基于结构化策略的隔离分布式云存储服务架构

高英 郑建林

华南理工大学计算机科学与工程学院，中国

gaoying@scut.edu.cn，helinz@qq.com

**摘要**：云安全威胁极大地阻碍了云计算的发展和云应用的推广，如何保护存储在云中的数据不仅是安全的核心问题，也是云计算和云应用发展面临的最大的挑战。在上述问题的基础上，创造性地提出了基于云存储服务概念的信任控制体系结构。基于结构化策略和独立分布式存储架构，云服务架构可以消除对云安全的担忧，实现高度安全、可靠和可用的云存储。实验证明，这是一种简单而有效的方法。

**关键词**：信任机制，安全性，结构，存储，云，服务。

1介绍

推动云计算之所以是非常困难的，因为对于客户来说，安全问题被认为是云计算的关键问题。 人们相信银行并且把他们的钱投入银行，因为银行是国有机构并且背后有政府的法律保障，但云提供商的数据中心的安全性没有得到任何第三方公信机构承认，在制度保证方面的可信度[1]不能确定。 因此用户害怕在云中心共享他们的数据。 云安全已经成为云计算发展中遇到的最大问题; 因此，如何提供一种可信的机制来满足云环境中的安全需求，是当前亟待解决的问题。

2结构化隔离分布式云存储服务的思想

**2.1解决云安全的方法**

相对于技术而言，云计算本质上更多是基于现有技术的创新模式或者是策略。 云计算改变了服务方式，但并没有推翻传统的安全模式[2]。

云计算的不同之处在于安全措施的变化，安全设备的部署地点以及安全责任的主体。 与传统方式类似，解决云计算安全问题的方式也是策略，技术和人员这三个要素的结合[3]。

Z. Du（Ed）：2012年国际MCSA会议论文集，AISC 191，第31-37页。

springerlink.com ©Springer-Verlag柏林海德堡2013

**2.2云存储服务的概念**

云存储是基础架构层中云计算概念的延伸，然而，它目前主要用作云计算平台的附属，而不是作为独立的标准化云服务。另外，云提供商已经建立了自己的数据中心;然而数据只存在于各自的中心。

云计算意味着信息技术基础架构可以作为服务运行，服务可以是从原始硬件租赁到使用第三方API的任何事情。同样，存储设施也应该被用作云环境的基本资源。云存储的未来应该是云存储服务。

随着下一代云存储技术的发展，存储（服务）是对现有存储方式的改变，它是一种特殊形式的架构服务。这种云存储服务对用户也是完全透明的。它并不是指特定设备或云提供商的云存储中心，而是指由分布在不同物理地理位置或各种提供商的存储中心的存储设备所组成的聚合体，用户可以使用多种不同的云服务商提供的服务 - 存储中心。

云存储服务的核心是云服务协议与云存储系统的结合，并采用开放的服务接口标准，从而实现云存储向云存储服务的转型。这是本文后面提出的架构的核心思想。

**2.3基于结构化策略的云存储服务孤立分布式体系结构解决方案**

基于结构化策略的孤立分布式云存储服务有3个方面的特点：结构化，分散化和孤立分布式存储。 实质上，它基于数据方案将数据分解为几个单独无意义的数据子集，然后使用属于不同云提供商的隔离的分布式云存储服务来存储这些子集。 对于云计算来说，它提供了一个安全可用的存储解决方案，并且是可以信任的（如图1所示）。

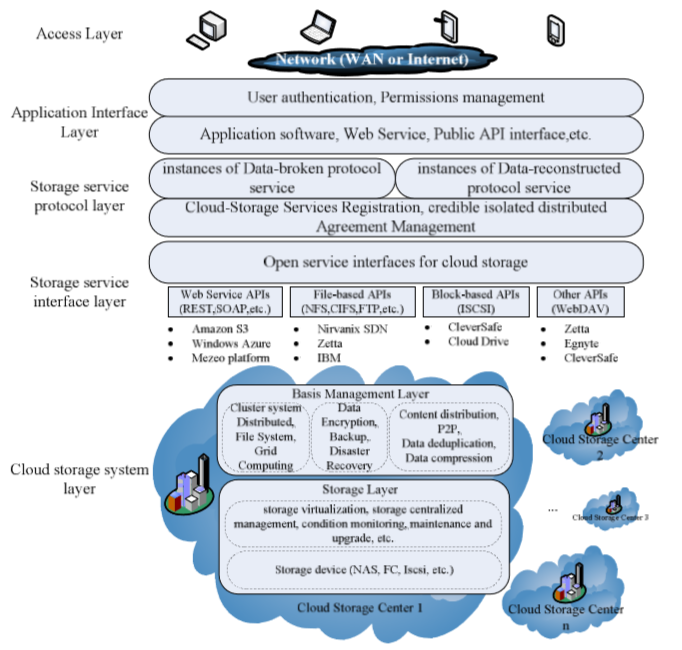


图1.基于结构化策略的孤立分布式服务体系结构

该体系结构是一种信任机制。数据存储在来自不同云提供商的孤立分布式云存储中心中，这样用户可以重新获得最初属于云提供商的数据的控制权，然后对分布式控制台进行更高级别的集中控制。再加上对数据保护的耐受性入侵(基于数据方案的数据破坏)，这样单独的无意义数据存储让用户不必再担心云存储的安全性。此外，独立分布式存储的方法将整体风险分散开来，每个云存储服务提供商自身安全性的叠加使系统具有更高的安全性，简单的结构化分布使其非常高效。在架构中，根据云存储服务协议（数据中断协议，数据重构协议和孤立分布式协议）的云计算应用可以在其自己的云计算环境中创建服务实例，以调用云存储服务的接口来存储数据，也可以直接使用实现上述结构化孤立分布式接口的第三方服务。

**2.4结构化隔离分布式存储设计**

结构化，孤立的分布式存储服务层涉及三个主要服务实例：数据中断，数据重建和隔离分布式存储服务。

数据中断的服务实例的输出是基于数据的逻辑结构的数据段。云应用程序提交的数据分为两类：强结构化数据和非逻辑数据。对于结构化数据，首先根据数据的结构进行分段，其次，继续分割每个数据段，然后形成多个具有相同方案的子集；对于非逻辑化的数据，它将首先根据默认的规则格式进行分段。

数据子集形成后，孤立的分布式存储服务将收到通知。通过访问不同云组织提供的用户认证和可用云存储服务，形成多子集到云存储多业务的映射，然后将映射和所有子集存储在不同的孤立分布式云存储中心。最后，存储的映射路径存储在云服务注册表中，并将反馈结果发送给用户（数据存储和修改操作都可以得到反馈结果）。 （图2）

对于数据重构服务的实例，当数据读取请求到达时，从云服务注册中心获取已存储的映射路径，然后根据结构化映射从孤立的分布式子集中请求需要的数据块，并基于结构恢复数据的原始内容。如果请求是数据更新或数据删除时，则实例将根据原始映射进行相应的操作，然后在删除后更新映射（当受影响的碎片计数更多时，它也主要存在于数据更新过程中） 或者更新受影响的数据段。

孤立的分布式存储服务基于底层的开放云存储服务接口运行。通过从属键值对，列，文档，图和其他数据的逻辑模型，该接口层包括：REST，Thrift和Map Reduce，GET / PUT，语言特定的API和SQL子集。

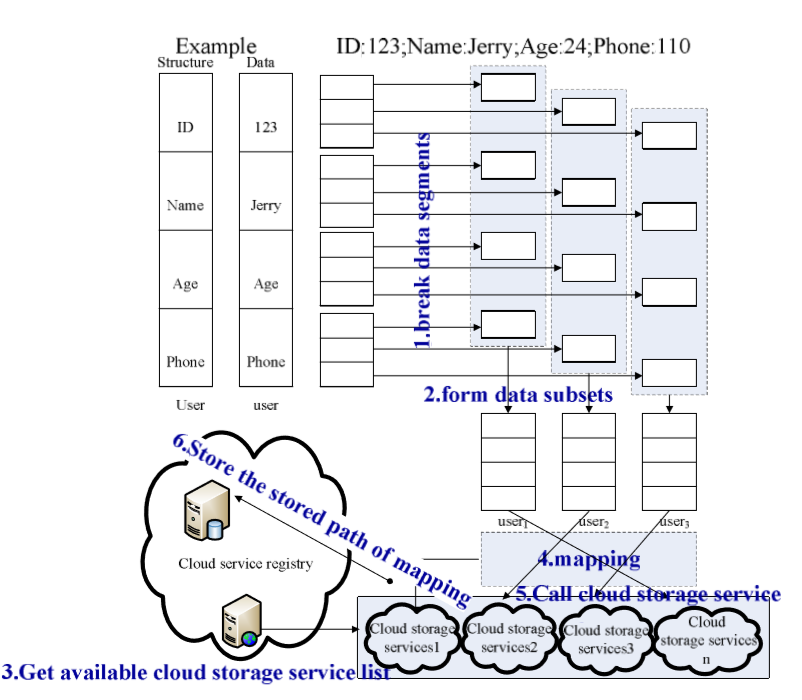


图2.结构化隔离分布式存储的基本过程

**2.5数据分离的方法**

早在云计算的到来之前，数据分离和重建在分布式系统已经取得了很大的进步。目前,最常用和有效的方法是基于信息传播算法（IDA或RS），但在本文提出的云存储服务体系结构中，数据的分离和重建是基于信息与自身设计的模式。这种简单高效的数据分散算法比IDA和RS具有更低的存储、计算和带宽成本。

**2.5.1数据分离的方法**

基于冗余编码技术的信息分散传播算法的基本思想是：文件F的长度L被分割成n个片段Fi（1<= i <=n），使得文件可以被任意m片段重建。在空间中，我们可以看到n片Fi的总和是原始文件的n / m倍，n / m大于或接近1。

IDA同样将需要分配的数据量D等分为多个片段（大小为m）。假设卷的大小N为|D|，则片段数N'为N / m。使用与阶段无关的n \* m向量矩阵乘法将数据编码为n段数据，并且其大小为N'。 RS和IDA的唯一区别在于，RS使用与阶段无关的(n-m)\* m范德矩阵来计算每条数据的校验和数值，然后根据数据左对齐的方式构成具有原始数据的n个片段类似于RAID 6的奇偶校验。在相同的n，m和故障片段数据编号的情况下，IDA和RS的时间成本基本相同，但随着n和m的增加，RS方法的工作效率将略高于IDA [4]。

**2.5.2基于模式的信息扩散算法**

基于结构的数据分散算法根植于信息的二重性：形式和内容。没有模式的数据不存在，只要数据存在于系统中，无论是存储还是应用，它都有自己的模式。数据需要解决并显示其合​​格意义，这是数据的内容；如何构建或解析数据是模式反映的数据的形式。因此，一个完整且有效的信息对象是形式与内容的统一体。

在应用过程中，数据都有自己的组织形式（这里称为模式），基于数据结构的分散算法实际上是使用数据附带的模式分散的，例如分散到一个长度L的信息I，使用数据字段，表或文件作为基本粒度，将数据的形式和内容剥离出来并将数据分散到n个数据子集Si（1 <= i <=n）具有相同的模式，Li对应于Si（1 <= i <= n）的长度，其中。

在空间上，基于模式的算法中数据子集的总长度等于原始文件的数量，所需存储空间小于信息分散算法（IDA，RS）；随着时间的推移，因为它基于数据模式，通过结构化阶段和数据解析阶段的集成进行简单的分段和重组，所以它非常高效。它不需要向量矩阵乘法编码，所以运算效率比信息分散算法要好。另外，为了避免不必要的重复，各个云提供商提供的底层云存储服务都采用了灾难恢复和纠错流程，因此可以极大地减少时间和空间的消耗。

3系统分析和验证

该体系结构不同于传统的保证信息安全的加密方法。它采用信息组织结构中的密码机制，使信息内容和形式分离，并结合孤立的分布式云存储服务解决方案建立信息信任系统，确保信息的可靠性，可用性和隐私性。

应用程序数据在结构化破坏之后的程序中成为单独的无意义子集，并且所有子集都被隔离地分布式地存储在由不同云提供商提供的云存储介质中，并且具有拉出的模式和映射文件。分离的一个或多个数据子体是无效的，只有使用所有必需子集的模式映射才能重构原始消息体。这种内容和形式的孤立共享，相互制衡的云提供商确保内部和外部信息的安全，用户无需担心外部黑客或内部员工或云用户对云数据提供商的入侵，盗窃和滥用等一系列威胁。为了突破体系结构的防线，至少需要通过以下检查点：用户安全认证、模式映射的安全密钥以及各云提供商的云存储服务安全认证。因此，这一防线非常扎实，可靠的云提供商也保护了系统安全的权威。该解决方案是通过LINUX环境中的系统仿真和验证，并进行信息分散算法比较。

在基于模式的扩散算法与信息扩散算法的对比中，实验的输入是：具有用户结构的数据文件（图2），长度为1M，n = [5,15]。 基于模式的分散算法的零散子单元为文件，其基本粒度为表; 信息分散算法中m = n-1。实验重复10次，得到离散和重构的时间代价（图3）。数据表明，基于图式的分散算法的效率优于信息分散算法，其效率与原始数据文件的大小之间存在线性关系。

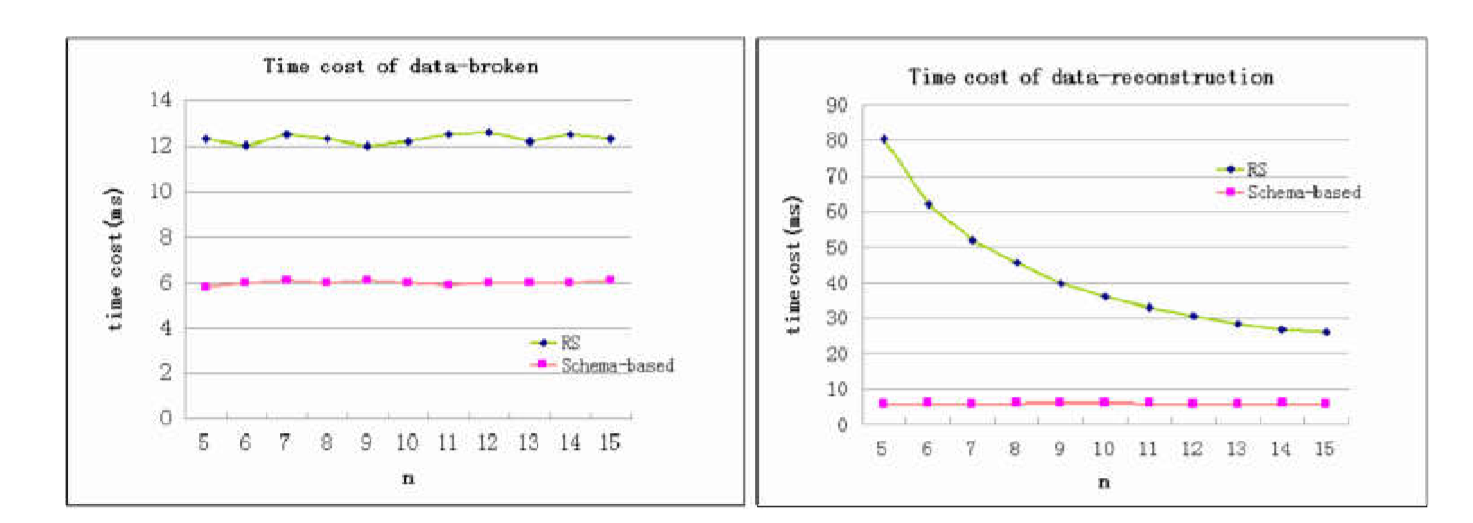


图3.数据中断和数据重构的时间成本

4结论

与分布式存储相比，本文创造性地提出了一种基于云存储服务概念的信任控制体系结构。 用户不再需要担心云存储的安全问题；它消除了云推广面临的最大障碍，对云的发展具有重要意义。

**感谢**：该研究得到了以下支持：

* 广东省科技计划项目（2009B010800047）。
* 广东省教育部大学 - 产业研究项目省（2009B080702037）。
* 广东省教育部大学 - 产业研究项目省（2010B090400535）。
* 中央大学基础研究基金（20112M0068）。

参考

[1] Yi，Y.，Chun，L.Y .:安全云计算环境问题。电脑知识与技术（5）（2009）

[2]从云计算到云安全。 信息系统工程（2009年）

[3] Yi，C.S .:关于云计算安全问题。 网络安全技术与应用（10），22-23（2009）

[4] Rabin，M.O .:安全信息的有效散布，负载平衡和容错。计算机械协会杂志（36），2-4（1989）