云计算
数据安全方案及其应用

张　华　金正平　李文敏　时忆杰

著

科学出版社

云计算数据安全方案及其应用

张　华　金正平　李文敏　时忆杰　著

科学出版社

北京

内 容 简 介

本书以作者及其课题组多年的研究成果为主体，结合国内外学者在云计算数据安全方面的代表性成果，以云计算中数据的生命周期为主线，系统介绍了云计算数据安全的理论和技术，主要内容包括云计算及数据安全概述、密码学基础、云存储安全、可搜索加密、远程认证、访问控制、外包计算等，并针对云计算在远程医疗信息系统中的应用，给出相应的安全解决方案. 全书重点从云计算数据生命周期的角度阐述了不同应用阶段数据安全的关键技术方案的设计与模拟实现.

本书既可作为对云计算数据安全方案感兴趣的读者的入门教材，也可作为云计算数据安全理论研究工作者的参考用书，适用于网络空间安全、密码学、信息安全、数学、计算机及相关学科的高年级本科生、研究生、教师和科研人员阅读参考.

图书在版编目(CIP)数据

云计算数据安全方案及其应用/张华等著. —北京：科学出版社, 2018.6
ISBN 978-7-03-057846-4

Ⅰ. ①云… Ⅱ. ①张… Ⅲ. ①云计算–网络安全 Ⅳ. ①TP393.08

中国版本图书馆 CIP 数据核字(2018) 第 129149 号

责任编辑：王丽平 / 责任校对：邹慧卿
责任印制：张 伟 / 封面设计：黄华斌

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
*
2018 年 6 月第 一 版 开本：720 × 1000 B5
2018 年 6 月第一次印刷 印张：24
字数：482 000

定价：149.00 元
(如有印装质量问题，我社负责调换)

前　言

云计算是当前信息技术领域的重要发展方向, 也是产业界、学术界、政府等都十分关注的焦点. 云计算体现了信息技术 (Information Technology, IT) 领域向集约化、规模化与专业化道路发展的趋势, 是 IT 领域正在发生的深刻变革. 但当前云计算的发展面临许多关键性问题, 其中安全问题首当其冲, 已成为制约其发展的重要因素. 特别地, 在云计算服务模式下, 用户将数据和应用托管至云端, 云服务的透明性使用户失去对数据的控制, 而云服务商可信性不易评估, 所以数据安全问题已成为云计算用户首要担忧的问题.

根据云计算数据的生命周期, 可以将其分成产生、存储、使用、共享、归档及销毁六个阶段. 以生命周期为主线, 云计算数据安全可以着重从以下几个方面考虑: 云存储安全、可搜索加密、远程认证、访问控制、外包计算等. 其中, 云存储安全主要解决在云计算模式下如何保护用户上传到云端的数据完整性和数据隐私的问题; 可搜索加密满足了在云端不可信条件下加密数据安全云检索的应用需求; 访问控制实现了对云端数据进行访问的用户授权过程; 远程认证主要致力于解决用户在获得合法授权后访问所需服务时身份的核对检验问题; 而外包计算主要是借助于云服务器超强的存储和计算资源, 通过代理等方式安全地实现资源受限用户的计算需求. 近年来, 云计算数据安全是该领域的研究热点, 出现了以学术论文和其他科研成果为代表的众多解决方案, 取得了较为显著的发展.

本书旨在系统阐述云计算数据安全的关键技术, 以作者及其课题组多年的研究成果为重点, 汇集了国内外学者在这方面的重要研究成果. 对相关领域的理论工作者、研究生以及高年级本科生的专业课题研究有一定的参考价值.

全书共八章. 第 1 章为绪论, 概述了云计算及其安全体系, 并从数据生命周期的角度, 分别阐述了云数据安全威胁、安全需求和数据安全总体框架. 第 2 章简要介绍了密码学的相关基础知识, 以便读者把握后续章节的具体方案, 更多内容可参阅其他相关文献. 第 3 章介绍了云存储的由来及其面临的安全挑战, 并就其数据的完整性验证和隐私性展开论述. 第 4 章是可搜索加密技术的展示及其在数据管理中的应用. 第 5 章列举了不同类型的远程认证, 包括基于 ElGamal 公钥密码体制、基于椭圆曲线公钥密码体制、基于双因素、基于三因素、单服务器、多服务器等不同技术手段或应用场景所设计的远程认证方案. 第 6 章展现了云计算数据安全中访问控制技术, 包含身份认证和授权两个方面, 其中前者主要讲述了单点登录技术, 而后者主要是基于不同工具的访问授权, 同时对访问控制中的身份管理进行

了简要介绍. 第 7 章讨论了外包计算模式下几类安全解决方案的设计, 包括基于特殊工具和具有特殊功能的属性加密方案以及安全多方计算协议. 最后, 作为在远程医疗方面的应用案例, 第 8 章简要介绍了远程医疗信息系统及其安全架构, 并针对该应用场景给出了认证、加密等安全解决方案.

最后, 作者对课题组成员郭子卿博士、于萍博士、赵少华硕士、龚云平硕士、尹亚平硕士等给予的密切配合, 以及北京邮电大学网络与交换技术国家重点实验室全体老师和学生的支持表示感谢. 另外, 本书的出版得到了以下项目的资助: 国家自然科学基金项目 (编号: 61502044)、中央高校基本科研业务费专项资金资助项目 (编号: 2015RC23) 等, 在此特别表示感谢.

由于时间仓促, 书中不妥之处在所难免, 恳请读者指正.

作　者

2017 年 3 月于北京

此为试读,需要完整PDF请访问: www.ertongbook.com

目　录

第1章 绪 论

随着云计算的快速发展, 其安全问题日趋显现. 特别是云计算下新的服务模式导致用户失去对数据的直接控制而引起的数据安全问题尤为突出. 本章从云计算的基本概念入手, 简要介绍云计算及其安全体系, 并以云计算中数据的生命周期为主线, 阐述云计算数据安全威胁、安全需求及其应对措施的总体框架.

1.1 云计算概述

从云计算概念的提出到不断推广和逐步落地, 其作为 IT 产业的革命性发展趋势已经不可逆转, 甚至被称为当今世界的第三次技术革命. 但到底什么是云计算, 却是众说纷纭, 有许多种定义, 让人云里雾里.

现阶段广为接受的是美国国家标准与技术研究院 (National Institute of Standards and Technology, NIST) 给出的定义[1]: 云计算是一种按使用量付费的模式, 这种模式提供可用的、便捷的、按需的网络访问, 进入可配置的计算资源共享池 (资源包括网络、服务器、存储、应用软件、服务), 这些资源能够被快速提供, 只需投入很少的管理工作, 或与服务供应商进行很少的交互. 这也是我们狭义上说的云计算. 而广义上的云计算是指服务的交付和使用模式, 指通过网络以按需、易扩展的方式获得所需的服务. 这种服务可以是 IT 和软件、互联网相关的, 也可以是任意其他的服务.

当然, 不管具体定义如何, 云计算应具有如下五个特征[1].

(1) 按需分配的自助服务. 消费者可以单方面地按需自动获取计算能力, 如服务器时间和网络存储, 从而免去了与每个服务提供者进行交互的过程.

(2) 无处不在的网络访问. 网络中提供许多可用功能, 可通过各种统一的标准机制从多样化的瘦客户端或者胖客户端平台获取 (如移动电话、笔记本电脑或掌上电脑).

(3) 资源虚拟化共享. 服务提供者将计算资源汇集到资源共享池中. 通过多租户模式共享给多个消费者, 根据消费者的需求对不同的物理资源和虚拟资源进行动态分配或重分配. 资源的所在地具有保密性, 消费者通常不知道资源的确切位置, 也无力控制资源的分配, 但是可以指定较精确的概要位置 (如国家、省或数据中心). 资源类型包括存储、处理、内存、带宽和虚拟机等.

(4) 快速且灵活性. 能够快速而灵活地提供各种功能以实现扩展, 并且可以快

速释放资源来实现收缩. 对消费者来说, 可取用的功能是应有尽有的, 并且可以在任何时间进行任意数量的购买.

(5) 计量付费服务. 云系统利用一种计量功能 (通常是通过一个付费使用的业务模式) 来自动调控和优化资源利用, 根据不同的服务类型按照合适的度量指标进行计量 (如存储、处理、带宽和活跃用户账户). 通过监控、控制和报告资源的使用情况, 提升服务提供者和服务消费者的透明度.

根据云计算服务对象范围的不同, 云计算有四种部署模式[1]: 私有云、社区云、公有云和混合云.

私有云 (Private Cloud)　云计算出现之前, 对于数据密集型或计算密集型任务, 用户需要建立数据中心来提供服务, 以满足其对数据存储、计算、通信能力的要求. 用户需对数据中心进行运维和安全管理, 对服务器上的数据和应用具有所有权和控制权. 云计算出现后, 这种传统的用户/服务提供者模式逐渐发展成私有云模式. 私有云是由一个用户组织 (如政府、军队、企业) 建立运维的云计算平台, 专供组织内部人员使用, 不提供对外服务. 私有云能够体现云计算的部分优势, 例如计算资源的统一管理和动态分配. 但是, 私有云仍要求组织购买基础设施, 建立大型数据中心, 投入人力、物力来维护数据中心的正常运转, 由此可见, 私有云系统提高了组织的 IT 成本, 而且使云的规模受到了限制. 由于私有云的开放性不高, 在几种部署模式中, 私有云的安全威胁相对较少.

社区云 (Community Cloud)　也称为机构云, 云基础设施由多个组织共同提供, 平台由多个组织共同管理. 社区云被一些组织共享, 为一个有共同关注点 (如任务、安全需求、策略或政策准则等) 的社区或大机构提供服务. 显然, 社区云的规模要大于私有云, 多个私有云可通过 VPN 连接到一起组成社区云, 以满足多个私有云组织之间整合和安全共享的需求.

公有云 (Public Cloud)　公有云的基础设施由一个提供云计算服务的大型运营组织建立和运维, 该运营组织一般是拥有大量计算资源的 IT 巨头, 例如 Google、微软、Amazon、百度等大型企业. 这些 IT 公司将云计算服务以 "按需购买" 的方式销售给一般用户或中小企业群体. 用户只需将请求提交给云计算系统, 付费租用所需的资源和服务. 对用户来说, 不需要再投入成本建立数据中心, 不需要进行系统的维护, 可以专心开发核心的应用服务. 目前, Amazon 的 EC2、Google App Engine、Windows Azure、百度云等都属于公有云计算系统. 由于公有云的开放性较高, 而用户又失去了对数据和计算的控制权, 因此, 与私有云相比, 公有云的数据安全威胁更为突出.

混合云 (Hybrid Cloud)　云基础设施由两种或两种以上的云 (私有云、社区云或公有云) 组成, 每种云仍然保持独立, 但用标准的或专用的技术将它们组合起来, 具有数据和应用程序的可移植性, 例如混合云可以在云之间通过负载均衡技术应付

此为试读,需要完整PDF请访问: www.ertongbook.com

突发负载. 由于混合云可以是私有云和公有云的组合, 某些用户选择将敏感数据和计算外包到私有云, 而将非敏感数据和计算外包到公有云中, 但在这种使用模式下, 服务在不同云之间的安全无缝连接较难实现.

最常使用的私有云和公有云之间的关系如图 1-1 所示.

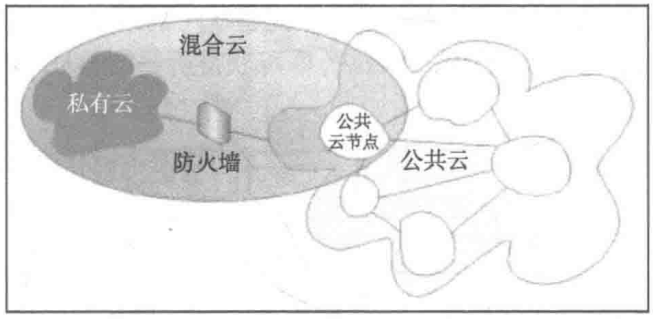

图 1-1 云计算的几种部署模式[2]

计算就要有计算环境, 一般计算环境都有硬件的一层、资源组合调度的一层 (即操作系统层), 以及计算任务的应用业务的软件层. 云计算与一般计算环境的三个层面类似, 云计算提供的三种服务模式就对应了计算环境的三个层面. 这三种服务模式[1]分别是基础设施即服务 IaaS(Infrastructure as a Service)、平台即服务 PaaS(Platform as a Service) 以及软件即服务 SaaS(Software as a Service).

IaaS 将计算、存储、通信资源封装为服务提供给云用户, 云用户相当于使用裸机, 就能够部署和运行任意软件. IaaS 提供计算资源最常用的方式是虚拟机 (Virtual Machine, VM), 典型服务有 Amazon 的 EC2 等. IaaS 提供存储资源的服务能够为用户提供海量数据存储和访问服务, 这种存储服务也被单独称为 DaaS(Data as a Service). 提供存储资源的典型服务有 Amazon 的 S3、Google 的 GFS 等. IaaS 可以提供高速网络和通信服务, 这种服务也被称为 CaaS(Communication as a Service), 提供网络和通信资源的典型服务有 OpenFlow.

PaaS 是在基础设施与应用之间的重要一层, PaaS 将基础设施资源进行整合, 为用户提供基于互联网的应用开发环境, 包括应用编程接口和运行平台等, 方便了应用与基础设施之间的交互. 典型的 PaaS 平台有 Google 的 MapReduce 框架, 应用执行环境 Google App Engine、微软公司的 Microsoft Azure Services.

SaaS 即云应用软件, 为用户提供直接为其所用的软件. SaaS 一般面向终端用户, 特别是 "瘦终端". 终端用户利用 Web 浏览器, 通过网络就可以获得所需的或定制的云应用服务. 终端用户不具有网络、操作系统、存储等底层云基础设施的控制权, 也不能控制应用的执行过程, 只有非常有限的与应用相关的配置能力. SaaS 使用户以最小的开发和管理开销获得定制的应用. 典型的 SaaS 服务有 Salesforce 公司的 CRM 系统、Google Docs 等.

云安全联盟 CSA[4]给出了云计算平台的体系结构, 涵盖了上述三种服务模式, 如图 1-2 所示. 由云计算的服务模式可知, IaaS 为云用户提供基础设施服务. PaaS 基于底层的基础设施资源, 为用户提供定义好 API 的编程模型和应用程序运行环境. SaaS 基于下层的基础设施或者编程模型和运行环境来开发, 为用户提供云应用软件. 从云用户的角度看, 云计算系统是一个完整统一的系统, 用户只需将服务请求提交到云计算系统的 "入口"(Web Portal), 即可获得所需的 IaaS、PaaS 和 (或)SaaS 服务.

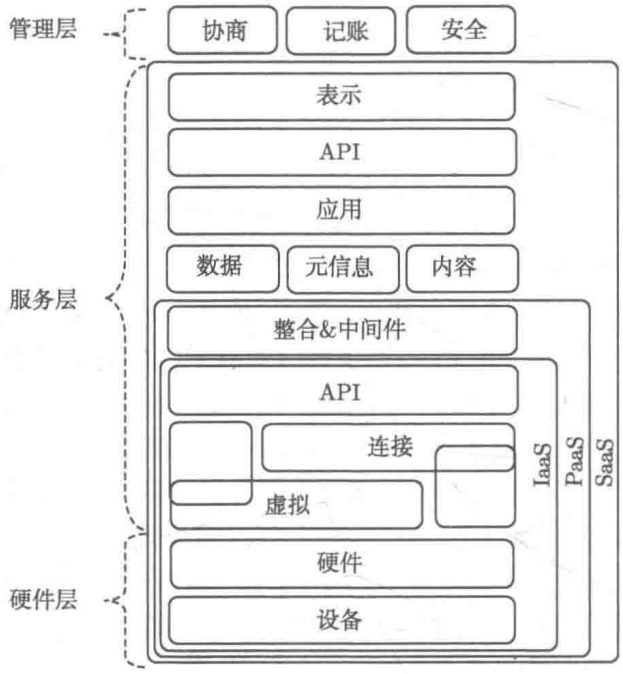

图 1-2　云计算平台的体系结构[3]

1.2　云计算安全体系

信息安全领域的发展历程已多次证明, 信息技术的重大变革将直接影响信息安全领域的发展进程. 云计算以动态的服务计算为主要技术特征, 以灵活的 "服务合约" 为核心商业特征, 这种变革为信息安全领域带来了巨大的冲击.

简单来说, 如图 1-3 所示, 云环境中安全风险来自于多个实体, 包括云服务提供商 (Cloub Service Provider, CSP)、云用户 (Data Owner, OWN)、第三方审计者 (Third Party Auditor, TPA)、外部攻击者等. 这些实体都可能给云环境带来安全风险, 如云服务商的非授权访问、云用户的租户间隔离失效、第三方审计者的泄露隐私、外部攻击者的入侵云端系统等.

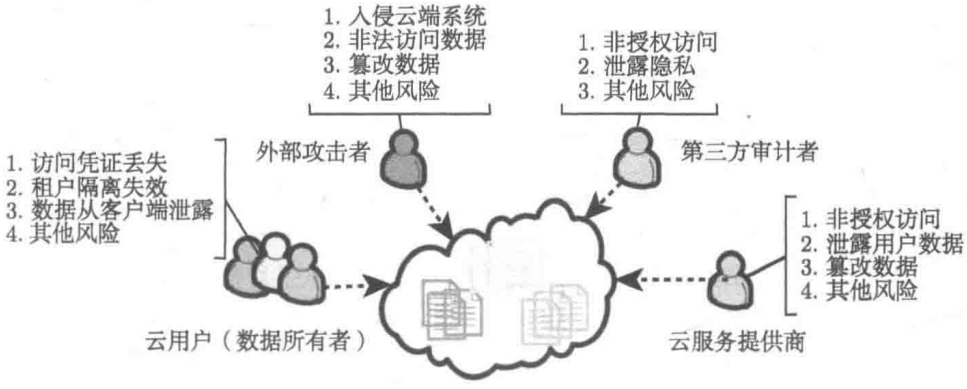

图 1-3 云环境中的访问实体及安全风险[5]

云安全联盟与惠普公司, 基于对 29 家企业、技术供应商和咨询公司的调查结果, 共同列出了云计算存在的安全问题[2], 如下所述.

(1) 数据丢失/泄露: 云计算中对数据的安全控制力度并不理想, API 访问权限控制以及密钥生成、存储和管理方面的不足都可能造成数据泄露, 并且还可能缺乏必要的数据销毁政策.

(2) 共享技术漏洞: 在云计算中, 简单的错误配置都可能造成严重影响, 由于云计算环境中很多虚拟服务器共享相同的配置, 这就需要为网络和服务器配置执行服务水平协议 (SLA) , 以确保及时安装修复程序以及实施最佳做法.

(3) 供应商可靠性不易评估: 云计算服务供应商对工作人员的背景调查力度可能与企业数据访问权限的控制力度有所不同, 很多供应商在这方面做得还不错, 但并不够, 企业需要对供应商进行评估并提出如何筛选员工的方案.

(4) 身份认证机制薄弱: 很多数据、应用程序和资源都集中在云计算环境中, 而如果云计算的身份验证机制很薄弱的话, 入侵者就可以轻松获取用户账号并登录客户的虚拟机.

(5) 不安全的应用程序接口: 在开发应用程序方面, 企业必须将云计算看作新的平台, 而不是外包, 在应用程序的生命周期中, 必须部署严格的审核过程, 开发者可以运用某些准则来处理身份验证、访问权限控制和加密.

(6) 没有正确运用云计算: 在运用技术方面, 黑客可能比技术人员进步更快, 黑客通常能够迅速部署新的攻击技术, 而在云计算中自由穿行.

(7) 未知的风险: 透明度问题一直困扰着云服务供应商, 账户用户仅使用前端界面, 他们不知道他们的供应商使用的是哪种平台或者修复水平.

当然, 随着云计算的发展, 新的风险和安全问题还将不断升级和出现. 也正因为如此, 沈昌祥院士[2]呼吁: 云计算安全应从技术防护、运营管理、法规保障 3 个方面解决问题. 同时, 从技术防护层面, 提出一个可信云计算体系架构, 即在安全

管理中心支撑下的可信计算环境、可信接入边界和可信网络通信三重防御架构, 如图 1-4 所示.

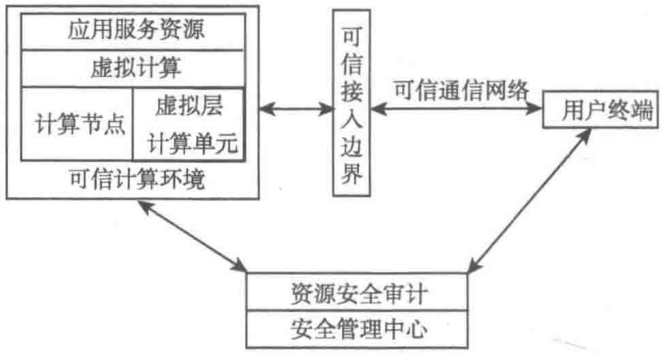

图 1-4　可信云计算体系架构[2]

冯登国研究员[6]也提出, 实现云计算安全至少应解决关键技术、标准与法规建设以及国家监督管理制度等多个层次的挑战. 建立以数据安全和隐私保护为主要目标的云安全技术框架, 建立以安全目标验证、安全服务等级测评为核心的云计算安全标准及其测评体系, 建立可控的云计算安全监管体系; 同时, 提出了一个参考性的云安全框架 (图 1-5) 建议, 包括云计算安全服务体系与云计算安全标准及其测评体系两大部分, 以便从技术层面支撑实现云用户安全目标: 数据安全、隐私保护、安全管理和用户自定义的安全对象.

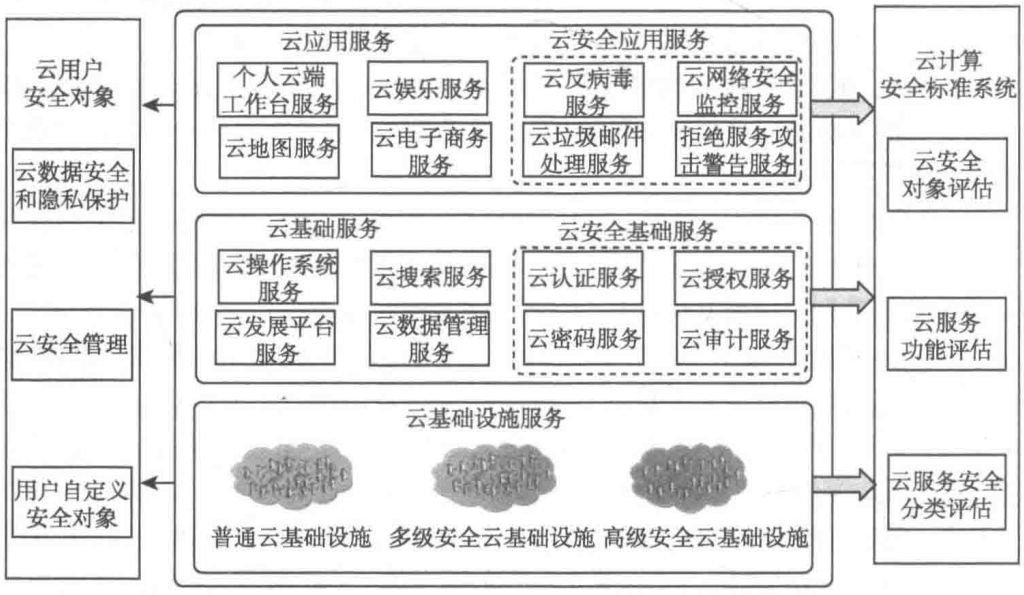

图 1-5　云计算安全技术框架[6]

1.3 云计算数据安全生命周期

诚如上述, 冯登国研究员[6]指出云用户的首要安全目标是数据安全与隐私保护服务. 特别地, 随着云计算的发展, 越来越多的用户开始将数据、应用迁移到云环境中. 当数据迁移到云环境后, 其安全保障依赖于云端系统. 然而, 云用户 (数据拥有者) 没有云端系统的管控权, 因此也就失去了对数据的管控权, 对于云端发起的各种内部攻击也无法阻止. 在这种情况下, 如何确保云环境中的数据安全面临新的挑战.

在云计算环境中, 用户数据的存在形式主要是静态数据和动态数据两种[5]. 静态数据以海量存储和便捷访问为主要目的, 例如长期存储的文档、图片、视频等. 这些数据在存储过程中不需要参与运算, 用户仅仅利用云的存储服务. 云端数据的机密性、完整性、可靠性以及隐私保护等是用户对存储数据关注的核心安全问题, 也是云存储安全技术的研究重点. 目前, 对静态数据的保护主要基于密码学技术, 例如基于密码学实现对数据的访问控制, 用密文检索技术来检索加密数据, 利用高效的完整性检测算法证明数据的存在性与可用性, 利用模糊关键词或匿名搜索来保护用户隐私, 用 VPN、SSH 等机制保证数据传输安全等. 动态数据是参与计算的数据, 例如数据库文件、程序文件、配置文件等. 如果用户既使用了存储服务, 又使用了计算服务, 则动态数据可以从云中存储服务器上直接调用. 利用加密技术保护动态数据十分困难, 目前对密文数据直接操作的 “全同态加密” 执行效率距离实际可用还相差很远. 因此, 数据在运算时依然要解密驻留在内存中, 很难利用密码学技术进行完整的保护. 对动态数据的保护多基于安全策略模型和机制进行, 例如访问控制模型和机制、沙箱机制等.

ORACLE 公司[7]把数据的生命周期用数据的访问频度来描述, 认为随着时间推移, 数据的访问频度也逐渐减小, 最终数据变为归档态 (整个的生命周期经历 3 个状态: 活跃态、次活跃态、历史态或归档态). 云安全联盟[4]提出了数据安全生命周期概念, 其中将数据安全的生命周期归纳为 6 个过程, 包括产生、存储、使用与共享、销毁及归档.

因此, 云计算数据的生命周期就是数据从其产生到销毁的过程. 数据在其生命周期可以用数据所经历的各环节来描述. 原始数据从产生后, 可能被修改、被移动、被归档, 最后被销毁. 数据生命周期模型如图 1-6 所示.

图 1-6 中数据的各处理环节有如下解释.

产生: 数据从无到有的创建过程, 数据可能在云客户端产生, 也可能在云服务端产生.

存储: 数据被存储在云环境中. 为了可靠性, 数据可能被存储在多个节点.

使用与共享：数据从云环境中被提取使用，数据可以被多人共同使用.

销毁：数据彻底不再有价值或不再继续被使用，为保证信息不泄露，使用一定手段使数据无效且无法恢复.

归档：数据因为访问度降低被暂时移除主存储区.

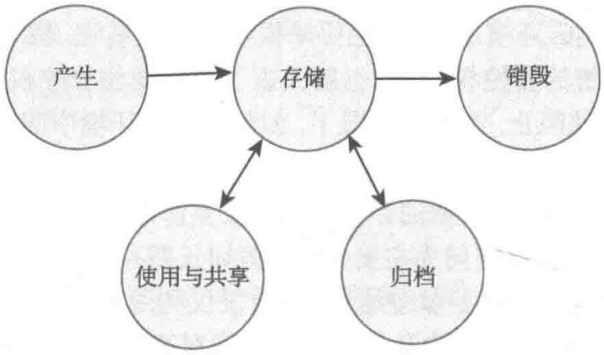

图 1-6　数据生命周期模型

图中圆形表示数据所处的处理环节. 单向或双向箭头表示数据在不同环节的流动

1.3.1　数据安全威胁

云计算环境下，数据面临多方面的威胁，从来源看，可以分为内部威胁和外部威胁.

内部威胁主要来自云服务提供商和云服务用户. 云服务提供商具有控制云数据资源的所有权，对其资源上存储的数据具有完全操作权限. 为了维护系统性能，可以将数据进行任意移动或备份，而数据拥有者却感知不到. 云服务提供商还具有较高的系统权限，可以操作具体系统上的用户资源，修改相应的用户配置信息，获取运行在内存中的明文信息等. 云服务用户拥有合法身份，在多用户共享的云计算平台上，不同用户之间也只是逻辑上的隔离，不同用户的数据也可能存储在同一物理设备上. 在多虚拟机的平台上，通过虚拟机漏洞，可以探测出目标用户的数据操作及数据.

外部威胁来自云环境中的各种恶意攻击者. 这些攻击者可以在云客户端，窃取用户的账号密码，伪装为合法用户，或者在数据的传输过程中进行中间人攻击，窃取用户网络传输中的数据，也可能修改这些数据，甚至制造一些无效数据，或者攻击云端系统，利用云服务系统漏洞入侵系统，直接威胁用户的数据安全.

基于云计算的数据存储服务是一种在线服务. 云计算环境下，数据不只是存储在云中，而是在云服务者和云用户之间流动. 另外，云计算是一种多人共享服务，不同用户数据可能存储在同一物理设备上. 这些云计算特点告诉我们，单纯从某一环节去保护数据安全是不够的，必须统筹考虑数据在数据生命周期不同阶段的安全威

此为试读，需要完整PDF请访问：www.ertongbook.com