经典译丛

Springer

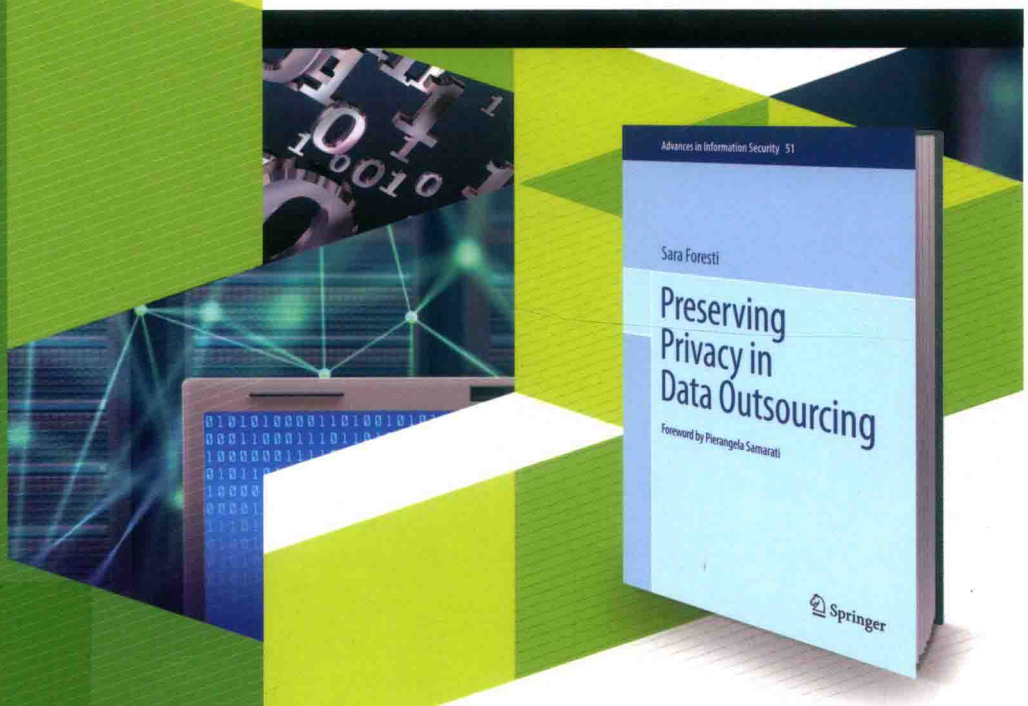
网络空间安全

数据外包中的隐私保护

Preserving Privacy in Data Outsourcing

【意】 Sara Foresti 著

唐春明 姚正安 盛 刚 译



中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

经典译丛·网络空间安全

数据外包中的隐私保护

Preserving Privacy in Data Outsourcing

[意] Sara Foresti 著

唐春明 姚正安 盛 刚 译

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内容简介

本书主要研究云计算环境中数据外包时的隐私性保护，讨论关于安全数据外包的研究现状，总结数据外包情况下的查询计算机制、隐私性保护和数据完整性。本书假设服务器是诚实但好奇的（即半诚实的），设计了一个访问控制系统，为数据外包中的访问控制机制更新提出了有效策略。为了保证解的安全性，该书也考虑了不同服务器的合谋风险。在数据完整性方面，提供了一个约束模型并设计了查询分布式数据的机制。

本书可作为高等院校网络信息相关专业研究生的教材或教学参考书，也可供数据外包和云计算研究领域的相关技术人员作为参考书使用。

版权贸易合同登记号　图字：01-2014-5137

图书在版编目（CIP）数据

数据外包中的隐私保护/（意）萨拉·福雷斯蒂（Sara Foresti）著；唐春明等译. –北京：电子工业出版社，2018.1
（经典译丛·网络空间安全）
书名原文：Preserving Privacy in Data Outsourcing
ISBN　978-7-121-33561-7
Ⅰ. ①数… Ⅱ. ①萨…②唐…Ⅲ. ①计算机网络-安全技术-研究 Ⅳ. ①TP393.08
中国版本图书馆 CIP 数据核字（2018）第 018036 号

策划编辑：马 岚
责任编辑：葛卉婷
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：787×1092　1/16　印张：12.75　字数：245 千字
版　　次：2018 年 1 月第 1 版
印　　次：2018 年 1 月第 1 次印刷
定　　价：49.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。
质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。
本书咨询联系方式：classic-series-info@phei.com.cn。

译者序[1]

在信息技术快速发展的今天，许多企业、机构或个人需要管理大规模的数据，需要付出高昂的人力、物力进行有效管理。云计算技术或大数据平台拥有非常丰富的资源用于数据存储和计算，能够为用户提供高效的数据管理服务，从而将企业、机构或个人从繁重的数据管理任务中解脱出来，专注于其核心业务。

然而，如果将数据存储在云计算或大数据平台的服务器上，企业、机构或个人由于失去了对数据的直接物理控制，引发了许多安全问题，如数据隐私泄露、结果正确性验证等。为了使用户能够安心地使用云计算或大数据平台提供的服务，就需要解决这些问题。

本书为作者的博士论文的扩展，对关系型数据外包中的几个安全问题进行了深入的研究，主要内容包括：

第一，提出了一个方法合并权限和加密，并将数据与访问控制外包。当指定一策略时，数据所有者无须参与该策略的执行。

第二，提出了一个结合分裂和加密的方法来有效地执行数据集合上的隐私性约束，特别关注了查询的执行效率。

第三，提出一个简单但有效的方法来描述权限和实施权限，用于在分布式计算的各数据持有者之间控制数据的泄露，以确保查询执行过程只泄露被明确授权可公开的数据。

本书由唐春明、姚正安和盛刚翻译并审校，参与本书整理工作的还有来自广州大学信息安全技术省重点实验室的博士生任燕、胡杏、陈月乃和硕士生张晓军等。在此，向所有为本书出版提供帮助的人士表示诚挚的谢意！

由于译者水平有限，书中翻译不妥之处，敬请广大读者和专家同行批评指正。

[1]本书符号的正斜体与原书保持一致。 —译者注

序

今天，数据外包作为一个成功案例出现，它允许组织和用户利用外部服务对资源进行分配。事实上，组织机构发现借助于外部服务器来管理 IT 服务和数据，从而专注于他们内部的主要核心业务，这样做更安全、更实惠。同样，用户正越来越多地借助于外部服务来储存和分配用户生成的内容，大量提供这种服务的YouTube，MySpace和Flickr所获得的成功就说明了这个事实。

在这种新的外包和储存/分布情形下，对用户来说，保证数据的适度安全性和隐私性是最重要的。然而，由于提供数据存储和访问服务的服务器不是完全可信的，这导致问题变得特别复杂。外包的数据通常包含敏感的信息，而这些敏感信息的发布应当受到严格的控制，甚至不允许被外部服务器访问。为了应对这个问题，现有的数据外包方案通常假设数据以加密的形式外包，同时附加数据的额外索引信息。这些额外信息允许对加密数据本身进行查询操作，从而避免了在查询计算中需要外部服务器对加密数据的解密。然而，这样的外包仅仅提供了一个基本的保护层，对外包数据中隐私保护的有效性、高效性、灵活性并没有提供一个完备的应对，许多挑战仍然有待解决。

首先，许多情况下对外包数据的访问具有选择性。我们怎么确保用户对外包数据的不同意见？除了数据，我们还能将管理和执行的授权外包给外部服务器吗？如果加密取决于授权，那么当授权改变时，我们如何避免需要重新上传资源的新版本？

其次，当加密和解密在计算上可行时，查询加密数据的代价不可避免地越来越昂贵，并且可能仅适用于有限类型的查询。此外，当敏感的信息不仅是数据本身，还包含数据之间的关联时，加密可能代表着一种过度保护。那么，我们可以离开加密，比如分割数据以打破敏感关联吗？数据如何分割？我们需要对在物理分片以及在服务器上的存储做什么样的假设？如何在碎片数据上执行查询工作？

第三，在某些情况下，可能还需要在存储数据的服务器上执行分布式查询，存储在不同的服务器因此需要在各服务器之间对查询计算进行协作和信息共享。我们怎么建立授权调节服务器之间的信息共享？在查询计算中，我们怎么衡量由派生关系所携带的信息？我们怎么定义和执行一个允许执行协同查询和执行不同的授权的查询计划？

本书所涉及的上述三个方面说明了该领域的研究状况，分析了所要解决的问题。本书探究了不同的方向并对它们的解决方案提出了可行的办法，也对某些开放性问题提出了意见和见解。本书代表着对安全和隐私方面感兴趣的学者和研究人员的宝贵灵感来源，特别是数据外包的情况，为他们所要考虑的不同问题提供一个良好的概述和分析，并给他们提供了解决办法。本书提供了一个完美的问题调查方法和醒目的开放性问题，还是今后研究的灵感源泉。

Pierangela Samarati

前言

随着个人信息大集合的可用性以及支持数据密集型服务的数据存储设备的日益普及，认为服务提供商将会被越来越多地要求对存储、高效以及信息的可靠传播等负责的观点得到了支持，从而实现“数据外包”架构。在数据外包架构中，数据和前端应用程序一起储存在完全负责其管理的外部服务器的网站上。在外部服务器发布数据会提高服务的可用性，减少数据拥有者管理数据的负担，同时数据外包引进了新的隐私和安全担忧，因为存储数据的服务器可能是诚实但好奇的。一个诚实但好奇的服务器会诚实地管理数据，但可能会因读取它们的内容而不被数据拥有者所信任。为了确保充分的隐私保护，一个传统的方法是加密外包的数据，这样可以防止外部攻击以及从服务器本身的侵入。然而，这种传统的解决方法的缺点是降低了查询执行效率和防止选择性信息的发布。于是，这就引进了为外包数据访问控制和隐私限制的定义和执行制定新的模式和方法，同时确保有效的查询执行的必要性。

在本书中，我们提出了一个详尽的方法来保护那些储存于不在数据所有者控制之下的系统中的敏感信息。在设计一个系统以确保被诚实但好奇的服务器存储和管理的数据的保密性时，主要考虑三个安全方面的要求。第一个要求是访问控制的执行，以限制授权用户访问系统资源的能力。在传统意义上，由一个可信的数据管理系统模块负责执行访问控制策略。在此所考虑的情况下，服务提供商在执行访问控制策略上是不可信的，并且数据所有者也不愿去调解访问请求来过滤查询结果。因此，我们提出了一个新的访问控制系统，该系统基于选择性加密，这样该策略的执行不需要系统中存在可信模块。第二个要求是隐私保护，以限制非授权用户对存储/公布的数据的可视性，同时最大限度地减少使用加密方法。数据收集通常包含一些能够识别出个人的信息，它们在存储和传播给其他方时需要保护。例如，医疗数据不能和患者的身份信息一起存储或公开。为了确保隐私保护以及限制加密的使用，在本书中我们第一次提出一个解决办法，即通过保密性限制，

用简单而强有力的方式来模拟隐私要求，隐私被定义为必须限制数据联合可见度的数据集。 然后我们提出了一个执行保密性限制的机制，它基于碎片和加密技术的结合使用： 碎片打破的关联将只对那些被授权、只有自己知道关联的用户可见。第三个要求是安全数据整合，以限制授权用户为了分布式查询评估而交换数据的能力。 事实上，通常需要储存用户个人信息的不同信息源合作以达到一个共同的目标。 然而，这样的数据整合和共享可能受保密性限制，因为不同的参与方会被允许访问数据的不同部分。 因此，我们提出了一个模型来方便表达数据交换的限制以及一个在分布式查询评估过程中执行的机制。

在本书中，为了在外包数据中执行访问控制，我们通过定义一个模型和一个机制， 通过引进一个执行隐私限制的分割和加密方法，以及通过设计来调节不同参与方之间的数据流的技术， 来解决这三个关于安全方面的要求。 主要的贡献可概述如下。

- 关于外包数据的访问控制执行，原创成果有： 联合使用选择性加密与密钥派生策略以应对访问控制的执行； 在密钥派生方面引进一个加密策略的极小性概念来正确执行访问控制策略而不降低它的有效性； 开发了一个在多项式时间内计算最小的加密策略的启发式方法； 引进了一个两层加密模型来管理策略更新。

- 关于执行隐私保护的模型的定义，原创成果有： 定义了保密性限制，这对模拟隐私要求来说是一个简单却完整的方法； 引进最小化碎片的概念，可以捕获一个碎片的性质来满足保密性限制，同时最大限度减少碎片数量； 开发了一个有效计算最小化碎片的方法，这是一个NP困难问题； 引进三个局部最优的概念，该概念基于构成解决办法的碎片的结构，基于碎片中的属性的亲和力以及一个查询评估成本的模型； 提出了三个不同的计算碎片方法以满足关于最优性的三个定义。

- 关于安全数据整合机制的设计，原创成果有： 权限的定义对数据交换限制建模来说是一个简单但完整的方法； 对权限和查询的建模通过基于图的模型的关系分布及其表示； 引进了一个在多项式时间内对工作权限的组成的方法； 定义了一个考虑数据交换限制同时设计一个查询执行计划的方法。

Sara Foresti

致 谢

这本书是我的博士学位论文发表的结果。我想借此机会对那些使我的博士学位论文工作得以实现的人表达我真挚的谢意。 我很抱歉，在这之前，不能用言语来表达我对所有这些人的感激之情。

首先，我要感谢我的指导老师，Pierangela Samarati，她一直在专门指导我。在这五年中，她向我介绍了她所做的科研工作以及对它的喜爱之情，这感染和激励了我。 很感谢她给了我许多机会，感谢她一直以来的支持、指导和建议。所有我所知道的关于安全和隐私都是从她那里学的，没有她，这本书根本写不出来。成为她的博士生是我的荣幸，她给了我和她一起工作的可能，这不仅仅是荣誉，更是我的乐趣所在。

还要特别感谢我的合作导师，Sabrina De Capitani di Vimercati，因为当我需要建议时，无论是技术方面还是其他方面，她总能及时出现，她也一直在听以及回答我的所有（笨的）问题，她的耐心以及帮助使我从不同的角度看待事物。

我尤其还要感谢Sushil Jajodia，他第一个设想把我的博士学位论文作为一本书出版的可能性，而且也是他使得这成为可能。我还要感谢他给了我机会去访问美国弗吉尼亚州的乔治·梅森大学的安全信息系统中心。我非常感谢他的支持以及提供了一个刺激愉快的工作氛围。

本书中已取得的大部分列举的结论要归功于与 Valentina Ciriani，Sabrina De Capitani di Vimercati，Sushil Jajodia，Stefano Paraboschi 和 Pierangela Samarati 等值得我感激的人的合作及许多有意义的讨论。我想向他们表示感谢，不仅是因为他们在我这本书不同部分的学习中给予的支持与帮助（因此在不同章节中分别表示感谢），而且因为给我机会向她们学习以及学习她们的经验。

我还要特别感谢 Vijay Atluri 教授，Carlo Blundo 教授，Sushil Jajodia 教授，Javier Lopez 教授，谢谢他们的宝贵意见，帮助改进本书的工作。

同时还要感谢 Susan Lagerstrom-Fife，感谢她在本书出版前期的指导，同时

还有 Jennifer Maurer。她们在我准备手稿时的支持是我完成工作的基本。

最后，但并非不重要，我还要感谢我的家人。用尽这书的所有纸都不足以表达我对他们的感激：他们的教导、支持以及她们的爱，曾经是而且将来也将一直都是我的一个基本参考点。要特别感谢的是Eros，当我需要的时候他一直都在，谢谢他的支持和耐心。

目录

第1章　引 言

由私人公司和公共组织机构储存、处理以及交换的数据的数量正在急剧增加。结果是，如今的用户诉诸于服务提供商来传播和共享他们希望对他人有用的资源的频率一直在增加。因此，对侵犯个人隐私的保护正成为最重要的问题之一，它必须在这样一个开放、合作的背景下加以解决。在本书中，我们定义了一个全面的当信息存储于不在数据所有者的直接控制之下的系统中时保护敏感信息的方法。在本章的其余部分，我们给出了本书的动机和一些概述。

1.1　目的

存储、处理以及技术交流的快速进化正在改变着由私人公司和公共组织机构所采用的传统的信息系统架构。这个改变是必要的，主要有两个原因：首先，由于正在增长的存储能力和现代化设备的计算能力使得组织机构所保持的信息量增长得非常快；其次，由组织收集的数据包含着敏感信息（如身份识别信息、金融数据、健康诊断信息等），这些信息的保密性必须得到维护。

无论是对攻击系统的外部用户还是恶意的内部人员，存储和管理这些数据集的系统都应当是安全的。然而，为了保证敏感数据的保密性，一个安全系统的设计、实现和管理的成本是非常昂贵的。这是由于在敏感数据的大集合的内部管理和存储费用的增加，因为它需要存储能力和熟练的管理人员。最近，数据外包和传播服务得到可观的增长，并且会成为未来网络的一个共同组件，这已由提供存储和配送服务且越来越成功的网络公司所证明（如MySpace，Flickr和YouTube）。该趋势的主要结果是公司将他们的数据储存在诚实但好奇的外部服务器中，公司依赖它们确保数据的可用性以及对所存储的数据执行基本的安全控制。然而，尽管这些外部系统在使得公布的信息的有用性方面是可信的，但是，在访问内容以及充分执行服务控制策略和隐私保护要求方面，却是不可信的。

很明显，用户以及公司在使用传播服务时将发现一个有趣的机会，对用户隐私保护提供有力的保证，以防那些恶意的用户想攻入系统以及服务提供商本身。事实上，除了众所周知的保密性和隐私泄漏的风险，还包括收集的信息的不当使用威胁着外包数据。服务提供商可以使用大部分由数据所有者收集和组织的数据集，这可能潜在地损害数据所有者的由其产品和服务组成的市场。

当设计一个系统来确保一个诚实但好奇的服务器储存和管理的数据的保密性时，主要有三个安全方面需要考虑，简要概述如下。

- 访问控制执行。传统架构分配一个关键的角色给参考监视器[7]来进行访问控制执行。参考监视器是负责对访问要求验证的系统组件。然而，本书所考虑的情况挑战着传统架构的一个基本原则，即由一个可信的服务器负责定义和执行访问控制策略。这个假设在这里不再成立，因为服务器甚至不知道由数据所有者所定义的访问（以及可能被修改）。因此我们需要重新思考在开放环境中的访问控制的概念，其中诚实但好奇的服务器负责管理数据收集，而在数据保密性上却不被信任。

- 隐私保护。由组织收集和维护的大量数据通常包括敏感的个人识别信息。这个趋势已引起个人和立法机构的关注，这迫使组织在信息存储、处理和与他人共享数据时对敏感信息提供隐私保护。事实上，最近的法规[22,78]明确地要求敏感信息应当加密或者与其他个人识别信息分开以确保其保密性。加密使得对存储数据的访问变得低效，因为它不可能直接在加密数据上执行查询，因此有必要去定义新的解决办法来保证数据的保密性和有效的查询执行。

- 安全数据整合。越来越多新的场景需要不同的参与方合作来共享他们的信息，他们每一个都拥有大量可独立管理的数据。因为每方拥有的数据集都包含敏感信息，不能再使用传统的分布式查询评估机制[23,64]。因此，我们需要一个调节各方之间数据流以及重新定义查询评估机制的方法来实现各参与方要求的访问控制限制。事实上，在各合作方之间的数据流可能会因为隐私限制而被禁止，因此使得查询执行的设计必须遵循高效原则和隐私限制。

现实生活中有许多这种应用需要一种机制，通过一种选择性和安全的方式来交换和披露数据的例子。在这我们列出三种可能的情况。

多媒体共享系统。 人们每天收集的多媒体数据的量正在快速地增加。结果是，为图像和视频提供存储和分配服务的系统正变得越来越流行。然而，这些数据可能是敏感的（如图像可以缩小人的大小），如果没有得到数据所有者的明确的授权，那么它们在因特网的广泛传播应当受到阻止。由于考虑到数据保密性，分配

服务可能是不受信任的，它不能执行由数据所有者定义的访问控制策略。因此，有必要去思考阻止敏感数据公开的另一种解决方案。

医疗系统。 越来越多的医疗系统收集关于过去的和现在的住院治疗、诊断以及更多关于病人的一般健康状况等敏感信息。这些数据关联着病人的身份信息，因此是敏感信息。它们的存储、管理及分配都受制于包括国家级和国际的规则。因此，任何医疗系统都应采用一个充分的隐私保护系统，以保证敏感信息永远不会和病人的身份信息储存在一起。

目前，医疗系统的功用已经被扩展，这是由于网络通信技术的演变和广泛传播，允许了各合作方之间的数据交换，如医务人员、药店、保险公司和病人自己本身。虽然这种解决办法提高了提供给患者的服务质量，但是它仍需要小心设计来避免非授权的数据泄漏。因此，很有必要去设计一个数据整合协议来保证数据的保密性。

金融系统。 金融系统储存着需要去充分保护的敏感信息。例如，公司收集的用于信用卡支付的数据是敏感的，无论在存储或管理时都需要保护（如信用卡账号以及相应的安全密码不能存储在一起），这是法律所要求的。此外，由于在线交易的广泛传播，需要系统去管理和保护的金融数据的数量正在急剧增加。与医疗系统一样，为了管理独立的数据收集，金融系统也仍需要和其他方合作，如政府办公室、信用卡公司和客户。

从以上描述可以明显看到，为医疗系统设想的安全问题同样适用于金融系统，在数据存储和数据交换方面都需要相同的解决办法和技术来保证数据的保密性。

1.2 本书的贡献

本书提供了当数据所有者把数据给一诚实但好奇的服务器管理或存储，且数据所有者不直接控制他们的数据时所日益突出的问题的分析。本书的贡献集中在前面所提到的关于安全的三个方面，即访问控制执行、隐私保护和数据安全整合。在本节的其余部分，我们将从更多细节来说明。

1.2.1 访问控制执行

本书的第一个重要贡献是提出对加密的外包数据的访问控制执行[15,41,44]。我们工作的贡献概述如下。