# 可信云存储安全机制

张寿华　杨文柱　著

科学出版社

(TP-8252.31)

# 可信云存储安全机制

科学出版社互联网入口
信息技术分社：010-64000430 销售：010-64031535
E-mail：it@mail.sciencep.com
销售分类建议：计算机应用、云存储、信息安全

www.sciencep.com

ISBN 978-7-03-061381-3

9 787030 613813 >

定 价 108.00 元

2019

# 可信云存储安全机制

张寿华　杨文柱　著

科学出版社

北京

## 内 容 简 介

本书内容主要来自可信云存储安全领域的最新研究成果，系统阐述了可信云存储安全机制、可信云存储安全关键技术、可信云存储的体系结构，以及可信云存储中的加密存储与访问控制、可搜索加密、完整性证明、可用性保护和数据删除等技术。

本书可供可信云存储安全领域的工程人员及高等院校相关专业的师生阅读，也可作为相关领域科研人员的参考用书。

**图书在版编目（CIP）数据**

可信云存储安全机制 / 张寿华，杨文柱著. — 北京：科学出版社，2019.6

ISBN 978-7-03-061381-3

Ⅰ. ①可… Ⅱ. ①张… ②杨… Ⅲ. ①计算机网络－信息存贮－信息安全－研究 Ⅳ. ①TP393.071

中国版本图书馆 CIP 数据核字(2019)第 105125 号

责任编辑：董素琴 王 哲 / 责任校对：郑金红

责任印制：吴兆东 / 封面设计：迷底书装

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

北京虎彩文化传播有限公司 印刷

科学出版社发行 各地新华书店经销

*

2019 年 6 月第 一 版 开本：720×1 000 1/16

2019 年 6 月第一次印刷 印张：15 1/4

字数：300 000

**定价：108.00 元**

（如有印装质量问题，我社负责调换）

# 前　言

云存储是以云计算为支撑的新兴网络存储技术，是将存储资源放到云上供用户存取的一种新兴存储方案。用户可以在任何时间、任何地方，通过任何可联网的设备连接到云端并方便地存取数据。目前，以亚马逊云服务、谷歌 Drive、百度云和阿里云为业界翘楚，它们为全世界的客户提供云存储解决方案。这极大地降低了移动终端的存储开销，方便了用户数据接入和数据分享。云存储的优势是可以满足信息爆炸时代用户对数据存储的需求，且性价比极高，但随之而来的是云存储环境下的数据安全问题。云存储环境下，数据存储于云端，处于用户不可控域中，也就是用户在云端存储自己的敏感数据时，无法对安全风险进行直接控制，导致了比传统存储系统更多的安全问题。

可信计算是在计算和通信系统中广泛使用的、硬件安全模块支持下的技术，其目的在于提高系统整体的安全性。可信计算经过了十余年的发展，已经在构建基础性安全方面体现了其技术优势。目前大量的笔记本电脑、微软 Windows 操作系统、Linux 操作系统内核、Intel、AMD 等都直接采用或支持可信计算技术。在我国，可信技术更受到了国家的支持，国家密码管理局、全国信息安全标准化技术委员会都组织了可信计算相关标准的制定工作。

如何通过某种技术手段来保护云存储基础设施，确保基础设施在运行过程中按照预期的设计和部署来工作，是云存储必须解决的关键问题。可信计算是解决这一问题的核心技术。本书以可信云存储为主线，对涉及的相关理论、技术、方案等进行详细的介绍，以期为从事可信云计算的业内同行和初学者提供一本较为全面的参考书。

本书是近年来国内外相关研究的简要总结，也是项目组几年来研究成果的归纳。内容深入浅出，既有简明的理论介绍，也提供了丰富的解决方案。本书共分为 7 章，第 1～4 章由张寿华负责撰写，第 5～7 章由杨文柱负责撰写。全书由张寿华负责统稿。第 1 章主要介绍云存储、可信计算的概念，以及二者之间的关系，并给出可信云存储系统设计一般原则。第 2 章主要介绍可信云存储安全相关技术，以及可信云存储平台的体系结构。第 3 章介绍可信云存储中的加密存储与访问控制技术，涉及基于可信平台模块(trusted platform module，TPM)的可信云存储数据加密模型、可信云存储中的可信加密磁盘、属性加密机制、基于属性加密的可信云存储数据访问控制、基于 TPM 的密钥管理等。第 4 章主要介绍可信云存储中的可搜索数据加密，包括对称可搜索加密、公钥可搜索加密等。第 5 章介绍可信云存储中的数据完整性证明，涉及完整性证明模型、可信云存储中的数据完整性证明机制、可信云存储中

的动态数据完整性证明机制等。第 6 章介绍可信云存储中的数据可用性保护，主要包括多副本技术、容灾备份技术等。第 7 章介绍可信云存储中的数据删除技术，主要涉及数据销毁和确定性删除。

本书的出版得到了国家科技支撑计划项目子课题“组织机构代码管理服务云平台和安全体系的研究”(2013BAK07B04)的支持，也得到了科学出版社的大力帮助，特此感谢。感谢项目组刘振鹏教授、李昆仑教授、杨晓晖教授等对本书撰写的支持与帮助。

由于作者水平有限，本书难免存在不足之处，恳请广大读者批评指正。

作　者

2019 年 5 月

# 目　录

# 第1章 绪 论

近年来，云存储服务的用户数量出现了迅速的增长，亚马逊云服务、谷歌 Drive、百度云和阿里云等云存储服务应用，为全世界的客户提供云存储解决方案，降低了智能手机等移动终端的存储开销，提供便利的数据接入和数据分享。云存储系统承载了大量的用户隐私敏感性数据。因此，云存储的安全性和隐私性成为制约其未来发展的关键因素。

云存储中的虚拟机技术由于对客户机的高分离性和对资源的高可控性，大大提高了系统的安全性；而可信计算技术更是在硬件层上通过建立一个可信任根，解决系统的可信性和安全性问题，因此紧密结合可信计算和虚拟机技术，可确保云存储环境中用户数据和应用的安全。

## 1.1 云存储的安全性

云存储技术是在云计算的概念上发展出来的一个新概念，云存储与云计算几乎是同时兴起的，旨在通过互联网为用户提供更加优质的存储服务[1]。云存储用户可以随时随地使用终端设备通过网络连接到云存储数据中心，方便地进行数据存取操作。云存储的优势可以满足信息数据爆炸时代的人类对数据存储的需求，而且云存储能够以十分高的性价比来实现。但是随之而来的是云存储环境下的数据安全问题。云存储环境下，数据存储于云服务端，数据处于用户不可控域中，用户对敏感数据进行存储时，无法对风险进行直接控制，导致了相较于传统存储系统更多的安全问题。

2016 年，多家主流云存储服务商发生宕机事件。3 月 11 日，电商巨头亚马逊官方网站发生宕机事故，时间长达 20 分钟，事故不仅导致亚马逊电子商务主网站无法访问，而且波及了亚马逊的其他服务，其中就包括了全球最强的亚马逊云计算服务以及一些数字内容服务等，对于亚马逊来说这是一个巨大的事故，并且这一事故造成了巨大的经济损失。谷歌 2016 年不但发生宕机事故，而且发生了两次。4 月 11 日，由于两个漏洞的问题，谷歌云也全面陷入了一次 18 分钟的宕机。8 月 8 日，谷歌云存储及文件备份服务器服务终端再次宕机。云存储服务提供商最担心的就是长时间中断。7 月 30 日，微信公众号甚至出现大规模故障，不仅公众号文章无法查看，所有分享到微信朋友圈的文章也都显示网页出错。微信方面官方回应表示，服务器升级，部分用户朋友圈更新出现延迟现象。1 月 18 日，微软 Office 365 出现宕机，

且此次宕机只是个开始，在之后的几个月里，微软连续发生宕机事件，部分用户的电子邮件服务甚至连续 9 天无法收发邮件，堪称 2016 年最长宕机。宕机的原因在于服务器升级出现问题，云基础设施进行升级的目的是防止未来发生服务中断，而对于微软而言，还没开始便已经结束了。对于一直以技术领先的苹果而言，也曾因服务器问题而导致用户无法享受到正常的服务。6 月 2 日，苹果服务器出现问题，包括应用商店在内的部分服务不能正常使用，直到 6 月 3 日，苹果官方才表示包括 AppStore、iCloud 等在内的所有服务恢复正常。因此，云存储的安全性和隐私性成为制约其未来发展的关键因素。

## 1.2　可信计算技术

最近十几年可信计算经历了一段高潮发展阶段。可信计算已经在世界范围内取得了丰硕的成果，它由于可以确保系统资源和数据的完整性，所以可以确保计算机系统在一定条件下无恶意代码。可信计算是一种计算机系统安全的共性技术，凡是采用计算机的系统，都需要可信计算技术。

可信计算这个词来源于可信系统，1985 年美国国防部制定了世界上第一个《可信计算机系统评价准则》(trusted computer system evaluation criteria，TCSEC)。在 TCSEC 中第一次提出可信计算机和可信计算基(trusted computing base，TCB)的概念，并把 TCB 作为系统安全的基础。作为补充，TCSEC 又相继提出了可信数据库解释(trusted database interpretation，TDI)和可信网络解释(trusted network interpretation，TNI)。之后的可信计算则是由 IBM、Intel、AMD、HP 和微软等许多业界巨头组成的可信计算组织(trusted computing group，TCG)推动和开发的技术[2]。TCG 提出了可信平台模块(trusted platform module，TPM)的概念，并于 2009 年完成了 TPM 的标准化规范 ISO/IEC 11889。

TCG 提出的 TPM 以核心可信度量根作为起点，在将控制权交给下一个环节前，先通过度量的手段评估下一个环节的安全性，在确定下一个环节可信之后再将控制权转交给下一环节，之后再对下下个环节进行度量，由此建立一条信任链，完成对整个系统的完整性度量工作。这些度量值将保存在 TPM 内部的平台状态寄存器(platform configuration register，PCR)中。这些寄存器的内容只能通过重置或扩展两种操作修改，具备防止重放攻击的特性。验证者可以通过远程证明的方式要求 TPM 对这些 PCR 值进行签名，以便对 TPM 所在平台的安全状态进行验证。此外，TPM 还能将一些安全敏感数据与计算机的状态绑定，即与 PCR 值绑定，从而保证该数据只有在计算机处于特定状态时才能被解封。TPM 通过上述核心安全功能为用户提供从底层系统到上层应用的密码保护和可信证明[3]。

## 1.3 可信计算与云存储安全

可信计算经过了十余年的发展，已经在构建基础性安全方面体现了其技术优势。目前大量的笔记本电脑、微软 Windows 操作系统、Linux 操作系统内核、Intel、AMD 都直接采用或者支持可信计算技术。在我国，可信技术更受到了国家的支持，国家密码管理局、全国信息安全标准化技术委员会都组织了可信计算相关标准的制定工作。在技术层面，可信计算提出了在系统中构建信任根和信任链的基本思路，将系统的安全性置于一种来自于管理和技术相结合的安全基础上，通过可信度量和可信报告等技术手段实现一种行为序列可预期和环境状态可证明的可信。

如何通过某种技术手段来保护云存储基础设施，确保基础设施本身在运行过程中是按照预期的设计和部署来进行工作的，这一需求恰恰是可信计算能够有效提供的[4]。可信计算技术强调信任度量，度量就是建立于信任根基础上不断地对被度量实体进行完整性校验的过程。TCG 的信任度量是静态的度量，它适合对配置和静态的存储状态进行度量。这也是云存储基础设施安全必须达到的一个要求，即当前运行的云存储基础设施是最初基础设施搭建者所配置的；同时，还需要确保当前云存储基础设施上所运行的部件是最初在搭建平台基础设施过程中所预期部署的那些部件，不存在预期而没有部署、预期而没有加载运行、预期而没有按照某种依赖关系加载运行等状况。这些需求都是可信计算技术能够解决的。

由于云存储是一种开放式资源提供模式，对于云存储服务的使用者而言，其本身会有相当多的个性化数据置于云存储服务中，又由于云存储服务方式是依赖于 Internet 这样的网络的，所以资源的使用者对于其个性化数据资源的载体是不可物理直接接触的，资源使用者将处于一种弱势的地位，其会担心云存储服务的提供者对其个性化数据采取超出规约的处理。这一问题势必会影响云存储被大多数用户所接受，因此需要找到一种方式来平衡用户和服务提供者所处的地位。或者说需要提供给用户一种手段来检测其服务使用的安全性和合理性。

在可信计算中，一种称为远程证明的协议被提出，其核心思想是创建一种可证明的思想，即向远程访问者提供当前平台环境安全可信、符合访问者要求的凭证的思想。这对于云存储是尤其重要的。

云存储需要的是均衡云存储资源提供者和云存储资源访问者之间对资源、数据的控制能力，可信计算平台中，TPM 是一种超越了平台所有者的特殊模块，它与平台所有者的关系是证明性、绑定性的，而不是所有性的。可信平台所提供的度量与证明能力，是一种来源于硬件的度量与证明手段，其已经相对超越了资源所有者的普通权限。可信平台向资源访问者所提供的证明信息可以抵抗资源所有者的伪造，可信计算可以帮助云存储来克服资源提供者与资源请求者之间的不均衡性。同时，

远程证明中的证明方不再直接是云存储的构造者，其所提供数据的可信性是由存在其硬件之上的可信安全部件所提供的，数据来自于 PCR 中，其度量过程由平台启动部件和 TPM 等协作完成，每个部件的加载都受到验证，且加载顺序的正确性也由 PCR 的迭代哈希(Hash)运算保证，因此，远程证明与启动度量的方法在云存储环境中将是非常重要的技术手段。

但是，对于云存储而言，可信计算的度量方法又不能够直接使用。因为系统资源提供的上下文环境是系统处于运行态的，而可信计算主要能解决的却是静态信任度量。静态信任链是在整个系统的初期阶段进行度量并建立起来的，对于云存储来说，除了要考虑系统初始启动时的静态可信，还要考虑系统启动后的动态可信。云存储不是频繁启动的个人计算机，因此更应强调动态可信问题。应该指出，基础设施的可信仍然是整个云存储系统可信的基础，只有基础设施本身是可信的，即基础设施的加载过程是可信的，基础设施加载的各个模块是预先可认证的，基础设施加载的模块是能够保障上层其他安全模块的安全与可信性的，才能确保基础设施向外提供的一系列的安全服务和其他系统级别服务是可信的。由于云存储对外提供服务的阶段是在整个基础设施建立之后才形成的，此时早已经过了云存储基础设施静态信任链的建立阶段，因此可信计算中所提出的静态可信度量对于度量云存储服务的可信度不再是一个合适的手段，需要寻找一种非交互式的离线度量机制来实施信任度量。

用户的敏感数据存储也从终端迁移到了超越用户控制范围的云端来存储。因此，旧的用户的风险控制手段也应从终端迁移到云端。用户的信赖对象从终端本地平台转移到了云端的存储平台。信赖对象的改变促使人们思考如何提高云存储的可信性，安全可信成为云存储成败的关键。因此如何构建面向存储服务的可信的云存储成为一个关键问题。

## 1.4　可信云存储系统设计一般原则

可信云存储系统就是结合使用云存储安全技术和可信技术，保证云数据的安全和可靠的云存储系统。

可信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，可信云存储系统设计时需要考虑如下几个方面。

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。因此可信云存储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，称为可信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。在可信云存储系统中，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。

## 1.5 本章小结

云存储环境下，数据存储于云服务端，数据处于用户不可控域中，用户对敏感数据进行存储时，无法对风险进行直接控制，导致了相较于传统存储系统更多的安全问题。可信计算是一种新的信息安全技术，它已经成为国际信息安全领域的一个新热潮，并且取得了令人鼓舞的成绩。从可信计算领域的发展趋势来看，将可信计算技术与云存储安全技术相结合为解决云存储安全提供了新的思路，并成为解决云存储安全问题的重要研究方向。本章首先介绍了云存储的安全问题，然后介绍了可信计算技术，分析了利用可信计算技术解决云存储安全的可能性，以及可信云存储系统的设计原则。

## 参考文献

[1] 张继平. 云存储解析. 北京: 人民邮电出版社, 2013.

[2] 慈林林, 杨明华, 田成平, 等. 可信网络连接与可信云计算. 北京: 科学出版社, 2015.

[3] 代炜琦. 云计算执行环境可信构建关键技术研究. 武汉: 华中科技大学, 2015.

[4] 张焕国, 赵波. 可信计算. 武汉: 武汉大学出版社, 2011.

# 第 2 章　可信云存储安全

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，国内外的研究者做了大量研究。

可信计算技术通过增强体系结构的安全性来提高计算平台的安全性。从当前本领域研究的发展形式分析，云存储技术与可信计算技术融合以更好地解决云存储中的安全问题将成为一个重要方向。

## 2.1　可信云存储的安全需求

除传统的身份认证(网络钓鱼、密码泄露等)、底层系统安全(安全传输、弱随机数、侧信道攻击等)、物理安全等安全需求外，可信云存储用户面临的安全需求主要包括数据的安全性、密钥管理分发机制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

数据安全是可信云存储系统中最重要的安全需求之一，可信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、完整性和可用性三个方面。

1)数据的机密性

可信云存储系统中数据的机密性是指无论存储还是传输过程中，只有数据属主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露数据的可能性。

2)数据的完整性

可信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

可信云存储的不可控制性滋生了可信云存储系统的可用性研究。与以往不同的

是可信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可信云存储系统的安全需求之一。

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。数据加密需要密钥，可信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。

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可信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。可信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。

## 2.2　可信云存储安全技术

为了保证可信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。在不同的系统中所使用的解决方案也不尽相同，特别是随着云存储和可信计算的发展与应用，一些在传统安全网络存储系统中所不关注的技术在可信云存储系统中也受到了重视[2-5]。

1. 融入可信计算

随着云存储的不断发展和壮大，人们逐渐认识到单纯使用软件的方法难以解决云存储中所有的安全问题。尝试建立一种以硬件安全芯片为信任根的可信云存储环境成为云存储安全研究的一个方向。可信计算为平台提供了信任链构建以及平台可信证明的机制。然而，云存储安全研究仍处于初级阶段，很多问题有待进一步探索。将可信计算技术融入云存储中可谓是目前云存储安全最重要的研究方向之一。

2. 密钥技术

在目前的可信云存储系统中，数据加密存储是解决机密性问题的主流方法。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。现有的可信云存储系统大都采用了粒度偏小或适中的加密方式，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可信云存储系统需要解决的重要问题。

1)密钥的生成机制

密钥生成的关键在于如何减少需要维护的密钥数量和如何高效处理密钥的更新。目前的系统所采用的密钥生成机制主要有以下三种。

(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 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，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(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，使相同数据明文经过加密后，生成的密文也相同的技术称为数据收敛加密技术。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，将每个文件块的哈希值与偏移量作为密钥，对文件块本身进行加密。

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哈希值有关，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。

懒惰权限撤销是指在基于共享的可信云存储系统中，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。

2)密钥的管理机制

目前的可信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/解密下层密钥。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。

可信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，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