云存储系统可问责机制研究

陈冰泉 *
CHEN Bing-quan

摘 要 目前，很多厂商开发了云存储系统供企业和个人使用，但在实际应用中，大部分用户仍然不能完全信任云存储，其核心担忧仍然是云存储的安全，即用户存储在云端的数据被泄露、恶意攻击、窃取的可能性；另外，用户也很关心云存储系统出现问题时，事故责任怎么划分。针对上述问题，本文首先对云存储的安全问题进行了综述，在此基础上研究了云存储可问责机制，从问责的视角提出了解决云存储安全问题的方案。

关键词 云存储 安全 问责

doi：10. 3969/j. issn. 1672 - 9528. 2015. 1.22

1 云存储概述

云存储是由云计算衍生出来的。云存储是在宽带网络、WEB2.0 技术、应用存储、集群、网格技术、分布式文件系统、CDN 内容分发、存储虚拟化、存储网络化管理等技术的发展与成熟基础上发展而来。从技术的角度出发，云存储是一个有别于传统的存储的数据中心，其存储介质不再是昂贵的专有存储设备，而是采用大量廉价的硬盘，大幅度降低了成本；同时结合虚拟化技术，为每个用户动态调整所拥有的存储容量，具有很高的灵活性。从使用的角度出发，云存储就是一个远程的硬盘，用户使用浏览器或其他客户端与远程资源交互。云存储不单是一堆硬件，而是网络、存储、服务器、应用、接口、客户端等有机构成的复杂系统。云存储的体系结构分层模型参见图 1。

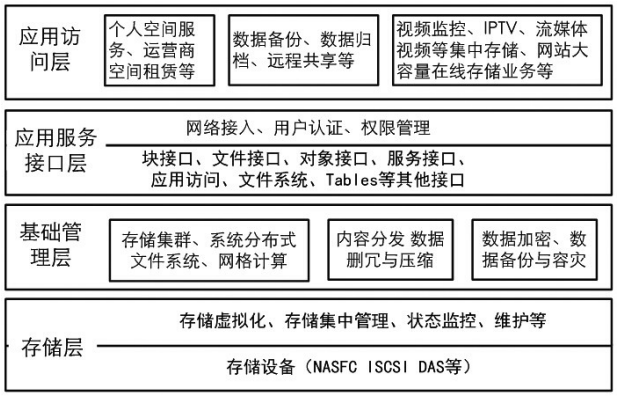

图 1 云存储系统结构模型

2 云存储的数据安全

云存储的用户群越来越庞大，但数据安全性问题仍然备受关注，严重阻碍云存储向企业用户的扩张。作为一种新的数据存储模式，云存储中的数据安全包含以下三个方面。一是云存储数据的机密性、完整性、可用性。数据机密性（私密性）就是防止信息经过通道被泄露的特性；数据完整性是指云存储中所有数据的值均正确的状态，如果存储的数据变成了不正确的状态，则称其丧失了数据完整性，如数据损坏；云存储的远程不可控制性带来了云存储的数据可用性问题。二是提供云存储服务的厂商不完全可信，可能做出损害用户数据安全的行为。云存储提供商有责任为用户提供可靠的存储服务，但是由于数据完全由云端保管，云存储提供商在缺乏监管的情况下有可能隐瞒用户数据的丢失损坏或泄露用户敏感数据。三是云存储行为监督问题。通常用户会与云存储提供商订立相应合约，但合约履行的情况缺乏实际监督。纠纷发生时，责任难划分，易于推卸责任。

上述因素使得传统网络安全和存储安全技术在云存储环境下不能直接适用，如云存储服务器本身不可信时，由其执行的访问控制也值得怀疑；普通的基于消息数字签名的完整性保护不能在数据在云端存储过程中执行实时的完整性检查。用户日益关注云存储的数据安全，技术的发展可解决一部分安全问题，但更多的问题是非技术问题，如何结合云服务模式的商业伦理与法律责任构建完善的问责机制对云存储服务市场的规范显得更加重要，本文将从云存储服务模式的商业伦理与法律责任，结合相关技术的角

* 工业和信息化部电子第五研究所 广东 广州 510610
基金项目：国家科技支撑计划项目“电子产品生态设计云服务平台”（2012BAH27F02）。

度探讨问责机制及其策略。

3 云存储系统问责机制

3.1 云存储与问责

目前关于问责的研究在国外学者远比国内走得远，很多国际组织将“问责”写入了云存储相关指南中；国内则处于起步阶段，缺乏系统性和切实可行的方法。

云存储不是一项全新的独立的技术，但却是一种服务模式的创新，其创新地提出了一套动态分配存储资源的运营机制。云存储的“动态效用”提供了廉价、快捷的弹性服务，与此同时，也带来了信任危机，引发了“问责”问题：一方面，用户数据主控权的丧失加之一直存在的网络安全问题，必然导致信任问题；另一方面，动态效用的复杂性增添了云存储商业部署的不确定性。这两方面的问题不能依靠单纯的技术手段来解决。问责机制是一种可选的解决方案。问责能协助建立云服务提供商与使用者之间的信任，是一种结合商业伦理与法律责任的义务，同时融合技术与法律两方面的支持的解决方案。

判断一个系统是否具备可问责的特性，即判断该系统是否提供了检测系统中参与者不当行为的手段，具体表现为：①能够可靠地发现系统中存在的错误，这是首要条件，否则问责无从谈起；②检测到错误后能追溯错误源；③过错方无法抵赖，也无法推卸责任。问责的实施可分为以下六个步骤（见图 2）：

图 2 云存储问责实施步骤

第一，信息记录。服务供应商判断并记录信息，包括：①事件数据，指一连串发生的动作及其相关信息；②角色数据，记录事件的主体，即谁导致了事件发生；③时间数据，记录事件发生时间；④地点数据，记录事件发生地址（虚拟 / 物理）。第二，感知与溯源。云端发生意外事件时，感知与溯源负责检测意外发生并触发记录动作。第三，日志记录 / 保存。确保日志的完整性，避免非法访问，考虑在某些情况下对敏感信息进行适当处理；此外，还应考虑日志回滚及灾备，以防日志丢失或被篡改。第四，报告通知：生成关于文件和审计结果的信息，标注可疑信息，并告知用户。第五，审计：一方面是中立第三方审计；另一方面，将某些审计设计成自动化的流程嵌入到系统。第六，优化与修正：系统安全漏洞的闭环，改进系统性能，提高系统的可靠性。

问责机制可将系统中发生的行为严格归属到其参与者，从而当过错发生便能够迅速定位过失，明确其责任，规范以合约承诺的服务双方的权益和义务的执行，密切结合云存储系统技术，解决单纯技术无法完美解决的云存储安全威胁。在具体实践上云存储的问责分为程序性措施与技术性措施，程序性措施包含选定云计算服务提供商、签署合同与服务等级协议、限制重要信息的流通、买保险、企业指派数据保护管理员、定期评估隐私安全性等；技术性措施包含数据加密保护、访问控制、隐私信息中介等。总而言之，云存储的问责机制是一种结合商业伦理与法律责任的义务，通过融合技术和程序实现的可行的数据安全解决方案。

3.2 云存储可问责解决方案

3.2.1 模型定义

模型的角色包括存储服务提供商 (Cloud Storage Provider, CSP)、用户 (User)、可信第三方 (Trusted Third Party ,TTP)：User 购买存储服务，将数据存储在云端；CSP 根据合约为用户提供存储服务，满足用户对数据的完整性和保密性的需求；TTP 依据标准化的指标对 CSP 的服务进行评估和认证，监控 CSP 的行为，为用户和云存储服务提供商生成双方都认可的凭单，在发生争议时来判定是谁的责任。

从对象被创建到其操作记录被删除的时间段为该对象的生命周期；用户和云存储服务提供商交互完成的每次操作都生成凭单项，构成凭单链表。凭单链表存储在云存储服务提供商处，方案借助可信第三方采用基于 Merkel 哈希树的算法确保凭单的完整性。方案同时给出具体可行的凭单生成协议与问责审计协议。基于可信第三方的云存储完整性问责模型的总体架构如图 3 所示。

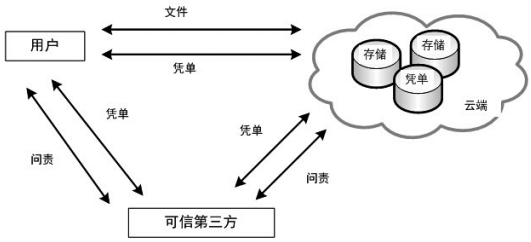

图 3 基于可信第三方的云存储问责模型

模型假设：

①云存储服务支持数据的版本管理，基本模型暂不考虑数据对象在不同用户间共享的应用场景。② User 删除对象后，CSP 在特定时间内仍保存该对象的操作记录，以记录用户的删除操作。③ CSP 不可信，可能对用户的数据进行删除、篡改；User 不可信，可能恶意破坏服务的可问责性，但二者不同谋；TTP 可信，保持中立。④ User 与 CSP 的通信安全可靠。⑤攻击者可部分或全部获取云端的用户数据，且对用户数据修改可不被 CSP 发现。⑥模型采用的密码学措施安全，同时假设各个角色的密码安全；

模型将系统中的数据流分成三类：①读写流，在用户和云端之间传递数据；②凭单流，为验证参与方的动作产生的数据流，为每个操作产生相应的证明凭据，用 TTP 审计；③审计流，检查 CSP 是否违反应该遵守的规则。User 和 CSP 服从 TTP 的审计结果，不可抵赖及推卸责任。

3.2.2 安全威胁描述

问责模型关注导致数据对象处于某状态的原因，即数据对象的各种操作。对象的每一个符合服务条款（也称为合法的）操作都同时涉及 User 与 CSP，表 1 列出了云存储服务的可问责模型可能遭受的威胁及可采取的问责方法。

表 1 云存储服务可问责可能遭受的威胁

威胁	问责方法
CSP 执行写失败	User 直接发现并拒绝
CSP/User 否认关于某个对象的更新、删除操作	对象操作审计
CSP 返回错误的读响应	对象操作审计
User 执行写错误	对象操作审计
CSP 恶意修改或丢弃某些对象	对象操作审计
CSP 否认创建对象	第三方判决
非授权的写	对象审计操作
非授权的读	可信第三方控制

表中给出的部分操作完成时可检测，无需问责，如 CSP 写失败；模型为不能被检测的行为提供了两类问责方法：①对象操作审计，即依据关于被质疑对象的可验证操作记录进行问责；②第三方判决，当争议不能通过对象操作审计得到解决时由 TTP 判决。

3.3 云存储问责机制实现

3.3.1 凭单定义

凭单是一种与用户和云端动作相关联的记录。凭单分两种：一是用于进行完整性保证和划分责任的持久性凭单；二是在 User、CSP、TTP 三方交互中动态生成的用来确认某些事件的临时性凭单。持久性凭单与每次写操作提交后产生的新版本数据文件对象一一对应，凭单随着对象版本的更新而增加，构成凭单链表，链表的头节点对应数据文件对象的最新版本，尾节点对应最初版本。TTP 记录关键的凭单信息来确保 CSP 会尽最大努力存储和保护用户的数据，CSP 存储凭单链表。凭单链表的连续性与完整性可确保 CSP 无法在逃避审计的情况下从凭单链表的中删除任何凭单。凭单的字段设置参见表 2。

表 2 凭单的字段信息

用户标识（user_id）	表征用户对象
对象标识（object_id）	表征数据文件对象
对象哈希值（object_hash）	数据文件内容对象的哈希值（MD5、SHA1）
动作（action_type）	用户发起的动作类型的标识
时间戳（time_stamp）	动作发生时间
预期版本号 p_version	用户请求操作时的预期版本号
密钥版本（key_version）	数据对象加密密钥的版本
用户签名（user_sig）	用户对凭单的签名
云端签名（csp_sig）	云端对凭单的签名
可信第三方签名（ttp_sig）	可信第三方对凭单的签名

3.3.2 用户动作及操作流程说明

①创建：User 新建数据文件对象，CSP 为每个对象分配全局唯一的对象编号 object_id 和 p_version。TTP 维护对象的 object_id 记录。

②更新：User 修改对象，更新其版本及凭单信息。

③读取：User 从 CSP 读取数据对象。

④删除：User 从 CSP 删除对象全部版本的内容，但不删除对象的凭单信息。

⑤审计：User 向 TTP 发送审计请求，TTP 收到请求后会对操作发起相应的审计。

3.3.3 数据安全保护

数据安全包含其机密性、完整性、可用性三方面。云存储可通过访问控制保护数据的机密性。模型认为 CSP

不可信，单纯的访问控制仍然存在安全威胁。因此采用对称加密和主副密钥结合，并由可信第三方和用户共同管控的方式实现访问控制。数据对象访问控制列表由TTP维护，审计时TTP还需判断用户的审计请求是否合法。模型采用的加密方案：数据对象拥有者为对象生成一个独立的在任何情况下不泄露的主密钥Secret Master Key，然后结合该密钥及对象（即密钥版本）生成对象的访问密钥，可共享：Access Key =Hash (Secret Master Key， Hash(object_id，key_version)。例如，用户读取或更新数据时，首先向对象拥有者提出申请，如果请求获得批准则告知TTP添加该用户到更新控制列表，同时对象拥有者共享该对象的访问密钥给请求者。

事实上，任何存储系统都不能完全确保数据的完整性。数据完整性保护主要通过冗余、备份、完整性检查、验证等技术手段实现，即使采取了各种技术措施，数据完整性仍然存在破坏的可能性。在基于合约的云存储中，对影响数据完整性的行为进行问责是一种可选的方案，将云存储提供的备份、拆分、容错、完整性检查等技术手段与可信第三方审计相结合，使得整个保护过程透明。云存储基于网络的动态效用模式增加了其可用性的威胁。可用性的保证必须结合数据完整性、机密性、硬件设施服务能力等。

3.3.4 问责审计

模型通过可信第三方完成对系统动作的记录的审计以划分事件责任，如User怀疑CSP自主篡改或删除数据时，User向TTP申请审计，TTP据判定依据首先检查申请的合理，如合理则对操作进行审计，然后将结果返回给用户，该审计流程参见图4。

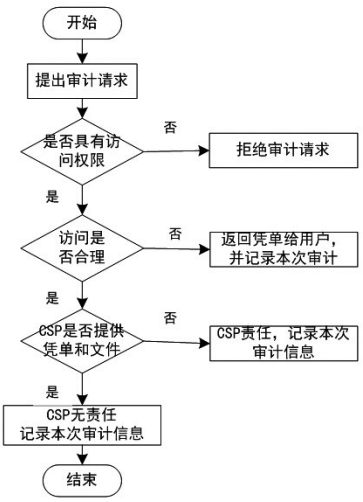

图4 审计流程

防篡改可信硬件也可作为方案中的TTP，可信硬件可替代独立可信机构，承担认证凭单及凭单链表的完整性任务。可信硬件的优势：简化系统结构；可信硬件与云服务器集成后，在云端进行对凭单的认证，大大降低了因网络带来的数据隐私泄露风险。

4 结语

针对云存储的安全问题，即用户存储在云端的数据被泄露、恶意攻击、窃取的等；针对用户云存储系统出现问题时，事故责任怎么划分的要求，本文首先对云存储的安全问题进行了综述，在此基础上研究了云存储可问责机制，从问责的视角提出了解决云存储安全问题的方案。

参考文献：

[1] Peter Mell,Timothy Grance, “ The NIST Definition of Cloud Computing ” [Z].NIST Special Publicaiton800-145,2011(9).

[2] 陈冰泉 . 云计算服务系统可靠性建模研究 [J]. 电子产品可靠性与环境试验 ,2014(4).

[3] 胡光永 . 基于云计算的数据安全存储策略研究 [J]. 计算机测量与控制 ,2011,19(10).

[4] 洪澄 , 张敏 , 冯登国 .AB-ACCS:—种云存储密文访问控制方法 [J]. 计算机研究与发展 ,2010(2).

[5] 王会波 . 安全存储与云存储安全 [J]. 信息安全与通信保密 ,2010(15).

[6] 范振华 . 云存储数据完整性验证和问责机制的研究 [D]. 保定 : 河北大学 ,2014.

[作者简介] 陈冰泉 (1985-)，女，四川大英人，工业和信息化部电于第五研究所数据中心工程师，质量工程师，硕士，主要从事云计算应用及其可靠性、系统可靠性相关研究工作。

（收稿日期：2014-12-06）