网络空间安全专业规划教材

总主编 ◎ 杨义先　　执行主编 ◎ 李小勇

云计算数据安全

Data Security in Cloud Computing

黄勤龙　杨义先　编著

北京邮电大学出版社
www.buptpress.com

网络空间安全专业规划教材

总主编　杨义先　　执行主编　李小勇

云计算数据安全

黄勤龙　杨义先　编著

北京邮电大学出版社
www.buptpress.com

内 容 简 介

针对迅速发展的存储云、移动云、社交云、健康云、物联云和车联云等典型云计算平台中的数据安全问题，本书首先介绍了云计算环境的安全问题和数据安全需求，然后重点介绍了云计算平台中数据安全的前沿技术，包括基于属性密码的数据加密存储和高效访问控制技术、基于代理重加密的数据安全共享技术、基于同态加密的加密数据分类技术、基于可搜索加密的密文搜索技术、基于洋葱模型的云数据库加密和访问控制技术，最后介绍了云计算中数据完整性、密文去重和确定性删除等技术，有助于读者了解各种云计算复杂应用场景下的数据安全技术。本书可作为高校网络空间安全相关专业本科生和研究生的教材，也可作为云计算数据安全研究人员的参考资料。

图书在版编目(CIP)数据

云计算数据安全 / 黄勤龙，杨义先编著. -- 北京 ：北京邮电大学出版社，2018.1
ISBN 978-7-5635-4384-7

Ⅰ. ①云… Ⅱ. ①黄… ②杨… Ⅲ. ①云计算—网络安全 Ⅳ. ①TP393.08

中国版本图书馆 CIP 数据核字(2017)第 227395 号

书　　名：云计算数据安全
著作责任者：黄勤龙　杨义先　编著
责 任 编 辑：毋燕燕　孙宏颖
出 版 发 行：北京邮电大学出版社
社　　址：北京市海淀区西土城路 10 号(邮编：100876)
发 行 部：电话：010-62282185　传真：010-62283578
E-mail：publish@bupt.edu.cn
经　　销：各地新华书店
印　　刷：北京鑫丰华彩印有限公司
开　　本：787 mm×1 092 mm　1/16
印　　张：11
字　　数：267 千字
版　　次：2018 年 1 月第 1 版　2018 年 1 月第 1 次印刷

ISBN 978-7-5635-4384-7　　定价：28.00 元

Prologue 序

Prologue

作为最新的国家一级学科，由于其罕见的特殊性，网络空间安全真可谓是典型的“在游泳中学游泳”。一方面，蜂拥而至的现实人才需求和紧迫的技术挑战，促使我们必须以超常规手段，来启动并建设好该一级学科；另一方面，由于缺乏国内外可资借鉴的经验，也没有足够的时间纠结于众多细节，所以，作为当初“教育部网络空间安全一级学科研究论证工作组”的八位专家之一，我有义务借此机会，向大家介绍一下2014年规划该学科的相关情况；并结合现状，坦诚一些不足，以及改进和完善计划，以使大家有一个宏观了解。

我们所指的网络空间，也就是媒体常说的赛博空间，意指通过全球互联网和计算系统进行通信、控制和信息共享的动态虚拟空间。它已成为继陆、海、空、太空之后的第五空间。网络空间里不仅包括通过网络互联而成的各种计算系统（各种智能终端）、连接端系统的网络、连接网络的互联网和受控系统，也包括其中的硬件、软件乃至产生、处理、传输、存储的各种数据或信息。与其他四个空间不同，网络空间没有明确的、固定的边界，也没有集中的控制权威。

网络空间安全，研究网络空间中的安全威胁和防护问题，即在有敌手对抗的环境下，研究信息在产生、传输、存储、处理的各个环节中所面临的威胁和防御措施，以及网络和系统本身的威胁和防护机制。网络空间安全不仅包括传统信息安全所涉及的信息保密性、完整性和可用性，同时还包括构成网络空间基础设施的安全和可信。

网络空间安全一级学科，下设五个研究方向：网络空间安全基础、密码学及应用、系统安全、网络安全、应用安全。

方向1，网络空间安全基础，为其他方向的研究提供理论、架构和方法学指导；它主要研究网络空间安全数学理论、网络空间安全体系结构、网络空间安全数据分析、网络空间博弈理论、网络空间安全治理与策略、网络空间安全标准与评测等内容。

方向2，密码学及应用，为后三个方向（系统安全、网络安全和应用安全）提供密码机制；它主要研究对称密码设计与分析、公钥密码设计与分析、安全协议

设计与分析、侧信道分析与防护、量子密码与新型密码等内容。

方向3，系统安全，保证网络空间中单元计算系统的安全；它主要研究芯片安全、系统软件安全、可信计算、虚拟化计算平台安全、恶意代码分析与防护、系统硬件和物理环境安全等内容。

方向4，网络安全，保证连接计算机的中间网络自身的安全以及在网络上所传输的信息的安全；它主要研究通信基础设施及物理环境安全、互联网基础设施安全、网络安全管理、网络安全防护与主动防御（攻防与对抗）、端到端的安全通信等内容。

方向5，应用安全，保证网络空间中大型应用系统的安全，也是安全机制在互联网应用或服务领域中的综合应用；它主要研究关键应用系统安全、社会网络安全（包括内容安全）、隐私保护、工控系统与物联网安全、先进计算安全等内容。

从基础知识体系角度看，网络空间安全一级学科主要由五个模块组成：网络空间安全基础、密码学基础、系统安全技术、网络安全技术和应用安全技术。

模块1，网络空间安全基础知识模块，包括：数论、信息论、计算复杂性、操作系统、数据库、计算机组成、计算机网络、程序设计语言、网络空间安全导论、网络空间安全法律法规、网络空间安全管理基础。

模块2，密码学基础理论知识模块，包括：对称密码、公钥密码、量子密码、密码分析技术、安全协议。

模块3，系统安全理论与技术知识模块，包括：芯片安全、物理安全、可靠性技术、访问控制技术、操作系统安全、数据库安全、代码安全与软件漏洞挖掘、恶意代码分析与防御。

模块4，网络安全理论与技术知识模块，包括：通信网络安全、无线通信安全、IPv6安全、防火墙技术、入侵检测与防御、VPN、网络安全协议、网络漏洞检测与防护、网络攻击与防护。

模块5，应用安全理论与技术知识模块，包括：Web安全、数据存储与恢复、垃圾信息识别与过滤、舆情分析及预警、计算机数字取证、信息隐藏、电子政务安全、电子商务安全、云计算安全、物联网安全、大数据安全、隐私保护技术、数字版权保护技术。

其实，从纯学术角度看，网络空间安全一级学科的支撑专业，至少应该平等地包含信息安全专业、信息对抗专业、保密管理专业、网络空间安全专业、网络安全与执法专业等本科专业。但是，由于管理渠道等诸多原因，我们当初只重点考虑了信息安全专业，所以，就留下了一些遗憾，甚至空白，比如，信息安全心

理学、安全控制论、安全系统论等。不过幸好，学界现在已经开始着手，填补这些空白。

北京邮电大学在网络空间安全相关学科和专业等方面，在全国高校中一直处于领先水平；从20世纪80年代初至今，已有30余年的全方位积累，而且，一直就特别重视教学规范、课程建设、教材出版、实验培训等基本功。本套系列教材，主要是由北京邮电大学的骨干教师们，结合自身特长和教学科研方面的成果，撰写而成。本系列教材暂由《信息安全数学基础》《网络安全》《汇编语言与逆向工程》《软件安全》《网络空间安全导论》《可信计算理论与技术》《网络空间安全治理》《大数据服务与安全隐私技术》《数字内容安全》《量子计算与后量子密码》《移动终端安全》《漏洞分析技术实验教程》《网络安全实验》《网络空间安全基础》《信息安全管理(第3版)》《网络安全法学》《信息隐藏与数字水印》等20余本本科生教材组成。这些教材主要涵盖信息安全专业和网络空间安全专业，今后，一旦时机成熟，我们将组织国内外更多的专家，针对信息对抗专业、保密管理专业、网络安全与执法专业等，出版更多、更好的教材，为网络空间安全一级学科，提供更有力的支撑。

杨义先

教授、长江学者、杰青

北京邮电大学信息安全中心主任

灾备技术国家工程实验室主任

公共大数据国家重点实验室主任

2017年4月，于花溪

前言

Foreword

云计算是近年来迅速发展的一种新型计算模式，它以服务的形式为用户提供丰富的计算和存储等资源，通过将计算任务分布在由大量计算机构成的资源池上，使各种应用系统能够根据需要获取计算能力、存储空间和各种软件服务。这种全新的应用模式，成为解决高速数据处理、海量信息存储、资源动态扩展、数据安全与实时共享等问题的有效途径，向人们展示了其强大而又独具特色的发展优势。因此，自2006年云计算概念诞生以来，得到了人们的高度关注，各种新概念、新观点、新技术和新产品层出不穷。

云存储正是在云计算概念上延伸和发展出来的一个新概念，其通过集群、分布式等技术将海量异构存储设备通过网络和应用软件等结合起来协同工作，并以按需访问的形式通过网络对外提供大规模的数据存储和访问服务。云存储一方面使用集群技术解决了传统技术的性能瓶颈问题，支持性能和容量的动态线性扩展，适用于海量数据的存储；另一方面为用户提供按需计费，缩减了用户对存储资源的投入，降低了管理成本。

云计算与云存储的透明性分离了数据与基础设施的关系，对用户屏蔽了底层的具体实现细节，但其服务模式也带来了安全隐患：云计算的服务模式允许用户将数据上传到云存储平台并共享给他人，同样也使得半可信的云平台能够访问到用户的数据，甚至能够在不经用户允许的情况下篡改用户的数据。此外，未授权的用户也有可能假冒合法用户访问云存储平台中的数据。传统的数据加密方案虽然可以保护云存储平台中数据的机密性，但却要求用户通过复杂的计算来解密数据，而且缺乏对密文修改权限的控制。另外，云计算平台中数据的加密也带来了分类、搜索、透明使用（如云数据库）等难题，数据的持有性和可恢复证明、密文去重和确定性删除等也是云计算数据使用过程中用户关心的安全问题。

本书针对云计算中数据的安全问题，介绍了数据的加密存储、访问控制、安全共享、密文分类、密文搜索和完整性验证等技术，防止云平台恶意泄露和修改用户的隐私数据，保护数据在上传、存储、共享等过程中的安全性，满足数据分类和搜索、云数据库等复杂应用场景。本书共分为9章，各章的具体安排如下。

第1章介绍了云计算的基本概念、基础架构，以及私有云、公有云和混合云3种部署模式。针对云计算的应用现状，介绍了存储云、移动云、社交云、健康云、物联云和车联云等典型的应用场景。

第 2 章介绍了云存储的基本概念和体系架构，以及用户和云存储提供商面临的主要安全问题。通过引入云安全的基本需求，重点介绍了云存储中的数据安全需求，包括数据机密性、访问控制、授权修改和可用性等。

第 3 章介绍了安全的基本理论，包括双线性对、困难问题、身份加密和广播加密算法，以及树形访问结构、秘密共享访问结构等。

第 4 章介绍了云计算数据访问控制的含义、属性加密的基本概念，以及基于属性加密的访问控制方案。接着介绍了基于属性加密的安全外包和基于属性加密的改进方案，后者包括层次化属性加密方案、支持策略更新的属性加密方案。最后介绍了结合属性签名、属性广播加密的访问控制方案。

第 5 章介绍了云计算数据安全共享的含义、代理重加密的基本概念，以及属性代理重加密的方案。重点介绍了条件代理重加密方案，包括基于关键字、基于访问策略和时间控制的条件代理重加密方案，以及代理重加密的综合应用方案。

第 6 章介绍了云计算加密数据分类的含义，以及典型的数据分类算法，包括朴素贝叶斯分类、K 最近邻分类和支持向量机分类算法。重点介绍了基于同态加密的隐私数据分类方法，包括基于朴素贝叶斯分类、K 最近邻分类和支持向量机的隐私数据分类。

第 7 章介绍了云计算加密数据搜索的含义，包括对称可搜索加密和公钥可搜索加密两类。分别介绍了基于线性扫描算法、倒排索引算法、布隆过滤器、模糊关键词检索以及可验证对称可搜索加密算法，单关键字、多关键字和连接关键词的公钥可搜索加密算法。

第 8 章介绍了云计算中数据库安全的含义，以及云数据库加密的分析。基于洋葱加密模型，重点介绍了基于同态加密、保序加密的云数据库透明加密方案，以及基于属性加密的云数据库密文访问控制方案。

第 9 章介绍了云计算的其他数据安全技术，包括数据持有性证明、数据可恢复证明、数据密文去重、数据确定性删除等，以及云计算数据安全的未来发展。

作者对参与本书编写的人员一并表示感谢，最后由北京邮电大学云计算与智能安全实验室(http://www.buptcsc.com)统稿和校对。本书的编写得到了国家自然科学基金面上项目“移动云存储中面向多用户共享的数据安全技术研究”(批准号:61572080)、国家重点研发计划网络空间安全专项“网络空间数字虚拟资产保护基础科学问题研究”(批准号:2016YFB0800605)、CCF-启明星辰鸿雁科研资助计划“移动云存储中数据访问控制关键技术研究”(批准号:2016012)的资助，特此表示感谢。

由于作者水平有限，书中不妥之处在所难免，恳请读者提出宝贵意见。

作　者

目录

Contents

第 1 章

云计算概述

1.1 云计算的概念

1.1.1 云计算定义

随着高速网络和移动网络的衍生，高性能存储、分布式计算、虚拟化等技术的发展，云计算服务正日益演变为新型的信息基础设施，并得到各国政府的高度重视。近年来，各国纷纷制定云计算国家战略和行动计划[1]，云计算在我国也得到了快速发展。2009 年以来，我国云计算开始进入实质性发展阶段，整个“十二五”期间，我国云计算一直保持超过 30%的年均增长力，成为全球增速最快的市场之一，云计算也成为国家“十三五”重点发展项目和战略性新兴产业。

提出云计算概念前，网格计算已有十多年的研究历史[2]，受到广泛关注。网格计算是一种分布式计算模式，将分散在网络中的空闲服务器、存储系统连接在一起，形成一个整合系统，为用户提供功能强大的计算及存储能力来处理特定的任务。对于使用网格的最终用户或应用程序来说，网格就像是一个拥有超强性能的虚拟计算机，其本质在于以高效的方式来管理各种加入了该分布式系统的异构松耦合资源，并通过任务调度来协调这些资源，合作完成一项特定的计算任务。云计算与网格计算的差别在于，网格计算由多台计算机构成网格，服务于一个特定的大型计算；云计算依托网络在互联网上由一个个集约化、专业化的云计算平台形成规模化的服务[3,4]。

云计算(Cloud Computing)的概念是由谷歌前 CEO 施密特在 2006 年 8 月举行的搜索引擎大会上最先提起的。此概念一经提出，即带来极具产业远景的计算模型架构的广泛探讨和热烈追捧，各国政府也纷纷投入了相当大的财力和物力用于云计算的部署，交通运输、电力、电信、石油石化等行业也启动了相应的云计算发展计划，以促进产业信息化。“云计算”目前仍是一个不断发展的词汇，不同领域的专家、学者对云计算研究的出发点各异，云计算的定义也不尽相同[5]。比较典型的定义如下。Salesforce 认为云计算是一种更友好的业务运行模式。在这种模式中，用户的应用程序运行在共享的数据中心，用户只需要通过登录和个性化定制就可以使用这些数据中心的应用程序，从而免除了软件购买、部署和维护的困扰和费用，降低了企业管理成本。IBM 认为云计算是一种共享的网络交付信息服务的模式，云服务的使用者看到的只有服务本身，而不用关心相关基础设施的具体实现[6]。云计

算是一种革新的IT运用模式，这种运用模式的主体是所有连接着互联网的实体，可以是人、设备和程序。这种运用方式的客体就是IT本身，包括我们现在接触到的，以及会在不远的将来出现的各种信息服务。而这种运用方式的核心原则是：硬件和软件都是资源并被封装为服务，用户可以通过互联网按需进行访问和使用。迄今为止，美国国家标准与技术研究院(NIST)对云计算给出的定义[7]，是目前接受度较高的定义，其具体描述是："云计算是一种模式。计算资源(包括网络、服务器、存储、应用软件及服务等)存储在可配置的资源共享池中，云计算通过便利的、可用的、按需的网络访问计算资源。计算资源能够被快速提供并发布，最大化地减少管理资源的工作量或与服务提供商的交互。"

1.1.2 云计算特征

云计算是分布式计算、网格计算、并行计算、效用计算、虚拟化、网络存储、负载均衡等传统计算机和网络技术发展融合的产物，是一种利用大规模低成本运算单元通过网络连接，以提供各种计算和存储服务的技术，也是需求推动、技术进步和商业模式转变共同促进的结果。云计算是一种基于因特网的超级计算模式，在远程的数据中心，几万甚至几千万台计算机和服务器连接成一片。因此，云计算甚至可以让用户体验每秒超过万亿次的运算能力，如此强大的运算能力几乎无所不能。用户通过台式计算机、笔记本式计算机、手机等方式接入数据中心，按各自的需求进行存储和运算。同时，云计算还是一个可以动态伸缩的弹性模型，这样可以根据应用和用户数量的不同，分配相当的计算资源。云计算平台里的硬件设施可以随时更新，这样可以保证云平台的可持续发展性。用户可以从各种终端设备随时随地获取相应的云服务，用户所得到的资源服务全部来自云平台，但是用户不知道这些服务具体运行在哪个位置，只要有一台计算机或一部手机，就可以通过互联网来得到我们想要的服务，甚至是超级计算这样的服务。由于云计算具有高容错性，这样就保障了服务的高可靠性，甚至比我们使用自己的计算机还可靠。云计算不针对特定的应用，在云平台的支撑下可以构造出千变万化的应用，同一个云可以同时支撑不同的应用运行。云平台是一个庞大的资源池，用户可以按需付费获取服务。云的特殊容错措施就决定了云可以用廉价的节点来组成，用户不用负担云平台的维护管理费用，就可以享受低成本的服务。

云计算的基本特征主要体现在以下6个方面[8]。

① 虚拟化。云计算将传统的计算、网络和存储资源通过提供虚拟化、容错和并行处理的软件，转化成可以弹性伸缩的服务。

② 弹性伸缩。云计算运用网络整合众多的计算机资源，构成技术存储模式，实现多种功能，包括并行计算、网格计算、分布式计算、分布式存储等。云具有漫无边际的属性，云计算则在构建基础设施的设备、信息基地、信息服务范围和信息用户方面具有超大规模的特点。云计算能够无缝地扩展到大规模的集群之上，甚至包含数千个节点同时处理。在用户看来，云的规模可以实现动态伸缩，满足不同用户不同时期的服务需要。

③ 提高工作效率。与原有的工作站单独计算的模式相比，云计算模式能在很短的时间内完成，实现效率的提升。

④ 资源使用计量。云计算的服务是可计量的，付费标准是根据用户的用量收费。在存储和网络宽带技术中，已广泛使用了这种即付即用的方式。

⑤ 按需自助服务。用户使用云计算平台上的服务就像使用生活的自来水、电和天然气一样，不受时空限制。享受云平台服务时，不受访问平台和系统的制约，只需拥有 Internet 和通过访问验证即可。

⑥ 经济性。在达到同样性能的前提下，组建一个超级计算机所消耗的资金很多，而云计算通过采用大量商业机组成集群的方式，所需要的费用与之相比要少很多。

1.2　云计算的基础架构

云计算其实是分层的，这种分层的概念也可视为其不同的服务模式。根据 NIST 的权威定义，云的服务模式包含基础设施即服务（Infrastructure as a Service，IaaS）、平台即服务（Platform as a Service，PaaS）和软件即服务（Software as a Service，SaaS）3 个层次[9]。基础设施即服务在最下端，平台即服务在中间，软件即服务在顶端，如图 1-1 所示。

图 1-1　云计算基础架构

1.2.1　基础设施即服务

基础设施即服务在服务层次上是最底层服务，接近物理硬件资源，首先将处理、计算、存储和通信等具有基础性特点的计算资源进行封装后，再以服务的方式面向互联网用户提供处理、存储、网络以及其他资源方面的服务，以便用户能够部署操作系统和运行软件。这样用户就可以自由部署、运行各类软件（包括操作系统），完成用户个性需求。底层的云基础设施此时独立在用户管理和控制之外，通过虚拟化的相关技术实现，用户可以控制操作系统，进行应用部署、数据存储，以及对个别网络组件（如主机、防火墙）进行有限的控制。

这一层典型的服务如亚马逊的弹性计算云（Elastic Compute Cloud，EC2）和 Apache 的开源项目 Hadoop。EC2 与 Google 提供的云计算服务不同。Google 只为在互联网上的应用提供云计算平台，开发人员无法在这个平台上工作，因此只能转而通过开源的 Hadoop 软件的支持来开发云计算应用。而 EC2 给用户提供一个虚拟的环境，使得可以基于虚拟的操作系统环境运行自身的应用程序。同时，用户可以创建镜像（AMI），镜像包括库文件、数据和环境配置，通过弹性计算云的网络界面去操作在云计算平台上运行的各个实例（Instance），同时用户需要为相应的简单存储服务（S3）和网络流量付费。Hadoop 是一个开源

的基于 Java 的分布式存储和计算项目，其本身实现的是分布式文件系统（HDFS）以及计算框架 MapReduce。此外，Hadoop 包含一系列扩展项目，包括了分布式文件数据库 HBase、分布式协同服务 ZooKeeper 等[10]。Hadoop 有一个单独的主节点，主要负责 HDFS 的目录管理（NameNode），以及作业在各个从节点的调度运行（JobTracker）。

1.2.2 平台即服务

平台即服务是构建在 IaaS 之上的服务，把开发环境对外向客户提供。PaaS 为用户提供了基础设施及应用双方的通信控制。具体来讲，用户通过云服务提供的基础开发平台运用适当的编程语言和开发工具，编译运行应用云平台的应用，以及根据自身需求购买所需应用。用户不必控制底层的网络、存储、操作系统等技术问题，底层服务对用户是透明的，这一层服务是软件的开发和运行环境，是一个开发、托管网络应用程序的平台。

典型的 PaaS 有谷歌公司大规模数据处理系统编程框架 MapReduce 和应用程序引擎 Google App Engine、微软提出的 Microsoft Azure 等。基于 Google App Engine，用户将不再需要维护服务器，用户基于 Google 的基础设施上传、运行应用程序软件。目前，Google App Engine 用户使用一定的资源是免费的，如果使用更多的带宽、存储空间等需要另外收取费用。Google App Engine 提供一套 API 使用 Python 或 Java 来方便用户编写可扩展的应用程序，但仅限 Google App Engine 范围的有限程序，现存很多应用程序还不能很方便地运行在 Google App Engine 上。Microsoft Azure 构建在 Microsoft 数据中心内，允许用户开发应用程序，同时提供了一套内置的有限 API，方便开发和部署应用程序。

1.2.3 软件即服务

软件即服务是指提供终端用户能够直接使用的应用软件系统。服务提供商提供应用软件给互联网用户，用户使用客户端界面通过互联网访问服务提供商所提供的某一应用，但用户只能运行具体的某一应用程序，却不能试图控制云基础设施。常见的 SaaS 应用包括 Salesforce 公司的在线客户关系管理系统 CRM 和谷歌公司的 Google Docs、Gmail 等应用。SaaS 是一种软件交付模式，将软件以服务的形式交付给用户，用户不再购买软件，而是租用基于 Web 的软件，并按照对软件的使用情况来付费。SaaS 由应用服务提供发展而来，应用服务提供仅对用户提供定制化的服务，是一对一的，而 SaaS 一般是一对多的。SaaS 可基于 PaaS 构建，也可直接构建在 IaaS 上。SaaS 具有如下特性[11]。

① 互联网特性。SaaS 应用一般通过互联网交互，用户仅需要浏览器或者联网终端设备就可以访问应用。

② 多租户特性。通过多租户模式实现多种使用方式，以满足不同用户的个性化需求。

③ 按需服务特性。支持可配置型和按使用付费。

④ 规模效应特性。一般面向大量用户提供服务，以取得规模效应。

SaaS 的典型代表有 Salesforce、Google Apps 或微软提供的在线办公软件。目前，成熟的服务提供商多采用一对多的软件交付模式，也称为单实例多租赁，即一套软件为多个租户服务。应用中一个客户通常是指一个企业，也被称为租户，一个租户内可以有多个用户。在数据库实现上，对应 3 种设计方式，每个租户独享一个数据库，或多租户共享数据库独立结