,Li Hui,Zhou Lidong

(Department of Computer Science, Huaihai Institute of Technology, Lianyungang 222005, China)

Abstract: As a new internet-based super computing model, the Cloud Computing technology has aroused great concern, and facing an increasing number of security threats at the same time. A secure date storage architecture based on Cloud Computing is presented in this paper to deal with the data security in the cloud storage systems and applications. This architecture researches the security issues from data storage and data security including the data storage model based on caching and data security model based on the third party secure publication. This method improves the high availability of data. Protection measures from the data storage to transmission are taken to ensure safety.

Key words: Cloud Computing, data security, architecture, research

随着互联网技术的快速发展,云计算(Cloud Computing)技术作为互联网技术中新兴的研究和应用领域,越来越受到人们的关注,并在近两年得到了迅速的推广和流行.如何高效、安全地保存和传输生成于云端的大量商业数据,也成为业界研究的重点.

越来越多的应用希望能够安全地存储、管理、共享和分析大量的复杂数据,来确定其模式和趋势,尤其是对既希望降低设备投入与管理成本、又需要容量可伸缩扩展性很高的在线服务提供商来说,云存储是很好的解决方案.云存储的资源分配和调度策略的特性决定了其安全挑战是数据拥有者不能控制数据被存放在哪里.因此,要确保商业应用云的安全,就需要维护在这种不受信任处理过程中的数据安全.

将数据迁移至云中,致使企业用户在数据安全性和可用性方面高度受制于其云存储服务器供应商.让众多企业将其数据迁移至云中非常艰难,可以说安全性和可用性的担忧是企业走向云存储模式的至高无上的因素.一旦上述问题得以解决,云存储也就适应了商业化信息存储库的需要.初始的备份可以在装置内完成,也可以在云存储上另作备份获得装置外的数据保护.考虑到只有新增的数据才会被迁移至云中,因此支持自动化增量备份的技术最为适宜.自动化增量备份将提供一个高效战略,即增量备份在降低宽带压力的同时、自动化特性也节省了雇员进行日常相关操作的时间.

为了实现在复杂的网络环境中保障云数据的安全存储及存储服务中的隐私,提高云存储中用户数据

收稿日期: 2012-07-27.

基金项目: 江苏省自然科学基金资助项目(11KJB520001)、连云港市科技攻关项目(CG1122).

通讯联系人: 施　珺,硕士,副教授,高级工程师,研究方向: 教育信息化,云计算. E-mail: shijlfg@126.com

的安全性、可信性,本文提出了一个云环境中的安全数据存储架构,提出了基于 Cache 的数据存储和基于第三方认证的数据安全模式,从数据存储和数据安全两方面提高云数据的可靠性与可信性.

1　云存储基础

1.1　基本概念

云计算是分布式计算技术的一种,是分布式处理、并行处理和网格计算的发展. 其最基本的概念是通过网络将庞大的计算处理程序自动拆分成无数个较小的子程序,再交给由多部服务器组成的运算系统,经过计算分析之后将处理结果回传给用户.

云计算是对分布式处理、并行处理、网格计算及分布式数据库的改进处理,其前身是利用并行计算,解决大型文体的网格计算和将计算资源作为可计量的服务提供的公用计算,是在互联网宽带技术和虚拟化技术高速发展后萌生出来的[1,2]. 云计算可为我们提供众多的服务,如 SaaS(软件即服务)、PaaS(平台即服务)以及 MSP(管理服务提供商)等,这些服务都可使用户更加专注于自己的创新,而不必担心一些繁琐的细节,极大地降低了成本.

1.2　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[3],如图 1 所示:

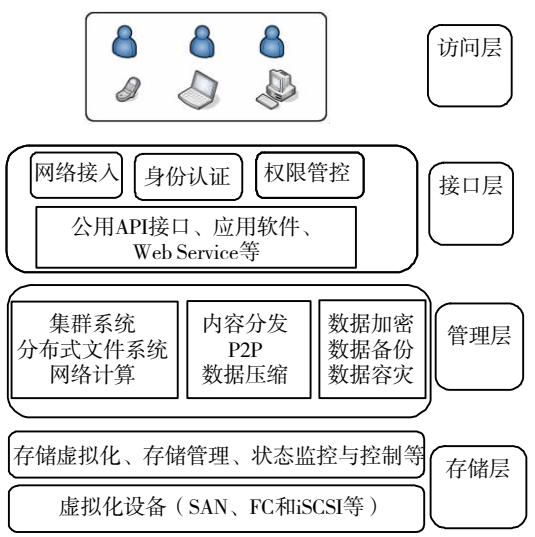

图 1　云存储系统结构

Fig.1　The system architecture of cloud storage

(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.

(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.

2　基于 Cache 的数据存储模式

2.1　云的架构设计

在基于云计算的数据存储服务中,云端应作为服务端存在,云对外仅提供一套接口以供客户端调用,云的内部实现依赖动态启动并分配的计算机节点. 将云设计为可伸缩规模的动念集群,可以有效利用计算资源和虚拟机技术的优势. 在云中,对各节点的控制需要一套完善的管理机制,尤其是云的伸缩以及节点启动、关闭、资源分配、数据转储、状态监控等.

云中节点的设计需要采用某一种拓扑结构,对于数据存储服务来说,星形结构的设计有利于充分发挥云的可伸缩性特点. 由于云中的节点是动态启动、分配和回收的,因而云设计是以一个中心控制服务器为中心的星形结构. 如图 2 所示,中心控制服务器控制和管理周围各运行节点的运行,而各节点均与数据池

保持通信.

2.2　Cache 设计

在数据存储服务中,数据先是存储于云中的节点上,但因为节点数量和节点本身资源是有限的,因而数据必须由节点转储到数据池以供长期存储. Cache 普遍存在的一种典型的应用为由某客户端数据上传至云中,此时数据仍留在节点上,还未转移至数据池,若此时服务端接收到数据获取请求,则可以不经数据池而直接将其由云中节点取出并处理. Cache 的设计原则是: 每一个云中的运行节点均可视作 Cache,在节点存储能力未达到上限或节点未收到明确的需要将其上数据转储至数据池的命令前,服务端总是将其视作 Cache 来处理. 当某客户端请求下载一个数据时,服务端将先检查 Cache 表,从而确定被请求下载的数据是否已经存在于某节点中. 如果找到,则视为 Cache 命中,指派使用该节点来同客户端交互,完成数据下载; 如果未找到,则另行分配节点. 基本的 Cache 管理机制如图 3 所示.

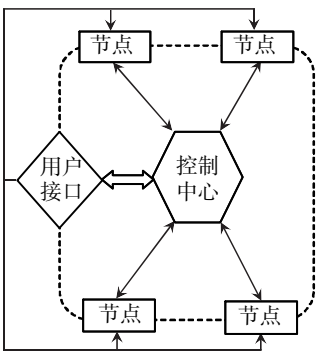

图 2　基于星形模型的拓扑结构

Fig.2　The topological structure based on star mode

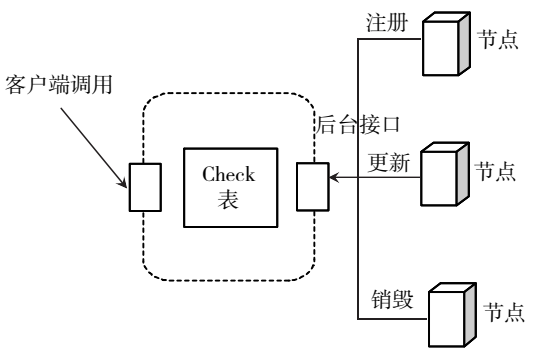

图 3　基本的 Cache 管理机制

Fig.3　The basic Cache management mechanism

控制中心负责处理用户接口的传输请求,经处理后,通过一定的机制、与云中各节点通信. 控制中心本身可以作为一个关键节点,除了与各节点间的通信外,中心节点还负责同用户接 $I=1$ 间的信息接收与反馈. 在起始通信完成后,即已分配和确定相应的节点后,控制中心把与客户端的通信通路交给节点,由节点直接完成和客户端的交互.

3　基于第三方的数据安全模式

3.1　应用于云计算的第三方安全数据发布

云计算方便远距离的数据存储,使资源利用达到最大化. 因此,对数据进行保护并只允许授权用户访问等保护措施就显得尤为重要. 这在本质上也就等同于保护第三方数据发布,同时对于数据外包和公开文档来说都十分必要. 目前针对第三方数据发布的安全问题已经存在许多技术. 我们假定这种数据被表示为 XML 文档的形式,由于目前网页上很多的文档都被表示为 XML 文档,因此这是一种合理的假设. 首先,我们提出一个访问控制框架(图 4),然后讨论基于第三方的安全数据发布. 在图 4 所示的访问控制框架中,安全策略依赖于用户角色和证书. 用户必须要拥有访问 XML 文档的证书,这个证书依赖于他们的角色. 例如,一个教授有权访问学生的所有信息,但是秘书只能访问管理信息. XML 规范被用于制定安全策略,为整个 XML 文档或文档的一部分授予访问权限. 在某种情况下,访问控制可能在 XML 树下传播.

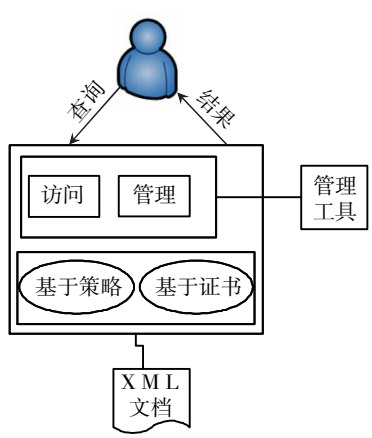

图 4　访问框架结构

Fig.4　The structure of access frame work

例如,如果根被授予访问权限,那么这并不一定意味着所有的结点都被授予访问权限. 如果一个用户被授予了文档类型定义(DTD)的权限,但他并不具有创建文档实例的权限. 一个人也可能被授予部分文档的访问权限. 例如,一个教授没有权力访问学生的医疗信息,但是他有权访问学生的成绩和学术信息.

关于 XML 文档的安全发布架构,我们的想法是引入不可信的第三方发布服务器. 文档拥有者为访问

对象指定访问控制策略. 当访问对象请求文档时,访问对象会从文档所有者那里获得策略,文档所有者把文件发送给发布服务器. 当访问对象请求一个文档时,发布服务器将会把相关的策略应用于访问对象,并把部分文档发送给访问对象. 现在,由于发布服务器是不可信的,它可能会给访问对象发送错误的信息. 因此,文档所有者将把文件、策略的各种组合和她/他的私有密钥加密. 用梅克来签名和加密,访问对象可以核实文档的真实性和完整性(图 5) .

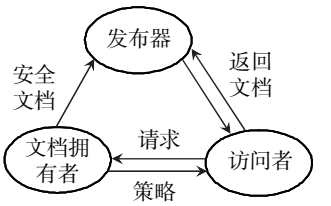

图 5　安全的第三方发布

Fig.5　Safety issued of third party

在云环境中,第三方发布服务器是在云中存储敏感和重要数据的服务器. 这种数据必须被保护,而且我们已经讨论的安全架构可以应用于任何需要保持数据真实性和完整性的场合.

3.2　云计算加密数据的存储

由于云中的数据会被放置在任何地方,因此数据被加密就显得很重要. 我们使用安全的协同处理部件,使加密的敏感数据得到更加高效的存储. 有些人可能会产生这样的疑问: 为提高数据安全,为什么不在目前云计算系统提供的硬件平台上(如 Open Cirrus) 运行自己的软件呢? Open Cirrus 是唯一一台可以支持横跨系统、应用、服务、开放源代码开发和数据中心研究的云计算试验台. 其实我们已经考察了这种选择. 首先,Open Cirrus 由于自身的经济模式只提供有限的权限. 其次,Open Cirrus 没有提供给我们需要的硬件支持(如安全的协处理器) . 只有通过插入一个安全的协处理器(SCP) 到云基础架构,系统才可以高效地处理加密数据(见图 6) .

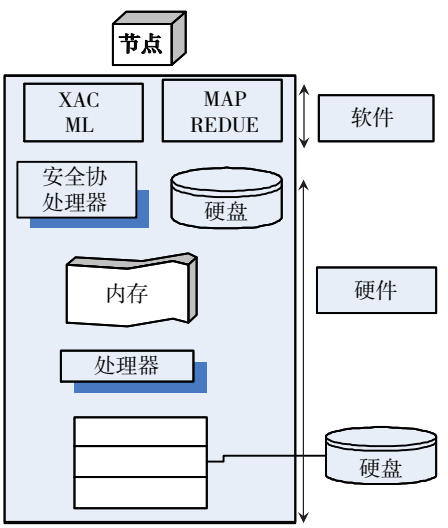

图 6　安全架构示意图

Fig.6　Security architecture schemes

简单来说,SCP 是一个防篡改的硬件、能够进行有限的通用计算. 例如,IBM4785 的加密协处理器是一个由包含 CPU、内存、特殊用途的加密硬件组成的单板机,并且通过了联邦信息处理标准(FIPS PUB 140 - 1) 的认证. 当在服务器上安装 SCP 后,它具有从服务器上完全隐藏的本地计算性能. 如果篡改行为被检测到,那么 SCP 就会清除内存. 由于 SCP 是难以破坏的,人们可以在其上运行整个敏感数据存储服务器,而将整个数据存储到 SCP 是不可行的,原因如下: 首先,由于 SCP 通常限制内存(只有 RAM 的几个兆字节以及几个不稳定的内存字节) 和计算能力. 虽然随着时间的推移性能会提高,但是一些固有问题,如热消散、用电(这必须被控制起来,来避免披露处理) ,将会拉开一般用途和安全计算之间的差距. 另一个问题是,SCP 上的软件运行必须是完全地被信任和被证实,这个安全性要求意味着 SCP 上运行的软件应该尽可能被简单地保存. 那么硬件如何帮助存储大量的敏感数据集呢? 我们可以用随机的私有密钥(Private keys) 给敏感数据集加密,来减轻密钥泄露的风险,我们也可以通过难以被破坏的硬件去存储一些加密密钥/解密密钥(如万能钥匙可以为所有其他的密钥加密) . 因为密钥在任何时候都不会在未加密状态而驻留在内存中,并且攻击者不能通过取得系统快照知道密钥. 而且,攻击者采取的任何通过软件或者硬件控制 SCP 的尝试行为都将会被清除. 因此,这就消除了一种解密任何敏感信息的途径. 这种框架将会促进安全数据存储和确保信息共享. 例如,SCP 可以被用作隐私维护信息集成,这对于信息共享是很重要的.

4　小结

云计算是网络时代发展的又一个高潮,随之伴生的云存储在其中起着推波助澜的重要作用. 如何构筑高效云存储、如何控制云存储系统成本、如何利用有效云存储系统,这些问题随着云存储技术的广泛应用将会逐渐得到解决,云存储问题的有效解决也将为云计算应用的普及打好坚实基础.

[参考文献]

[1] Feng D. Network storage key technology of research and progress [J]. Mobile Communications, 2009,33(11):35-39.

[2] Chuan Y, Xu J P. Recommendation algorithm combining the user-based classified regression and the item-based filtering [C] // Processing of the International Conference on Electronic Commerce. Melbourne, 2006: 574-578.

[3] 边根庆,高松,邵必林. 面向分散式存储的云存储安全架构 [J],西安交通大学学报:自然科学版,2011,45(4):41-45.

[4] 蔡文,杨春燕,林伟初. 可拓工程方法 [M]. 北京:科学出版社,2000: 130-136.

[5] 张光卫,康建初,李鹤松,等. 基于云模型的全局最优化算法 [J]. 北京航空航天大学学报:自然科学版,2007,33(4): 486-491.

[6] 张光卫,李德毅,李鹏,等. 基于云模型的协同过滤推荐算法 [J]. 软件学报,2007,18(10): 2 403-2 411.

[7] 代劲,何中市,胡峰. 基于云模型的连续属性决策表简化算法 [J]. 南京大学学报:自然科学版,2009,45(5):638-644.

[责任编辑:黄 敏]

(上接第137页)

[6] Gao D, Han S, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(6): 989-1 005.

[7] Zhang L, Tong M H, Marks T K, et al. SUN: A bayesian framework for saliency using natural statistics [J]. Journal of Vison, 2008,8(7): 1-20.

[8] Torralba A, Oliva A, Castelhano M S, et al. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search [J]. Psychological Review, 2006, 113(4): 766-86.

[9] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look [J]. ICCV, 2009: 8.

[10] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2 376-2 383.

[责任编辑:黄 敏]