计算机研究与发展
Journal of Computer Research and Development
ISSN 1000-1239/CN 11-1777/TP
50(1): 136-145, 2013

安全云存储系统与关键技术综述

傅颖勋[1] 罗圣美[2] 舒继武[1]
[1](清华大学计算机科学与技术系 北京 100084)
[2](中兴通讯股份有限公司 南京 210012)
(mooncape1986@126.com)

Survey of Secure Cloud Storage System and Key Technologies

Fu Yingxun[1], Luo Shengmei[2], and Shu Jiwu[1]
[1](*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084)
[2](*Zhongxing Telecom Equipment Corporation, Nanjing* 210012)

Abstract With the rapid development of cloud storage, more and more people prefer to store their owner data in remote cloud storage to avoid troublesome data management in local storage systems. The most famous feature of cloud storage is the concept that storage as a service, users can store their own data into clouds by public APIs. However, because of losing absolute control of data, users storing their own data in cloud storage will suffer a series of security problems, such as data peeping, data tampering, and so on. In order to solute those security problems and improve the quality of secure cloud system based on enhance its security, researchers have investigated lots about cloud security problem in recent years, which established a research branch of the cloud storage- -secure cloud storage system. This paper introduces the security demand of secure cloud storage system; expounds the current status of cloud storage system; summarizes the key technologies of currently secure cloud storage systems, such as encryption key's distribution and management, attribute-based encryption, searchable encryption, ciphertext-based data deduplication, provable data possession and proof of retrievability mechanism, data assured delete, etc. At the end of paper we discuss the future research directions of secure cloud storage system.

Key words cloud storage; secure cloud storage system; data security; key management; ciphertext searching; assured delete

摘 要 随着云存储的迅猛发展,越来越多的用户选择使用云存储存放自己的资料.云存储的最大特点在于存储即服务,用户可以通过公有API将自己的数据上传到云端保存.但由于用户丧失了对数据的绝对控制权,一些数据安全的隐患也由此产生.为了消除安全隐患,并在保证安全性的同时尽可能地提高系统的服务质量,近年来国内外机构作了大量研究,从而开启了云存储中的一个研究方向——安全云存储系统.首先介绍了云存储系统的安全需求,然后阐述了安全云存储系统的研究现状,并总结了现有安全云存储系统中的一些关键技术的现状与不足之处,其中包括密钥分发与管理、基于属性的加密机制、基于数据密文的搜索机制与删冗机制、数据的持有性证明与恢复以及数据的可信删除等;最后指出了安全云存储系统未来的研究方向.

收稿日期:2012-08-20;**修回日期**:2012-09-12;**网络发表日期**:2012-12-05
基金项目:国家杰出青年科学基金项目(60925006);国家"八六三"高技术研究发展计划基金项目(2012AA011003);中兴通讯股份有限公司合作项目(One1112300007)

关键词 云存储;安全云存储系统;数据安全;密钥管理;密文搜索;可信删除

中图法分类号 TP309

随着计算机技术和互联网应用的迅速发展,数据正以几何级数的方式增长,人们对存储空间的需求也越来越大.在这一趋势下,近年来云存储的提出与发展以及存储即服务的理念为人们提供了大量廉价的存储空间,同时也向传统的数据存储方式发起了挑战.

尽管云存储有着价格低廉、部署方便等优点,其推广过程却十分缓慢.从 Twinstrata 公司 2012 年最新的云存储的应用调查来看:只有 20%的人愿意将自己的私有数据放在云存储中;相比之下,大约有 50%的人愿意使用云存储来进行数据备份、归档存储以及灾难恢复等作业[1].由此可见,数据的安全问题是云存储推广的重大障碍之一,云存储系统对安全机制有着十分迫切的需求.

数据的安全性包含 CIA(confidentiality, integrity and availability)3 个方面,即机密性、完整性和可用性.机密性是指任何人或团体在非授权的情况下不得查看到数据明文;完整性是指数据在存储过程和传输过程中未被篡改,或者能够检测出此数据已被篡改;可用性是指用户可以通过云存储接口随时使用自己的数据.为了解决数据的安全问题,国内外对此作了大量的研究,分别从机密性、完整性和可用性 3 方面提出了一些新的系统架构来保证数据的安全性.

目前保证数据机密性的主流方法是将数据进行加密.数据被加密后,用户只需要保护好自己的密钥就可以保证数据在存储过程和传输过程中的机密性.但由于云端保存的所有数据都是加密的,云存储在无法偷窥用户数据内容的同时,也无法以传统的方式提供一些常见的功能,例如数据搜索、数据删冗等.在安全云存储系统中如何提供这些功能也非常值得研究.

本文从云存储系统中的安全问题与需求出发,详细介绍了目前已有的安全云存储系统的现状与关键技术,并指出未来安全云存储系统的研究方向.

1 云存储系统的安全需求

云存储[2]是在云计算[3]概念的基础上发展起来的一种新的存储方式,它是指通过网格计算、集群文件系统、分级存储等现有技术,将网络中大量的存储设备通过硬件/软件的方式集合在一起,并对外提供标准的存储接口,以供个人或企业调用并存储数据的存储方式.相比传统的存储方式,云存储的出现使得一些企业或个人不需要购买价格高昂的存储设备,只需要支付较少的费用便可以享受无限的存储空间.

随着云存储理念的深入发展,越来越多的企业开始搭建属于自己的云存储平台,并通过一些特定的接口为企业或个人提供存储服务,例如 Amazon 的 S3[4],Microsoft 的 Azure[5]等.云存储平台的出现使得许多企业或研究机构利用它来开发自己的系统,这些系统也被称之为云存储系统.近年来,云存储系统泄漏用户数据事件的不断涌出,使得如何保证云存储系统的安全性已成为一个不可忽视的问题[6].

与传统的存储方式相比,云存储中的安全需求不仅是保证数据的安全性,而且还包含了密钥分发以及如何在数据密文上进行高效操作等功能需求.

1.1 数据的安全性

数据安全是云存储系统中最重要的安全需求之一.云存储系统中数据的安全性可分为存储安全性和传输安全性两部分,每部分又包含机密性、完整性和可用性 3 个方面:

1) 数据的机密性

云存储系统中的数据机密性是指无论存储还是传输过程中,只有数据拥有者和授权用户能够访问数据明文,其他任何用户或云存储服务提供商都无法得到数据明文,从理论上杜绝一切泄漏数据的可能性.

2) 数据的完整性

云存储系统中数据的完整性包含数据存储时和使用时的完整性两部分.数据存储时的完整性是指云存储服务提供商是按照用户的要求将数据完整地保存在云端,不能有丝毫的遗失或损坏.数据使用时的完整性是指当用户使用某个数据时,此数据没有被任何人伪造或篡改.

3) 数据的可用性

云存储的不可控制性滋生了云存储系统的可用性研究.与以往不同的是云存储中所有硬件均非用户所能控制.因此,如何在存储介质不可控的情况下提高数据的可用性是云存储系统的安全需求之一.

1.2 密钥管理分发机制

一直以来,数据加密存储都是保证数据机密性的主流方法[7-8].数据加密需要密钥,云存储系统需要提供安全高效的密钥管理分发机制保证数据在存储与共享过程中的机密性.

1.3 其他功能需求

由于相同密文在不同密钥或加密机制下生成的密文并不相同,数据加密存储将会影响到云存储系统中的一些其他功能,例如数据搜索、重复数据删除等,云存储系统对这些因数据加密而被影响的功能有着新的需求.

2 安全云存储系统概述

用户对云存储的不信任引发了云存储系统中的安全问题.近年来,随着云存储的推广与普及,虽然有越来越多的人开始使用云存储存放自己的资料,但云存储系统中的安全问题却并没有得到缓解.为了解决云存储系统中的安全问题,国内外的研究者作了大量研究,逐渐在云存储系统的研究中形成一个新的方向——安全云存储系统.

2.1 安全云存储系统设计的一般原则

安全云存储系统是云存储系统的一个子集,它指的是包含了安全特性的云存储系统[9].安全云存储系统的设计者常常会提出一些安全方面的假设,然后根据这些假设建立系统的威胁模型与信任体系,最终设计并实现系统或原型系统.一般来说,安全云存储系统设计时需要考虑如下几个方面[8].

1) 安全假设.在安全领域中,最好的假设是除自己以外的所有实体都不可信.但是在云存储系统中,数据被存放在云端,拥有者对数据丧失了绝对控制权,使得这一假设只存在理论上的可行性.因此,云存储安全系统的设计者需要针对不同的应用场景提出相应的安全假设,并以此为前提来保证系统的安全性.

2) 威胁模型和信任体系.设计者基于安全假设相关实体进行分析,由此得出相关实体是否可信,然后将这些实体模型化或体系化,由此得出相应的威胁模型和信任体系.

3) 保证系统安全的关键技术.设计者往往会根据自己系统的应用场景与特征,采取一些相关技术来保证系统的安全性,这些技术也称为安全云存储系统的关键技术.

4) 系统性能评测.系统的安全与高效是一对矛盾体,在保证系统安全性的同时必然会在一定程度上降低系统效率.在安全云存储系统中,设计者需要对系统的安全与效率进行均衡,使得系统能够在适应所需的安全需求的同时,为用户提供可接受的性能.

2.2 安全云存储系统的现状

从存储系统的技术支撑与发展来看,文件系统是构建云存储系统的重要部分.CFS[10-11]是最早的加密文件系统之一,它是一个用户态的虚拟加密文件系统,可以挂在在其他文件系统之上,为使用者提供文件/文件名加密保护的功能.此后,NCryptfs[12],ECFS[13],Cepheus[14],TCFS[15]等都是在 CFS 的基础上研究开发的.NCryptfs 是一个内核态的加密文件系统,它将 CFS 的思想从用户态提升到内核态,同时为用户提供了方便的共享机制.ECFS 在加密数据的基础上提出了校验数据散列值(Hash value)的方式,提供了数据的完整性保护功能.Cepheus 提出了三方架构的模式,提出一个可信的第三方服务器进行用户密钥的管理,引入了锁盒子机制进行用户分组管理,同时提出了懒惰权限撤销的思想.TCFS 提出了多级密钥的加密方式,使用一个主密钥加密原来的文件密钥.

随着网络存储系统的发展,加密文件系统的理念也逐渐网络化、系统化,最终演变成安全网络存储系统.一般的安全网络存储系统至少包括客户端与服务器两部分,客户端由系统的使用者进行操作,为用户数据提供数据加解密、完整性校验以及访问权限控制等功能;服务器作为数据及元数据的存储介质,对数据没有任何的访问或使用权限.

安全网络存储系统中比较典型的有 Plutus[16],SAND[17]以及 Corslet[18]等.Plutus 是 Cepheus 思想在网络存储系统中的扩展.在 Plutus 系统中,客户端负责所有的密钥分发与管理,在共享过程中为用户数据与元数据提供端到端的机密性和完整性保护.SAND 提出了一种密钥对象的数据结构,为网络存储系统提供了端到端的安全解决方案.Corslet 是一个栈式文件系统,通过引入可信第三方服务器,消除了用户对底层存储系统的依赖,在不可信的网络环境下为用户提供端到端的数据私密性、完整性的保护以及区分读写的访问权限控制功能.

云存储的廉价、易扩展等特性使得它一出现就成为人们研究的热点,由于用户将数据存放在云存储中便意味着丧失了对数据的绝对控制权,云存储系统对安全性有着十分迫切的需求.在这种需求的驱使下,Microsoft 于 2009 年提出了 Cryptographic

Cloud Storage[9]. Cryptographic Cloud Storage 系统以加密的方式为数据提供机密性保护、以审计的方式为数据提供持有性保护,同时为系统提供了细粒度的访问控制功能,并在系统的原型设计中使用了可搜索的加密机制(searchable encryption)、基于属性的加密机制(attribute-based encryption)、数据持有性证明(probable of data possession)等技术,在提高系统整体性能的同时增强了用户的体验效果.

同业界一样,学术界也很重视云存储系统的安全问题. 2010 年,Tang 等人在 FADE[19] 系统中提出了一种解决云存储系统中数据可信删除(assured delete)的方法;Mahajan 等人在 Depot[20] 系统中提出了一种最小化云存储中可信任(可用性方面)实体的方式,只要有一个正确(可访问)的客户端或服务器上有用户需要的数据,用户就可以通过网络获取到正确的数据;Shraer 等人在 Venus[21] 系统中提出了基于一个核心集(core Set)的信任体系,通过三方架构的方式为用户提供安全功能. 2011 年,Bessani 等人在 DEPSKY[22] 中提出了云中云的思想,在一定程度上减轻了数据机密性问题和运营商锁(vendor data lock-in)的问题.

3 安全云存储系统的一般架构

云存储按照其体系结构可分存储层、基础管理层、应用接口层和访问层[23]. 在具体的安全云存储系统中,由于应用场景和研究目标的不同,其系统架构也各不相同. 图 1 总结归纳了现有安全云存储系统的通用架构,具体的安全云存储系统只需根据自身的特点实现部分或全部的功能.

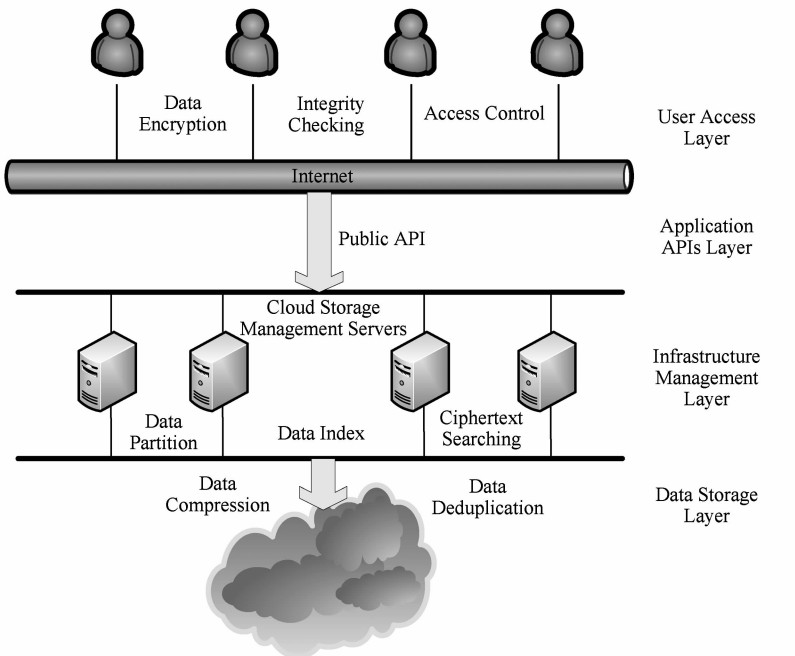

Fig. 1 Security cloud storage system generic architecture.

图 1 安全云存储系统通用架构

在一般的安全云存储系统中,数据访问层进行加密,然后通过应用接口层的公有 API 接口上传至云存储管理服务器,也就是基础管理层. 基础管理层可提供数据分块存储、建立数据索引、支持数据密文搜索等功能提高系统效率和用户体验. 最后,基础管理层将数据密文和其附加信息(一般为元数据,用来保证系统功能的正确性和高效性)通过安全高速的内部网络保存至存储层. 存储层可以对上层存进来的数据进行一定的压缩、删冗处理,以节省成本、提高存储空间的利用率.

现有的安全云存储系统一般分为客户端、服务器和云存储服务提供商 3 个组件,其中客户端属于访问层,服务器属于基础管理层,云存储服务提供商属于存储层. 客户端与服务器之间通过公有 API 及不可信的网络进行数据交互,服务器与云存储之间通过高速的可信网络传递数据. 用户数据和访问权限信息的机密性、完整性都由客户端保障,服务器可以记录一些数据的相关信息为用户提供数据同步、数据搜索等功能,但是在任何情况下服务器均无法获得用户数据的明文. 云存储服务提供商的作用

相当于过去的磁盘(或磁盘阵列),用来机械式地存取数据.

4 安全云存储系统的关键技术

为了保证安全云存储系统的正确性和高效性,不同系统的设计者往往会根据自己系统的特征,为系统添加一些特定的解决方案,这些解决方案便称为安全云存储系统中的关键技术.在不同的系统中所使用的关键技术也不尽相同.特别是随着云存储的发展与应用,一些在传统安全网络存储系统中所不关注的技术在安全云存储系统中却受到了重视.现有的云存储系统中所使用到的关键技术大致可分为以下几类.

4.1 安全、高效的密钥生成管理分发机制

在目前的安全云存储系统中,数据加密存储是解决机密性问题的主流方法[7-8].数据加密时必须用到密钥,在不同系统中,根据密钥的生成粒度不同,需要管理的密钥数量级也不一样.若加密粒度太大,虽然用户可以很方便地管理,却不利于密钥的更新和分发;若加密粒度太小,虽然用户可以进行细粒度的访问权限控制,但密钥管理的开销也会变得非常大.现有的安全云存储系统大都采用了粒度偏小或适中的加密方式,在这种方式下系统将会产生大量密钥.如何安全、高效地生成密钥并对其进行管理与分发是安全云存储系统中需要解决的重要问题.

4.1.1 密钥的生成机制

密钥生成关键在于如何减少需要维护的密钥数量和能够高效处理密钥的更新.目前的安全云存储系统所采用的密钥生成机制主要有以下 3 种.

1) 随机生成

随机生成密钥是最直接产生对称密钥的方式,CRUST[24]和 Plutus 等系统均采用了这种方式产生对称密钥对数据进行加密.这种加密方式具有良好的私密性和可扩展性,数据内容不容易被破解,但是密钥不能用作其他用途(例如数据的完整性校验),生成的数据密文随机性较强,不利于系统的重复数据删除操作.

2) 数据收敛加密

使用数据明文的某种(或多种)属性生成密钥对数据本身进行加密,使得相同数据明文经过加密后,生成的密文也相同的技术被称为数据收敛加密技术[25].Corslet 系统利用收敛加密的思想提出了一种数据自加密的方式(如图 2 所示),通过每个文件块的散列值与偏移量作为密钥,对文件块本身进行加密.

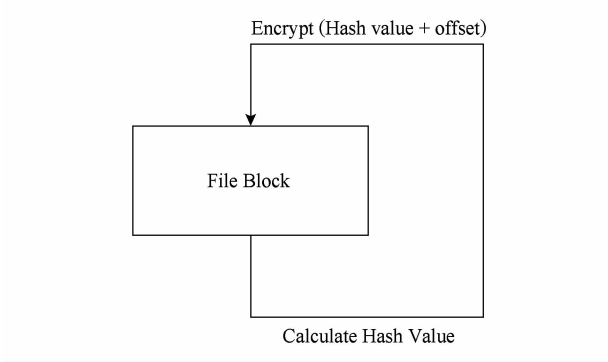

Fig. 2 Data self-encryption.

图 2 数据自加密

数据收敛加密的好处主要体现在以下几个方面:

① 若密钥的生成方式与数据的散列值有关,生成的密钥则可以用来校验数据的完整性,从而节省了存储空间;

② 修改数据的同时会修改密钥,因此特别适合懒惰权限撤销;

懒惰权限撤销[14]是指在基于共享的安全云存储系统中,若某个用户的访问权限被撤销,系统并不立即更换密钥对数据重新进行加密,而是采用触发的方式,当某个特定的事件发生时才对数据重新加密,例如,使用自加密技术后,若某个用户的访问权限被撤销,系统只需在访问控制信息中删除此用户的相关信息,待下次写操作发生时再对数据重新加密即可;

③ 相同内容的文件加密后密文依然相同,非常适合在系统中进行重复数据删除操作.

3) 通过特殊计算生成

在一些特定的应用场景中,为了提供一些特殊的功能,有时对文件密钥的生成也有一些特殊的要求,例如 Vanish 系统为了提供可信删除机制,要求密钥能够分成 m 份,用户只需要取得其中 n 份就能够解密文件.通过特殊计算生成的密钥通常是为了实现某个特定的功能,丧失了一定的通用性.

4.1.2 密钥的管理机制

目前的安全云存储系统大都采用分层密钥管理方式,其基本思想是将所有的密钥以金字塔形式排列,上层密钥用来加密/解密下层密钥.这样层层加密后,用户只需要管理位于金字塔尖的密钥,其他的密钥均可以放在不可信的环境中,或者以不可信的方式进行分发传递.因此,分层密钥管理方式可以在保证系统安全性的前提下,将大量的密钥交给不可

信的实体进行管理,用户及可信实体只需要保存极少量的密钥就可以达到以前的效果,大大提高了用户的方便性.

安全云存储系统大都采用 2～3 层的密钥管理方式.一般来说,无论某个系统将密钥分为多少层,我们都可以将它看成两层——顶层和其他层.现有系统在管理与分发顶层密钥时大都采用了 PKI[26] 体系中的公私钥算法,或是直接交给一个可信的第三方进行.相对地,其他层密钥可以直接存放在云存储中,合法用户在需要时从云存储中下载即可.

通过分层密钥管理的方式,安全云存储系统中的众多密钥可以被高效地组织起来,在保证数据私密性和完整性的同时,能够大量减少用户在密钥管理方面的开销,提高系统的效率,也有利于用户身份认证、访问授权等功能的安全实现.

4.1.3 密钥的分发机制

安全云存储系统大都具有共享功能,从而有了密钥分发的需求.一般来说,安全云存储系统中的密钥有以下 3 种分发方式:

1) 通过客户端进行分发

通过客户端对密钥进行分发是一种较老的分发方式.在这种方式下,服务器在任何情况下都不接触任何形式的密钥,因此安全程度很高.这种方式的缺点是要求客户端一直在线,一旦数据拥有者下线,数据的被共享者将因为无法获取密钥而不能访问数据.

2) 密文形式通过云存储进行分发

密钥经加密后存放在云存储中,数据被共享者访问数据时需要先从云存储中获取到数据密文和加密后的密钥,然后通过某种约定的方式(例如公私钥加解密方式)解密出密钥明文,随即再解密出数据明文.这种密钥分发的方式目前是业界中的主流方法,SpiderOak[27], Wuala[28] 等系统都是采用这种方式进行密钥分发.这种方式的优点是充分利用云存储的存储资源,采用了成熟的加解密技术,并可以随时对密钥进行发放;其缺点是过于依赖云存储,同时密钥冗余量太大,存储资源浪费较严重.

3) 通过第三方机构进行分发

密钥分发除了通过客户端和云存储进行之外,还可以通过与客户端和服务器独立的"第三方"进行.FADE 系统和 Corslet 系统使用一个可信的第三方服务器,用来集中管理分发密钥;Vanish[29] 系统通过 DHT 网络进行密钥分发.通过第三方机构的密钥分发方式结合以上两种方式的优点,但对应用场景的依赖较强,因此大都出现在某些特定的应用中.

4.2 基于属性的加密方式

在公私钥加密体系中有一种特殊的加密方式:基于属性的加密方式(attribute-based encryption)[30-33].基于属性的加密方式以属性作为公钥对用户数据进行加密,用户的私钥也和属性相关,只有当用户私钥具备解密数据的基本属性时用户才能够解密出数据明文.例如:用户 1 的私钥有 A,B 两个属性,用户 2 的私钥有 A, C 两个属性,若有一份密文解密的基本属性要求为 A 或 B,则用户 1 和用户 2 都可以解密出明文;同样,若密文解密的基本属性要求为 A 和 B,则用户 1 可以解密出明文,而用户 2 无法解密此密文.

基于属性的加密方式是在公钥基础设施(PKI)体系的基础上发展起来的,它将公钥的粒度细化,使得每个公钥都包含多个属性,不同公钥之间可以包含相同的属性.基于属性的加密机制有以下 4 个特点[34]:1)资源提供方仅需要根据属性加密数据,并不需要知道这些属性所属的用户,从而保护了用户的隐私;2)只有符合密文属性的用户才能解密出数据明文,保证了数据机密性;3)用户密钥的生成与随机多项式或随机数有关,不同用户之间的密钥无法联合,防止了用户的串谋攻击;4)该机制支持灵活的访问控制策略,可以实现属性之间的与、或、非和门限操作.

安全云存储系统中基于属性的加密方式其研究点在于:如何在系统中使用这种新的加密机制提高其服务效率与质量,而不是加密方式本身.基于属性的加密方式其特点使得它非常适合于模拟社区之类的应用[9].但是,目前基于属性的加密方式其时间复杂度很高、系统面向群体的安全需求很少的特点,使得这种加密方式目前的安全云存储系统中的应用并不广泛.随着安全云存储系统研究的进一步深入、属性加密方式的时间复杂度的降低,未来的安全云存储系统中一定会广为使用这种新的加密机制.

4.3 基于密文的搜索方式

一些云存储系统中添加了数据搜索的机制,使得用户可以高效、准确地查找自己所需要的数据资源.在安全云存储系统中,为了保证用户数据的机密性,所有数据都以密文的形式存放在云存储中,由于加密方式和密钥的不同,相同的数据明文加密后所生成的数据密文也不一样,因此无法使用传统的搜索方式进行数据搜索.

为了解决这个问题,近年来一些研究机构提出

了可搜索加密机制(searchable encryption)[35-38],能够提供基于数据密文的搜索服务.目前可搜索加密机制的研究可分为基于对称加密(symmetry key cryptography based)的 SE 机制和基于公钥加密(public key cryptography based)的 SE 机制两类.基于对称加密的 SE 机制主要是使用一些伪随机函数生成器(pseudorandom function generator)、伪随机数生成器(pseudorandom number generator)、散列算法和对称加密算法构建而成,而基于公钥加密的 SE 机制主要是使用双线性映射等工具,将安全性建立在一些难以求解复杂性问题之上.基于对称加密的 SE 机制在搜索语句的灵活性等方面有所欠缺,并只能支持较简单的应用场景,但是加解密的复杂性较低.而基于公钥加密的 SE 机制虽然有着灵活的搜索语句,能够支持较复杂的应用场景,但搜索过程中需要进行群元素之前和双线性对的计算,其开销远高于基于对称加密的 SE 机制.

在安全云存储系统中,基于对称加密的 SE 机制比较适用于客户端负责密钥分发的场景:当数据共享给其他用户时,数据所有者需要根据用户的搜索请求产生相应的搜索凭证,或将对称密钥共享给合法用户,由合法用户在本地产生相应的搜索凭证进行搜索.基于公钥加密的 SE 机制则更加适用于存在可信第三方的应用场景:用户可以通过可信第三方的公钥生成属于可信第三方的数据,若其他用户想要对这些数据进行搜索,只需要向可信的第三方申请搜索凭证即可.目前 SE 机制的难点与发展方向在于如何提高效率且支持灵活查询语句,以及如何保留数据明文中的语义结果.随着可搜索加密机制的逐步完善,安全云存储系统中对数据密文搜索的关联度、准确度以及效率方面将会越来越高,越来越多的安全云存储系统将会选择添加 SE 机制进行搜索.到那时,安全云存储系统的应用范围将更加广泛.

4.4 基于密文的重复数据删除技术

在一般的云存储系统中,为了节省存储空间,系统或多或少会采用一些重复数据删除(data deduplication)技术[39]来删除系统中的大量重复数据.但是在安全云存储系统中,与数据搜索问题一样,相同内容的明文会被加密成不同的密文,因此也无法根据数据内容对其进行重复数据删除操作.比密文搜索更困难的是,即使将系统设计成服务器可以对重复数据进行识别,由于加密密钥的不同,服务器不能删除掉其中任意一个版本的数据密文,否则有可能出现合法用户无法解密数据的情况.

目前对数据密文删冗的研究仍然停留在使用特殊的加密方式,相同的内容使用相同的密钥加密成相同的密文阶段,并没有取得实质性的进展.Storer 等人在 2008 年提出了一种基于密文的重复数据删除的方法[40],该方法采用收敛加密技术,使得相同的数据明文的加密密钥相同,因此在相同的加密模式下生成的数据密文也相同,这样就可以使用传统的重复数据删除技术进行对数据进行删冗操作.除此之外,近年来并无真正基于相同明文生产不同的密文的问题提出合适的解决办法.

重复数据的删除是安全云存储系统中很重要的部分,但目前的研究成果仅限于采用收敛加密方式,将相同的数据加密成相同的密文才能在云存储中进行数据删冗操作.因此,如何在加密方式一般化的情况下对云存储中的数据进行删冗是安全云存储系统中的一个很有意义的研究课题.

4.5 基于密文的数据持有性证明

在安全云存储系统中,用户数据经加密后存放至云存储服务器,但其中许多数据可能用户在存放至服务器后极少访问,例如归档存储等.在 Twinstrata 公司 2012 年的调查报告中,这类应用在云存储系统的使用中占据不小的比例.在这种应用场景下,即使云存储丢失了用户数据,用户也很难察觉到,因此用户有必要每隔一段时间就对自己的数据进行持有性证明检测,以检查自己的数据是否完整地存放在云存储中.

目前的数据持有性证明主要有可证明数据持有(provable data possession, PDP)和数据证明与恢复(proof of retrievability, POR)两种方案.PDP 方案通过采用云存储计算数据某部分散列值等方式来验证云端是否丢失或删除数据,文献[41]最早提出了远程数据的持有性证明,通过基于 RSA 的散列函数计算文件的散列值,达到持有性证明的目的.在此之后,许多文献各自采用了同态可认证标签[42]、公钥同态线性认证器[43]、校验块循环队列[44]以及代数签名[45]等结构或方式,分别在数据通信量、计算开销、存储空间开销以及安全性与检查次数等方面进行了优化.POR 方案在 PDP 方案的基础上添加了数据恢复机制,使得系统在云端丢失数据的情况下仍然有可能恢复数据.最早的 POR 方案通过纠删码提供数据的可恢复机制[46],之后的工作在持有性证明方面作了一定的优化,但也大都使用纠删码机制提供数据的可恢复功能.

云存储的不可信使得用户有着数据是否真的存放在云端的担忧,从而有了数据持有性证明的需求.现有的数据持有性证明在加密效率、存储效率、通信效率、检测概率和精确度以及恢复技术方面仍然有加强的空间.此外,由于不同安全云存储系统的安全模型和信任体系并不相同,新的数据持有性证明应该考虑到不同的威胁模型,提出符合相应要求的持有性证明方案,以彻底消除安全云存储系统中用户数据在存储过程中是否完整的担忧.

4.6 数据的可信删除

云存储的可靠性机制在提高数据可靠性的同时也为数据的删除带来了安全隐患:数据存储在云存储中,当用户向云存储下达删除指令时,云存储可能会恶意地保留此文件,或者由于技术原因并未删除所有副本.一旦云存储通过某种非法途径获得数据密钥,数据也就面临着被泄露的风险[19].为了解决这个问题,2007 年 Perlman 等人在文献[47]中首次提出了可信删除(assured delete)的机制,通过建立第三方可信机制,以时间或者用户操作作为删除条件,在超过规定的时间后自动删除数据密钥,从而使得任何人都无法解密出数据明文.Vanish[29]系统中提出了一种基于 DHT 网络的数据可信删除机制:用户在发送邮件之前将数据进行加密,然后将加密密钥分成 n 份存放在 DHT 网络中,邮件的接收者只需要拿到 $k(k\leqslant n)$ 份密钥就能够正常地解密,所有的密钥在超过规定的时间后将自动删除,使得在超过规定的时间后任何人无法恢复数据明文.

FADE[19]系统在文献[47]的基础上提出了一种基于策略(policy-based)的可信删除方式:每个文件都对应一条或多条访问策略(访问策略类似于属性加密(attribute-based encryption)机制中的属性,例如 Bob 可以访问和 2013 年之前是两条不同的策略),不同的访问策略之间可以通过逻辑“与”和逻辑“或”组成混合策略,只有当文件的访问者符合访问策略的条件时才能解密出数据明文.在具体的实现中,首先随机生成一个对称密钥 K 加密文件,然后为每个访问策略生成一个随机密钥 S_i,并按照混合策略的表达式对对称密钥 K 进行加密.第三方可信的密钥管理服务器(key manager)为每一个 S_i 生成一个公私钥对,客户端使用此公钥加密 S_i 后,将数据密文、对称密钥 K 的密文以及 S_i 的密文保存在云存储端.当数据删除操作发生或策略失效时,密钥管理服务器只需要删除相应的私钥就能够保证数据无法被恢复,从而实现了数据的可信删除.

云存储不可控的特性产生了用户对数据的可信删除机制的需求,目前在数据可信删除方面的研究还停留在初始阶段,需要通过第三方机构删除密钥的方式保证数据的可信删除.因此在实际的安全云存储系统中,如何引入第三方机构让用户相信数据真的已经被可信删除,或是采用新的架构来保证数据的可信删除都是很值得研究的内容.

5 总结与展望

安全云存储系统是云存储领域中的一个重要研究方向.本文介绍了云存储系统的安全需求、安全云存储系统的现状和一般架构,详细阐述了在现有安全云存储系统中所使用到的一些关键技术,并指出了未来安全云存储系统的研究方向.

云存储的服务性质让用户失去了对数据的绝对控制权,从而产生了云存储环境中特有的安全隐患.为此,安全云存储系统的设计者根据不同的应用场景,提出安全假设并建立相应的威胁模型与信任体系,采用合适的关键技术,设计并实现了各式各样的安全云存储系统.然而,任何安全机制都是有代价的,安全云存储系统无论是在效率上还是方便性上都要低于一般的云存储系统.

从总体上看,未来安全云存储系统的研究方向是在保证用户数据和访问权限信息安全的前提下,尽可能地提高系统效率,并提供一般云存储系统所具备的功能.目前安全云存储系统在密文的搜索、重复数据删除、数据持有性证明等功能的支持上仍有待加强,需要大家共同研究与探索.由于安全云存储系统有着无可比拟的安全优势,相信在不久的将来,它会逐步取代一般云存储系统,在实际应用中接受用户的检验.

参 考 文 献

[1] Twinstrata [EB/OL]. [2012-05-10]. http://www.twinstrata.com

[2] Cloud storage [EB/OL]. [2012-05-10]. http://en.wikipedia.org/wiki/Cloud_storage

[3] Cloud computing [EB/OL]. [2012-05-10]. http://en.wikipedia.org/wiki/Cloud_computing

[4] Amazon simple storage service [EB/OL]. [2012-05-10]. http://aws.amazon.com/s3

[5] Microsoft Azure [EB/OL]. [2012-05-10]. http://www.microsoft.com/windowsazure/

[6] A survey of data disclosing in 2010 by Verizon [EB/OL]. [2012-05-10]. http://netsecurity.51cto.com/art/201008/215676.htm

[7] Rafaeli S, David Hutchison D. A survey of key management for secure group communication [J]. ACM Computing Surveys, 2003, 35(3): 309-329

[8] Riedel E, Kallahalla M, Swaminathan R. A framework for evaluating storage system security [C] //Proc of the 1st Conf on File and Storage Technologies. Berkley: USENIX Association, 2002

[9] Kamara S, Lauter K. Cryptographic cloud storage financial cryptography and data security [C] //Proc of the 14th Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2010: 136-149

[10] Blaze M. A cryptographic file system for UNIX [C] //Proc of the 1st ACM Conf on Communications and Computing Security. New York: ACM, 1993: 9-16

[11] Blaze M. Key management in an encrypting file system [C] //Proc of Summer 1994 USENIX Technical Conf. Berkley: USENIX Association, 1994: 3-3

[12] Wright C P, Martino M C, Zadok E. NCryptfs: A secure and convenient cryptographic file system [C] //Proc of the USENIX 2003 Annual Technical Conf. Berkley: USENIX Association, 2003: 197-210

[13] Bindel D, Chew M, Wells C. Extended cryptographic file system [EB/OL]. [2012-05-10]. http://www.cims.nyu.edu/~dbindel/oceanstore/ecfs.pdf

[14] Group sharing and random access in cryptographic storage file systems [EB/OL]. [2012-05-10]. http://people.cs.umass.edu/~kevinfu/papers/fu-masters.pdf

[15] Cattaneo G, Catuogno L, Sorbo A D, et al. The design andimplementation of a transparent cryptographic file system for Unix [C] //Proc of the 2001 USENIX Annual Technical Conf. Berkeley: USENIX Association, 2001: 199-212

[16] Kallahalla M, Riedel E, Swaminathan R, et al. Plutus: Scalable secure file sharing on untrusted storage [C] //Proc of the 2nd Conf on File and Storage Technologies. Berkley: USENIX Association, 2003

[17] Miller E L, Freeman W E, Long D D E, et al. Strong security for network-attached storage [C] //Proc of the Conf on File and Storage Technologies. Berkley: USENIX Association, 2002: 1-13

[18] Xue Wei, Shu Jiwu, Liu Yang, et al. Corslet: A shared storage system keeping your data private [J]. Science China: Information Science, 2011, 54(6): 1119-1128

[19] Tang Y, Patrick P, Lee C, et al. FADE: Secure overlay cloud storage with file assured deletion [C] //Proc of the 6th Int Conf on Security and Privacy in Communication Networks. Berlin: Springer, 2010: 380-397

[20] Mahajan P, Setty S, Lee S, et al. Depot: Cloud storage with minimal trust [C] //Proc of the 9th Symp on Operation Systems Design and Implementation, Berkley: USENIX Association, 2010: 307-322

[21] Shraer A, Cachin C, Cidon A, et al. Venus: Verification for untrusted cloud storage [C] //Proc of the 2010 ACM Workshop on Cloud Computing Security Workshop. New York: ACM, 2010: 19-30

[22] Bessani A, Correia M, Quaresma B, et al. DEPSKY: Dependable and secure storage in a cloud-of-clouds [C] //Proc of the 6th Conf on Computer System. New York: ACM, 2011: 31-46

[23] Grossman R L, Gu Y, Sabala M, et al. Compute and storage clouds using wide area high performance networks [J]. Journal Future Generation Computer Systems, 2009, 25(2): 179-183

[24] Geron E, Wool A. CRUST: Cryptographic remote untrusted storage without public keys [C] //Proc of the 4th Int IEEE Security in Storage Workshop. Piscataway: IEEE, 2007: 3-14

[25] Douceur J R, Adya A, Bolosky W J, et al. Reclaiming space from duplicate files in a serverless distributed file system [C] //Proc of the 22nd Int Conf on Distributed Computing Systems. Piscataway: IEEE, 2002: 617-624

[26] PKI [EB/OL]. [2012-05-10]. http://datatracker.ietf.org/wg/pkix/charter/

[27] Spideroal [EB/OL]. [2012-05-10]. https://spideroak.com/

[28] Wuala [EB/OL]. [2012-05-10]. https://www.wuala.com/

[29] Geambasu R, Kohno T, Levy A A, et al. Vanish: Increasing data privacy with self-destructing data [C] //Proc of the 18th Conf on USENIX Security Symp. Berkley: USENIX Association, 2009: 299-316

[30] Goyal V, Pandey O, Sahai A, et al. Attribute-Based encryption for fine-grained access control of encryption data [C] //Proc of the 13th ACM Conf on Computer and Communications Security. New York: ACM, 2006: 89-98

[31] Cao N, Wang C, Li M, et al. Privacy-preserving multi-keyword ranked search over encrypted cloud data [C] //Proc of the IEEE INFOCOM 2011. Piscataway, NJ: IEEE, 2011: 829-837

[32] Curtmola R, Garay J, Kamara S, et al. Searchable symmetric encryption: Improved definitions and efficient constructions [C] //Proc of the 13th ACM Conf on Computer and Communications Security. New York: ACM, 2006: 79-88

[33] Ballard L, Kamara S, Monrose F. Achieving efficient conjunctive keyword searches over encrypted data [C] //Proc of the 7th Int Conf on Information and Communications Security. Berlin: Springer, 2005: 414-426

[34] Su Jinshu, Cao Dan, Wang Xiaofeng, et al. Attributed based encryption schemes [J]. Journal of Software, 2011, 22(6): 1299-1315 (in Chinese)
(苏金树, 曹丹, 王小峰, 等. 属性基加密机制[J]. 软件学报, 2011, 22(6): 1299-1315)

[35] Goldreich O, Ostrovsky R. Software protection and simulations on oblivious RAMs [C] //Proc of the 22nd Annual ACM Symp on Theory of Computing. New York: ACM, 1996: 431-473

[36] Song D X, Wagner D, Perrig A. Practical techniques for searches on encrypted data [C] //Proc of the 2000 IEEE Symp on Security and Privacy. Piscataway: IEEE, 2000: 44

[37] Waters B, Balfanz D, Durfee G, et al. Building an encrypted and searchable audit log [C] //Proc of the 11th Annual Network and Distributed System. Geneva: Internet Society, 2004

[38] Li M, Yu S, Cao N, et al. Authorized private keyword search over encrypted personal health records in cloud computing [C] //Proc of the 31st Int Conf on Distributed Computing Systems. Piscataway, NJ: IEEE, 2011: 383-392

[39] Data Deduplication [EB/OL]. [2012-05-10]. http://en.wikipedia.org/wiki/Data_deduplication

[40] Storer M W, Greenan K, Long D D, et al. Secure data deduplication [C] //Proc of the 4th ACM Int Workshop on Storage Security and Survivability. New York: ACM, 2008: 1-10

[41] Deswarte Y, Quisquater J J, Saidane A. A remote integrity checking [C] //Proc of IFIP ICCIS'03. Berlin: Springer, 2003: 1-11

[42] Ateniese G, Burns R, Curtmola R, et al. Provable data possession at untrusted stores [C] //Proc of the 14th ACM Conf on Computer and Communications Security. New York: ACM, 2007: 598-609

[43] Ateniese G, Kamara S, Katz J. Proofs of storage from homomorphic identification protocols [C] //Proc of the 15th Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2009: 319-333

[44] Xiao Da, Shu Jiwu, Chen Kang, et al. A practical data possession checking scheme for networked archival storage [J]. Journal of Computer Research and Development, 2009, 46(10): 729-736 (in Chinese)
(肖达,舒继武,陈康. 一个网络归档存储中实用的数据持有性检查方案[J]. 计算机研究与发展, 2009, 46(10): 1660-1668)

[45] Chen Lanxiang. Using algebraic signatures for remote data possession checking [C] //Proc of the 2011 Int Conf on Cyber-Enabled Distributed Computing and Knowledge Discovery. Piscataway, NJ: IEEE, 2011: 289-294

[46] Juels A, Kaliski B S. Pors: Proofs of retrievability for large files [C] //Proc of the 14th ACM Conf on Computer and Communications Security. New York: ACM, 2007: 584-597

[47] Perlman R. File system design with assured delete [C] //Proc of the 3rd IEEE Int Security in Storage Workshop. Piscataway, NJ: IEEE, 2007: 83-88

Fu Yingxun, born in 1986. PhD candidate. His research interests include network storage and cloud storage security systems.

Luo Shengmei, born in 1971. Master. His research interests include telecommunication network, VAS, mobile Internet, cloud computing, and so on.

Shu Jiwu, born in 1968. PhD, professor and PhD supervisor. Senior member of China Computer Federation. His main research interests include network storage and cloud storage, storage security, parallel process technologies, and so on.