

云计算中图像数据处理的隐私保护

任奎

(纽约州立大学布法罗分校计算机科学与工程学院, 纽约 布法罗 14260)

摘要: 近年来, 随着图像数据与云计算平台的快速增长, 各种各样的图像处理应用蓬勃发展。与此同时, 针对用户外包数据中敏感信息的滥用所产生的安全问题也成为了人们关注的焦点。在现实中, 一旦上传到云平台, 用户隐私信息的安全只能单纯依赖于云服务提供商的信用。为解决这一问题, 对于图像数据处理中的安全性要求与技术难点进行了研究与分析, 并提出了多个在保护外包图像数据安全性的同时实现图像处理应用的功能性要求的解决方案。之后针对云计算中的图像数据的隐私保护问题, 介绍并分析了多种包括同态加密体系、安全多方计算以及差分隐私在内的当前最新技术。

关键词: 云计算; 图像处理; 安全; 加密; 隐私保护

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00020

Privacy-preserving image processing in cloud computing

REN Kui

(Computer Science and Engineering, State University of New York at Buffalo, Buffalo 14260, United States)

Abstract: Enjoying the rapid growth of image data and cloud computing platforms, various image processing applications have emerged and flourished in recent years. Meanwhile, the privacy concerns over the abuse of sensitive information contained in outsourced data also arise in public. In fact, once uploaded to the cloud, the security of users' private information purely depends on the reliability of the cloud service providers (CSP). To solve this problem, the security requirements and technical challenges in privacy-preserving image processing based on different cloud computing architectures were studied, and several solutions to protect the security of outsourced data while enabling functionality of image processing applications were proposed. Several state-of-the-art techniques for secure image processing were introduced and analyzed, including homomorphic encryption (HE) scheme, secure multiparty computation (SMC) protocol, and differential privacy (DP).

Key words: cloud computing, image process, security, cryptography, privacy preservation

1 引言

近年来, 随着数据挖掘以及图像处理技术的广泛应用, 越来越多基于图像处理技术的应用出现在人们的日常生活中。这些技术的创新与发展, 以及随之产生的庞大计算量, 让人们将视角转向同样快速发展的云计算平台。作为即付即用 (pay-per-use)、设施即服务 (computing-infra-

structure-as-a-service) 商业模式下的代表, 云计算平台为用户提供了一种经济且极具灵活性的运算平台^[1]。实际上, 不光中小用户, 甚至有互联网巨头, 包括微软、雅虎等, 开始将自身的数据处理任务交付给第三方云平台, 以获取商业利益。

然而, 云计算平台的引入从不同程度上削弱了用户的隐私保护。这些外包到云计算平台上去

收稿日期: 2016-01-01; 修回日期: 2016-01-08。通信作者: 任奎, kuiren@gmail.com

处理的数据不可避免地将用户的敏感信息泄漏给了云服务提供商（CSP）。这其中包括了用户的个人身份、家庭住址，甚至是财务资料。实际上，云服务提供商作为公共平台的安全常常面临不同层面的威胁，这其中就包括了共享计算资源所带来的安全性隐患；广泛使用在云平台中的虚拟机（SVM, support vector machine）软件共享数据技术多次被发现新的漏洞，并被用作为窃取用户敏感信息的有效手段^[2]。不仅如此，由于各种系统接口调配以及内部员工操作不当引起的用户数据泄漏屡见不鲜。换句话说，作为外包平台，云技术并不是一个毫无风险的保险柜。因此，在现有的架构下，用户一旦上传数据到云服务器上，就失去了对于数据安全的最终掌握，完全将对于敏感信息的隐私保护交于他人之手。近年来，互联网企业中经常出现的用户信息泄漏事件也一次次敲响了警钟。可以说，用户对于外包数据安全性的顾虑，已经成为阻碍云计算技术进一步发展的最大障碍。

作为一个快速发展的研究领域，云计算安全吸引了越来越多学者的关注^[3-6]。作为大数据量、高复杂性计算的典型代表之一，图像数据的安全计算外包更是其中的热门方向。其中一个主要的研究方向为在密文域上对不同图像特征值的检测与提取，例如，图像全局特征尺度如 RGB 直方图、色彩布局描述子（color layout descriptor）、色彩结构描述子（color structure descriptor）^[7]、局部特征如不变特征转换（SIFT）^[8]以及基于形状的图像特征等^[9]。与此同时，还有对于不同制式的加密图像数据的控制，例如脸部识别系统^[10]、心电图信号^[11]，以及指纹识别系统^[12]。除此之外，在安全图像检索领域^[13]，一种利用选择性隐蔽传输（oblivious retrieval）协议针对公共图像数据库的隐私保护图像搜索被提出，并在这之后，将之扩展到了密文图像数据库^[14]。

本文对云计算中图像数据处理中的特征提取运算构建数据流与系统模型，并对设计目标的多个方面进行了理论与实例分析。在此基础上，进一步介绍了目前在图像处理的隐私保护中使用的同态加密、安全多方计算、差分隐私技术等主要技术，并分析了其优势与不足。

2 图像数据处理的隐私保护

2.1 数据流与系统模型

系统模型如图 1 所示，针对图像处理中的图像特征值提取，本文将构建两个主要实体：用户与云计算平台，并描述它们之间的数据流。其中，用户可以是公司或个人，其掌握有大量的需要进行处理（特征值提取）的图像数据。在这种情况下，用户使用云计算平台作为本地计算的补充，即将本地难以处理或计算成本较高的图像特征值提取计算任务外包至云计算平台处运行。然而这些图像数据可能含有大量敏感（私密）信息，如医疗图像、个人照片、资料档案等^[15]。因此，出于对用户个人隐私乃至商业利益的保护，用户选择在本地将图像数据加密后再将密文上传至云计算平台。另一个主要实体云计算平台则是由云计算提供商所运行的一系列服务器构成。在这里，云计算平台只能接受由用户上传的图像密文，再对其执行加密图像的图像特征提取计算。之后，云计算平台在算得所要求的特征值密文后，将其回复给用户。最后由用户使用自己持有的特定对称密钥、非对称密钥将收到的特征值密文解密，以得到其对应明文特征值。

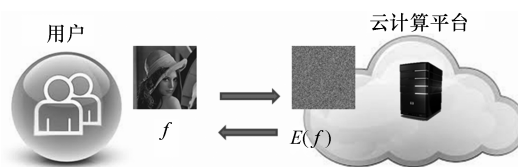


图1 系统模型

这一系统模型可分为两个主要阶段：一个是数据预处理，在这一阶段，对于图像 I ，用户通过 $encode(I) \rightarrow C$ 将 C 传送至云计算平台，这里， $encode(\cdot)$ 表示对于原始图像进行处理以提供密文域特征提取计算操作功能的编码算法（包括预处理与加密算法）的总称，在设计中，这样的编码算法应该是轻量化并且支持尽可能多种的特征提取算法，因此，用户只要将图像数据编码一次即可，在这种结构下，针对已编码图像数据的计算工作量主要由云计算平台所承担；另一个阶段是图像特征提取，云计算平台对于已编码的图像数据执行特征提取算法以得到密文域的图像特征。

综上所述,在这种系统模型下,用户可以得到最大的灵活性与可扩展性来执行大规模的图像特征提取计算。事实上,如果由用户执行部分特征提取计算并将加密的特征上传给云计算平台,那么用户的灵活性将会被极大地削弱,用户需要针对不同的特征提取计算进行多次的图像处理以及加密。不仅如此,哪怕是特征提取算法中微小的参数调整都会导致用户需要对整个图像数据集进行重新处理与加密。

2.2 设计目标与实例分析

这里分析一下系统设计的目标是什么:首先要实现的是功能性,即要求提出的设计能够保证图像特征提取算法在密文域的正确执行;其次是安全性,即要求提出的设计要能提供尽可能强的隐私保护;第三是高效性,即要求用户端与云计算平台端各自的计算复杂度以及他们之间的通信复杂度要在实际应用中可以接受。这样才可能在现实中实现。下面,针对这3个设计目标进行详细分析。

1) 功能性

本文所给出的系统模型对于所执行的图像特征提取算法亦可主要分为两类:全局特征描述子,如RGB直方图、色彩布局描述子(color layout descriptor)、色彩结构描述子(color structure descriptor)^[7]以及局部特征描述子,如尺度不变特征转换(SIFT, scale-invariant feature transform)^[8]。这里对RGB直方图的功能性要求作简要分析,以此作为全局特征提取的典型示例,如图2所示。在颜色特征提取中,直方图描述子是最基础同时也是最为人知的特征描述子。通过直方图描述子,可以求得多种广泛使用的颜色描述子,如颜色结构描述子(CSD)、颜色布局描述子(CLD)等。明文域的直方图运算十分简单,这里不再赘述。但是如果想在密文域执行这一计算,其对加密方法功能性的要求就变得十分苛刻,为了在密文域计算直方图标量,需要让密文间可以进行“比较”,以使每个像素值可以正确分配到直方图量的对应位置中。乍看之下,这一功能性要求似乎与加密的根本功能相互矛盾,如果密文之间可以比较,那么密文的安全性也就无从谈起了。然而,仔细分析直方图的功能需求,会发现其需要的并不一定是明文状态下密文比较的结果,密文状态

下的密文比较结果足够满足要求。通过这一发现,就可以从满足功能性的角度出发设计出可在密文域上实现的直方图特征提取算法(具体算法见文献[7])。

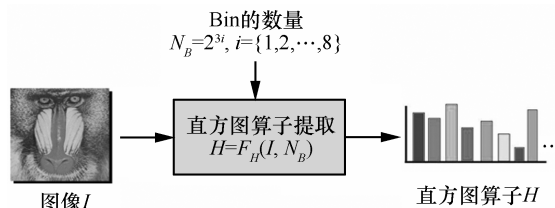


图2 RGB直方图

2) 安全性

根据上文给出的系统模型,假设云计算平台为“诚实但好奇”,云计算平台会根据协议要求,正确地执行密文域的特征提取计算以避免丢失商业信誉所带来的经济损失。然而不管是被入侵的平台管理员账号或是外部攻击者,都可能通过监测密文域数据的操作与传输来推导明文域的敏感信息。为了实现功能性设计目标,提出的系统设计就必然无法实现传统密码学定义的绝对安全性。因此,针对图像隐私保护的安全性分析并不能完全等同于传统的密码学安全性分析。举例来说,如果对图像中的每一个像素点分别进行加密,这样一来,似乎整个图像的安全性可以完全依赖于像素点加密方式的强度,即如果像素点安全性得到保障则图像安全性得到保障。然而如果把安全性与功能性需求放在一起分析,结果将大不相同。这里,以简述不变特征转换的安全性分析来作为局部特征提取的典型示例^[8]。如图3所示,作为局部特征提取算法,不变特征转换首先要提取出图像中关键点的位置,并通过描述关键点周围点的情况来生成所对应的局部特征描述子。在图3中,大小不一的圆圈即为这种局部特征描述子的一种表现形式。由于云计算平台要生成这些局部特征,其不可避免地要推导出这些关键点在图像中的具体位置。然而由图3可以知道,通过这些关键点的位置,即使是密文域的观察者也可以清楚地得到图像中内容的大概形状,这就造成了功能性与安全性的矛盾。如果进一步分析,会发现似乎这个问题与上文描述的直方图像素点数值比较问题的矛盾类似,那么是不是可以用相同

的方式解决呢？即对像素点位置亦进行加密，以实现关键点位置的密文域提取，这样似乎可以同时实现安全性与功能性。然而遗憾的是，这样做会极大地损害系统设计的第三个目标——高效性。在算法复杂度推导中，会发现针对关键点位置的隐藏将会引入大量额外的计算复杂度。而这些额外计算复杂度独立于加密算法的使用。也就是说不论使用何种加密算法，如果想同时实现安全性与功能性，根据图1所示的系统模型，就必然引入难以实现的巨大计算复杂度（具体推导见文献[8]）。为了解决这一问题，提出了基于多方云计算平台的系统模型，通过引入多个独立的云计算平台来同时实现功能、安全、高效3个目标^[8]。

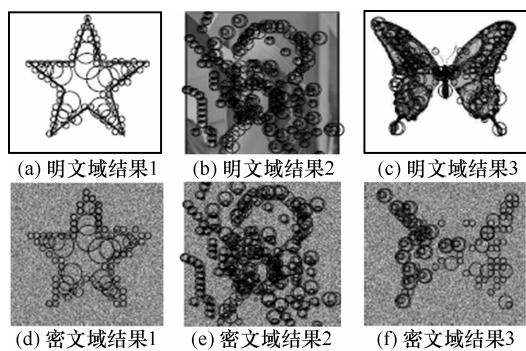


图3 SIFT特征描述子示例

3) 高效性

在安全云计算的时间复杂度分析中，需要从3个方面分析系统的整体效率：用户端和平台端的计算复杂度以及两者之间的通信复杂度。为了实现用户端的灵活与可扩展性，在大部分的设计中，用户端仅承担加密与解密的操作。在大多数的系统设计中，为了实现更复杂的功能性要求，不得不采用更复杂的加密算法，因此加重了用户端的负担。对于有些同态加密算法来说，其加密解密的计算复杂度甚至要超出用户本地计算图像处理任务的复杂度。当这种情况发生时，图像处理计算的外包的高效性也就无从谈起了。所以不仅仅需要设计适合计算任务的加密算法，还需要注意在这3个方面合理分配计算与通信复杂度，从而实现更高效的系统设计。

3 基于同态加密的图像处理

同态加密从提出之日起就一直受到密码学界

最广泛的关注，自从Gentry对其中的全同态加密（FHE, fully homomorphic encryption）产生决定性突破后^[16]，更有多种版本的全同态加密方法被提出。遗憾的是，当前的基于各种传统假设的全同态加密还远未达到实际要求的运算效率。在这种情况下，部分同态加密（somewhat homomorphic encryption）成为实现图像数据处理隐私保护最为强大的工具之一。本文首先简略介绍最具实用性的部分同态加密架构之一，随后会针对图像处理的功能性目标来分析这种加密方法的优缺点。

这里介绍一种基于环上误差学习（RLWE）的部分同态加密架构。本文通过一系列算法来定义部分同态加密架构： $SHE = (SH.Gen, SH.Enc, SH.Add, SH.Mult, SH.Dec)$ 。其中， $SH.Gen$ 通过随机选取环元素产生 a_1 、 e 以及私钥 sk ，并得到公钥 $pk = (a_0 = -a_1s - te, a_1)$ 。 $SH.Enc$ 、 $SH.Dec$ 以及加法与乘法的同态性质^[6]如下

$$SH.Enc(pk, m) = ct = (c_0 = a_0u + tg + m, c_1 = a_1u + tf) \quad (1)$$

$$SH.Dec(ct, sk) = m = \sum_{i=0}^{\delta} c_i s^i \pmod{t} \quad (2)$$

$$SH.Add(ct, ct^*) = (c_0 + c_0^*, \dots, c_{\max} + c_{\max}^*) \quad (3)$$

$$SH.Mult(ct, ct^*) = (\sum_{i=0}^{\delta} c_i v^i) (\sum_{i=0}^{\delta} c_i^* v^i) \quad (4)$$

值得注意的是，上述这些性质只在有限次数的同态加法和乘法中有效。

对于同态加密架构，除了图像特征提取外，在图像领域最常见的应用是在各种特征匹配算法的安全计算中。其中一个典型示例就是隐私保护的脸部识别^[17]，如图4所示。假设用户将脸部照片 I 加密后上传至云端数据库，数据库在密文域对图像 I 进行操作，从中提取出密文状态下的特征向量 $\vec{w}$ 。随后将其分解成多个特征脸 $e_i(i, j)$ 并与数据库中存储的特征脸进行比对（计算欧氏距离），其中距离最小的特征脸即为最为匹配的脸部照片。同样类似的特征匹配计算还有各种生理匹配算法、分类算法（classification）以及聚类算法（clustering）。在实现这些算法中最常遇见的挑战就是乘法次数的限制。由于可容纳多次乘法的部分同态加密架构会快速地增加计算复杂度，避免这个问题的常见做法是将多方安全计算、混乱电

路^[18]等技术结合在一起使用。例如，将乘法、阶乘等同态性质难以达到或非常昂贵的计算步骤通过这些替代技术来实现，从而达到功能性与高效性的平衡。

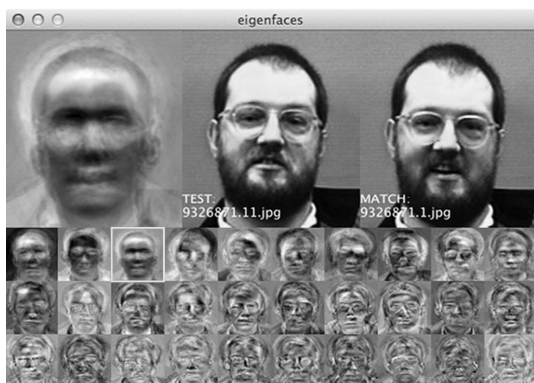


图4 基于特征脸的面部匹配示例

4 基于安全多方计算的图像处理

安全多方计算^[18]协议一直被认为可以用来计算任意函数。然而，由于所需要的庞大计算以及通信复杂度，使其难以在实际中大规模应用。但是，安全多方计算在兼容性以及算法的简便性等方面的优点，使其在云计算安全设计中仍然扮演着重要的角色。其中，安全双方计算（secure two-party computation）被一些云安全计算解决方案作为不少解决方案的基础协议来与同态加密、混沌电路等技术相结合。不仅如此，有时在一些图像处理计算问题的解决中，引入额外的第三方云服务器不失为一种在安全性与功能性之间的有效平衡。

由于不变特征转换具有相当高的计算复杂性，如密文域的数值比较、极值点的发现与隐藏、密文域上三角函数的计算等复杂计算。对于这些计算问题，当前的同态加密架构难以高效计算，有些甚至难以解决。因此，如果尝试改变单独云计算平台的系统架构，通过引入多个独立云计算平台来解决问题，这样所提出的解决方案就可以同时兼顾各个方面的要求。如图5所示^[8]，通过引入额外的云计算平台来利用安全多方计算技术，用户通过简单的一次性密码本（one time pad）加密方式，可以将一份明文图像转化为两份同样大小的密文，之后，两个云计算平台就可以对于

各自收到的不同密文进行操作来实现相应的功能性。这一设计与安全多方计算技术对于算数性（arithmetic）加密方式的利用遵循了同样的方法论，不仅如此，对于安全多方计算技术的使用，还为平衡计算复杂度与通信复杂度提供了新的角度。以图5为例，在类似安全多方计算的架构下，再引入额外通信复杂度（将密文上传给两个云平台）后，用户可以利用多种算数性加密方式来实现密文域的功能计算，并同时享受这些加密方式的低计算复杂度。

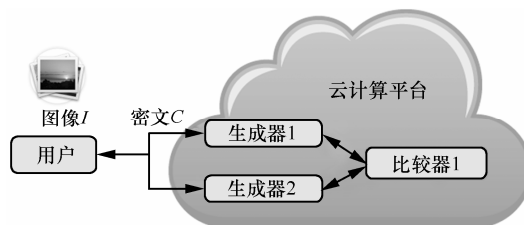


图5 通过引入额外的云计算平台来利用安全多方计算技术

5 基于差分隐私的图像处理

差分隐私（differential privacy）作为一种针对保护隐私的数据发布技术在2006年由Dwork提出^[19]。近年来，本地差分隐私（local differential privacy）概念以及相关技术的提出为具备隐私保护的大数据收集技术提供了新的方向^[20]。不同于传统密码学对于安全的定义，差分隐私并不能百分百地阻止信息的泄漏，而是通过统计的方式提供理论上的量化边界来约束敏感信息的泄漏。在此基础上，本地差分隐私进一步细化了对于个人隐私的保护而不仅仅是像传统差分隐私一样将自己“藏在”统计结果中。这一安全概念的提出与发展在传统密码学之外为一些特定问题的解决提供了新的思路。基于差分隐私技术的图像处理系统架构如图6所示，不同于现有的加密方式，差分隐私可以使用“噪声干扰（noise perturbation）”这种“加密”方式来保护单个数据的安全性。而最终通过聚合大量含有“噪声”的数据来“抵消”其中的“噪声”，从而提取出其中的特征信息^[19]。本地差分隐私的主要缺点是其需要海量的数据集来实现相比一般统计结果可接受的统计精度（百万级以上）。但是在解决云计算平台上的图像数据处理这一问题时，数据量的巨大需求所带来的负

面影响减小了,这一变化更使得本地差分隐私技术在海量数据处理时高效的优势得到有效利用,两种需求的契合让差分隐私在这一领域的应用十分具有潜力。

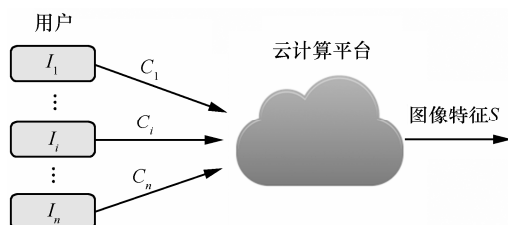


图6 基于差分隐私技术的图像处理系统架构

6 结束语

基于云计算平台的图像数据处理的隐私保护近年来被学界所关注。相比其他计算任务,图像数据处理具有数据量大、计算复杂度高等技术上的挑战。包括同态加密、安全多方计算、混乱电路在内的多种技术被用来实现其安全性、功能性以及高效性的设计要求。本文从系统模型开始,通过设计目标的3个方面结合实际问题具体分析,最后介绍了当前几种主要技术方案,并分析了其中的优缺点。云计算的快速发展与图像数据的迅猛增长相辅相成,只有有机地结合两者的特点才能做出创新又切合实际的研究成果。

参考文献:

- [1] ARMBRUST M. A view of cloud computing[J]. Communications of the ACM, 2010, 4 (53): 50-58.
- [2] MODI C, PATEL D, BORISANIYA B, et al. A survey of intrusion detection techniques in cloud[J]. Journal of Network and Computer Applications, 2013, 36(1):42-57.
- [3] LU W. Secure image retrieval through feature protection[C]//IEEE International Conference on Acoustics. c2009:1533-1536.
- [4] HSU C Y. Image feature extraction in encrypted domain with privacy-preserving SIFT[J]. IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society, 2012, 21(11): 4593-4607.
- [5] HSU C Y. Homomorphic encryption-based secure SIFT for privacy-preserving feature extraction[C]//Society of Photo-optical Instrumentation Engineers Conference Series. c2011.
- [6] NAEHRIG M. Can homomorphic encryption be practical? [C]//The 3rd ACM Workshop on Cloud Computing Security Workshop. c2011:113-124.
- [7] QIN Z, YAN J, REN K, et al. Privacy-preserving outsourcing of image global feature detection[C]//IEEE Global Communications Conference. c2014:710-715.
- [8] QIN Z, YAN J, REN K, et al. Towards efficient privacy-preserving image feature extraction in cloud computing[C]//ACM International Conference on Multimedia. c2014: 497-506.
- [9] WANG S, NASSAR M, ATALLAH M, et al. Secure and private outsourcing of shape-based feature extraction[C]//Information and Communications Security. c2013:90-99.
- [10] ERKIN Z, FRANZ M, GUAJARDO J, et al. Privacy-preserving face recognition[M]//Privacy Enhancing Technologies. Berlin: Springer Berlin Heidelberg, 2009.
- [11] BIANCHI T. Privacy-preserving fingercode authentication[C]//12th ACM Workshop on Multimedia and Security, Rome. c2010:231-240.
- [12] BARNI M, FAILLA P, LAZZERETTI R, et al. Privacy-preserving ECG classification with branching programs and neural networks[J]. IEEE Transactions on Information Forensics & Security, 2011, 6(2):452-468.
- [13] LU W. Enabling search over encrypted multimedia databases[C]//ISQT-SPIE Electronic Imaging Symposium on Media Forensics and Security I, San Jose, CA, c2009.
- [14] KHAN M K, ZHANG J, ALGHATHBAR K. Challenge-response-based biometric image scrambling for secure personal identification[J]. Future Generation Computer Systems, 2010, 27 (4): 411-418.
- [15] PANDEY S, VOORSLUYS W, NIU S, et al. An autonomic cloud environment for hosting ECG data analysis services[J]. Future Generation Computer Systems, 2012, 28 (1): 147-154.
- [16] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//The 41st Annual ACM Symposium on Theory of Computing. c2009:169-178.
- [17] Processing-eigenfaces. An example of using the p-eigenfaces processing library to perform face recognition[EB/OL]. <https://github.com/atduskgreg/Processing-Eigenfaces>.
- [18] YAO A C. Protocols for secure computation[C]//The 23rd Annual Symposium on Foundations of Computer Science. c1982:160-164.
- [19] DWORK C. Differential privacy[M]//Automata, Languages and Programming. Berlin: Springer Berlin Heidelberg, 2006:1-2.
- [20] DUCHI J C, JORDAN M, WAINWRIGHT M J. Local privacy and statistical minimax rates[C]//The 54th Annual Symposium on Foundations of Computer Science Berkeley, 2013:429-438.

作者简介:



任奎(1978-), 男, 安徽巢湖人, 纽约州立大学布法罗分校计算机与科学工程系教授, 主要研究方向为云计算中的数据安全、计算服务外包安全、无线系统安全、隐私保护、物联网系统与安全。