doi:10.11959/j.issn.1000-436x.2016078

隐私计算研究范畴及发展趋势

李凤华[1]，李晖[2]，贾焰[3]，俞能海[4]，翁健[5]

（1.中国科学院信息工程研究所信息安全国家重点实验室，北京 100195；
2. 西安电子科技大学网络与信息安全学院，陕西 西安 710071；3. 国防科学技术大学计算机学院，湖南 长沙 410073；
4.中国科学技术大学信息科学技术学院，安徽 合肥 230026；5. 暨南大学信息科学技术学院，广东 广州 510632）

摘 要：随着移动互联网、云计算和大数据技术的广泛应用，电商、搜索、社交网络等服务在提供便利的同时，大数据分析使用户隐私泄露的威胁日益凸显，不同系统隐私保护策略和能力的差异性使隐私的延伸管理更加困难，同一信息的隐私保护需求随时间变化需要多种隐私保护方案的组合协同。目前已有的各类隐私保护方案大多针对单一场景，隐私缺乏定量化的定义，隐私保护的效果、隐私泄露的利益损失以及隐私保护方案融合的复杂性三者之间的关系刻画缺乏系统的计算模型。因此，在分析隐私保护研究现状的基础上，提出隐私计算的概念，对隐私计算的内涵加以界定，从隐私信息的全生命周期讨论隐私计算研究范畴，并从隐私计算模型、隐私保护场景适应的密码理论、隐私控制与抗大数据分析的隐私保护、基于信息隐藏的隐私保护以及支持高并发的隐私保护服务架构等方面展望隐私计算的发展趋势。

关键词：隐私计算；度量；形式化描述；隐私感知

中图分类号：G643.0 **文献标识码：**A

Privacy computing：concept, connotation and its research trend

LI Feng-hua[1], LI Hui[2], JIA Yan[3], YU Neng-hai[4], WENG Jian[5]

(1. State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100195, China;
2. School of Cyber Engineering, Xidian University, Xi'an 710071, China;
3. College of Computer, National University of Defense Technology, Changsha 410073, China;
4. School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China;
5. College of Information Science and Technology, Jinan University, Guangzhou 510632, China)

Abstracts: With the widespread application of mobile Internet, cloud computing and big data technologies, people enjoy the convenience of electronic business, information retrieval, social network and other services, whereas the threats of privacy leaks are ever growing due to the use of big data analytics. The differences of privacy protection strategy and ability in different systems bring more difficulties in privacy extended management. In addition, the requirements of protecting the same information at different time need the combination of various privacy protection schemes. However, nearly all the current privacy protection schemes are ing at a single case, which lacks systematic and quantized privacy characterization. Furthermore, there is no systematic computing model describing the relationship between the protection level, profit and loss of privacy leaks and the complexity of integrated privacy protection methods. Based on the analysis on the research status of privacy protection, the concept and connotation of privacy computing is proposed and defined. Then the privacy computing research category will be discussed from the whole life cycle of information privacy protection. Finally, some research directions of privacy computing are given, including privacy computing model, context adaptive cryptology for privacy protection, big data a lytics resisted privacy control and protection, privacy protection based on information hiding and system architecture for high concurrent privacy preserving services.

Key words: privacy computing, quantification, formal description, privacy awareness

收稿日期：2016-03-01；**修回日期：**2016-04-13
通信作者：李晖，lihui@mail.xidian.edu.cn
基金项目：国家自然科学基金—广东联合基金资助项目（No.U1401251**）；移动互联网安全学科创新引智基地（“**111**”计划）基金资助项目（**No.B16037**）**

Foundation Items: The National Natural Science Foundation of China-Guangdong Joint Program (No.U1401251), 111 Project (No.B16037)

1 引言

移动互联网、云计算和大数据等技术的快速发展，孕育并产生了各种新的服务模式和应用，例如滴滴打车、百度外卖等基于位置的服务。这些服务和应用一方面采集用户的相关信息，另一方面为用户提供精准化、个性化的服务，给人们的生活带来了极大便利。然而，所采集信息中往往含有大量包括病史、收入、身份、兴趣及位置等在内的敏感信息，对这些信息的共享、收集、发布、分析与利用等操作会直接或间接地泄露用户隐私，给用户带来极大的威胁和困扰。因此，用户隐私保护已成为人们广泛关注的焦点。

当前，各国政府、工业界和学术界都对隐私保护问题开展了一系列的工作。美国于 1974 年通过了《联邦隐私法案》，该法案是美国最重要的一部保护个人隐私权方面的法律，用来严格限制政府侵犯公民个人隐私。1980 年，世界经济与合作发展组织（OECD）在美国隐私保护条例的基础上进行扩充，制定了 8 条隐私保护条例，后来演变为隐私保护原则，目前，这些原则已经成为一个被广泛接受的隐私保护标准。欧盟于 1995 年发布的《欧洲数据保护指导条例》中就包含了 OECD 原则，许多非欧盟国家或地区（如澳大利亚、新西兰、加拿大及中国香港地区等）也通过了基于 OECD 原则的法律。美国《电子通信隐私法》中规定：如果提供商未被授权访问内容，并在未经用户同意的情况下，提供商擅自公开用户信息，用户具有提起民事诉讼的权利。美国《联邦信息安全管理法案》、《美国爱国者法案》、《美国医治保险携带和责任法案》、《经济与临床健康信息计数法》等法律法案中，也明确列出了对隐私信息进行保护的条款。《第 95/46/EC》号法令中，规定了个人数据处理方面对个人的保护以及此类数据的自由范围，例如限制将个人数据传输到欧盟以外。欧盟于 2012 年发布了通用数据保护法案，制定了包括数据确权、删除、流转范围、违规使用等方面的法规。美国政府在 2012 年 2 月宣布推动《消费者隐私权利法案》的立法程序，2016 年，美国联邦通讯委员会 FCC 向国会提交了用户隐私保护法令。目前，世界上 50 多个国家和地区制定了保护个人信息的相关法律。

针对隐私保护问题，学术界开展了大量的研究工作，并在社交网络、位置服务、云计算、大数据、智能医疗、智能电网、智能交通等方面提出了诸多具体的隐私保护方案。然而，目前，隐私保护的学术研究尚缺乏对隐私的概念、度量等名词的定义和描述。仅有少量工作从经济学角度展开研究，如文献[1]提出了一种隐私演算（privacy calculus）模型，对用户泄露的个人信息与随之带来的经济和社会利益进行评估。若这些个人信息能被公正地使用，不产生负面影响，用户将愿意透露个人信息以换取一定的经济或社会利益。文献[2~4]从风险收益的角度研究用户网上隐私信息披露行为，分析隐私披露行为与感知收益、隐私忧虑间的关系，认为信息披露行为带给用户的感知收益要必然高于其所负担的风险。

但以上研究未对隐私进行形式化描述和定量分析，使隐私信息在不同系统、不同用户间共享、交换和分析过程中难以被准确刻画和量化，阻碍各类计算和信息服务系统对隐私进行统一评价。

针对这一问题，本文提出隐私计算的概念，阐述隐私计算的范畴、内涵和框架，并给出涉及的关键理论与技术，旨在建立科学的隐私保护研究体系。

2 隐私保护技术的研究进展

2.1 抗大数据分析的隐私保护

对于大数据中的结构化数据而言，由于攻击者可以从多种渠道获得关联信息进而推测出用户的隐私信息，导致用户隐私泄露。因此，亟待解决抗大数据关联分析和深度挖掘的隐私保护问题。

2.1.1 数据扰乱

数据扰乱是当前最常用的隐私保护技术之一。基于特定策略修改真实的原始数据，数据扰乱使攻击者无法通过发布后的数据来获取真实数据信息，进而实现隐私保护。数据扰乱通常包含数据泛化、数据扭曲、数据清洗、数据屏蔽等。数据泛化通常指使用数值型或枚举型的属性值来替换真实数据，使发布数据中所含信息的粒度降低[5]。数据扭曲主要通过将随机数值与原始数据进行叠加来修改真实数据的数值以实现隐私保护，叠加方法通常采用加性叠加[6]或乘性叠加[7~9]。数据清洗主要是基于隐藏数据潜在关联规则的原理，通过修改或移除某些特定的数据使相应的频繁项集的支持度降低[10,11]。数据屏蔽通常利用符号来代替隐私属性值，采用基于概率分析的修正方法[12,13]，实现隐私保护的同时能提高数据分析的精度。

2.1.2　静态数据发布的匿名模型

目前，采用的隐私保护算法通常基于几种典型的隐私保护模型。其中，最早由 Sweeney 提出的 k-匿名(k-anonymity)[14]，其基本思想是将用户数据匿名化，使用户隐私无法被识别出。但该模型的缺陷是其只约束了准标识符属性，没有对敏感属性进行约束，因此容易遭受同质性(homogeneity attack)攻击[5]。针对 k-匿名的缺陷，又相继提出一些改进的方法，包括 l-多样性(l-diversity)[5]和 t-贴近性(t-closeness)[15]等。l-多样性要求发布数据集中的每个等价类至少包括 l 个敏感属性值的“代表”。t-贴近性则要求敏感属性值在每个等价类中的分布与数据集中的全局分布差异不超过 t。后续的研究主要是基于 k-匿名中的等价类概念，进一步增加对等价类中敏感属性的约束，实现降低隐私泄露的风险[16,17]，主要模型有置信度边界(confidence bounding)模型[18]、(a, k)-匿名模型[19]、(k, e)-匿名模型[20]、(e, m)-匿名模型[21]、(n, t)-closeness 匿名模型[22]等。然而，这些传统的隐私保护模型通常只适用于特定场景下的数据发布或仅是一次性发布的场景。而在现实中，大多数数据发布场景都具有数据连续、多次发布的特点，这就需要防止攻击者对多次发布后的数据进行关联分析，从而破坏数据的匿名特性[23]。因此需要拓展现有的隐私保护模型以适用于大数据环境下的数据隐私应用场景。

2.1.3　动态数据发布的隐私保护模型

考虑到多种数据发布场景的特点，研究者提出了多版本发布、相继发布、连续数据发布和联合数据发布等技术。其中，多版本发布(multiple release publishing)[24]主要是针对相同的原始数据集，依据不同用途或针对不同的发布对象，通过选择不同的属性来同时发布匿名后的数据。然而，在获得了 2 次或以上的发布数据集，攻击者可通过关联多个数据集来推断出某些敏感属性。相继发布(sequential release publishing)[25]指数据发布者已经发布了数据集 $T_1,T_2,\cdots,T_{p-1}$，当前若需要发布数据集 T_p，数据拥有者只能对数据集 T_p 进行匿名化。连续数据发布(continuous data publishing)[26,27]是指数据发布者已经发布了数据集 $T_1,T_2,\cdots,T_{p-1}$，当需要发布数据集 T_p 时，所使用的原始数据集是基于 T_{p-1} 对应的原始数据集进行增删改操作后生成的数据集。因此，若某个用户存在于连续多个数据集中，该用户可能会遭受联合攻击。联合数据发布(collaborative data publishing)[28,29]的场景是数据分布式存储在 2 个组织时，需要联合后再共同发布给第三方。此时，数据除了要避免泄露给第三方，还要防止泄露给数据的其他拥有者。因此需要采取一些方法以在执行匿名化操作时避免泄露敏感信息。除上述发布场景外，还存在数据敏感属性有多个值的情形，文献[30，31]研究了这种多敏感属性的数据集匿名化问题。

2.1.4　隐私保护的数据挖掘

隐私保护数据挖掘(PPDM, privacy preserving data mining)是指针对采用数据扰乱等匿名技术处理后仍有足够精度和准确度的数据集，数据挖掘者在不接触实际隐私数据的情况下仍然可以进行有效的挖掘[32,33]。PPDM 主要关注两方面的问题：一是原始数据中的敏感信息，二是敏感规则。这些敏感规则通常隐含在原始数据集当中，须通过数据挖掘才能识别。依据隐私保护技术的特性和挖掘环境的不同，选取合适的隐私保护与数据挖掘技术，形成有效的隐私保护数据挖掘算法。文献[34]提出了一种用于集中式环境下的隐私保护分类挖掘算法，算法不需要重构原始数据的分布函数，通过在干扰后的数据集上构造朴素贝叶斯分类器[35]实现分类挖掘。该算法具有一定的局限性，只适用于朴素贝叶斯分类挖掘。文献[36]提出了在一种数据水平分布条件下的隐私保护关联规则挖掘算法，该算法采用安全并集计算协议和安全求和计算协议，进一步增强了安全性。文献[37]提出了一个适用于数据水平分布条件下的隐私保护聚类挖掘算法，该算法支持处理规模较大的数据库，且不会泄露中间通信过程中候选的簇中心信息，在保护数据隐私的前提下可得到更准确的聚类结果。然而，已有的隐私保护数据挖掘算法无法满足具有异构、跨平台等特点的大数据环境下海量数据的高效、精确挖掘要求，设计满足大数据需求的数据挖掘算法还有很多亟待解决的问题。

2.2　基于密码方法的隐私保护

2.2.1　同态加密

同态加密允许用户直接对密文进行特定的代数运算，得到的仍是加密的结果，将其解密所得到的结果与对明文进行同样的运算结果一样。优势在于可以对加密数据进行分析和检索，因此可以由第三方来处理隐私数据。

同态加密的概念最早由 Rivest 等[38]在 1978 年首次提出，随即成为了密码学界的开放难题，国外

学者相继研究了满足乘法或加法的同态加密算法，还提出了能同时满足有限次乘法与加法的同态密码，直到 2009 年，Gentry 才构造出了第一个全同态加密方案[39]，解决了困扰密码学界 30 多年的难题。全同态加密被认为是解决云计算安全的最好方法，利用全同态加密方案对用户数据进行加密，再将密文发送到云端，云端可以在不解密的情况下进行检索和比较等操作，避免了数据存储方泄露数据的危险。但现有的全同态加密方案计算复杂度相对较高，无法应用到实际系统中。

2.2.2 安全外包计算技术

安全外包是云计算隐私保护中不可或缺的关键技术之一。在科学计算领域，研究者针对各种具体科学运算安全外包方案进行了大量研究。针对大规模数值计算中的安全外包问题[40]，Atallah 等提出了一种安全的序列比较外包方案[41]。随后，线性代数运算（如大规模矩阵乘法等）安全外包协议、基于单服务器模型的高效矩阵乘法安全外包、线性方程组安全外包等安全外包方案不断被提出。在理论研究方面，Gennaro 等[42]首次提出了非交互式可验证计算的概念，并基于混淆电路和全同态加密技术给出了一个适用于任意运算可验证外包的框架。在密码学领域，Chaum 等[43]于 1992 年提出了“wallet with observers”的概念，即服务提供商通过在客户计算机中安装安全硬件从而帮助资源受限的用户完成复杂的密码运算，形成了外包密码运算的雏形。随后，研究者对各种具体密码运算的安全外包进行了大量研究。研究各种特定函数的高效安全外包算法具有非常重要的实用意义。

2.2.3 密文搜索

可搜索加密技术(searchable encryption)弥补了传统加密无法对密文进行直接操作的缺陷，具有重要的理论意义和应用价值。可搜索加密目前以理论研究为主，分为对称可搜索加密(SSE, symmetric searchable encryption)[44]和非对称可搜索加密(ASE, asymmetric searchable encryption)[45]。在实现数据共享方面，公钥加密比对称加密更有优势。对称加密大多适用于单用户场景，而公钥加密可应用于多用户场景。

带关键字搜索公钥加密方案(PEKS, public key encryption with keyword search)是公钥可搜索加密的代表性方案，随后的研究在此基础上进行扩展和改进，包括支持对关键字的联接词、子集、范围比较等查询[46]、带关键字搜索的公钥加密和代理重加密(PRE, proxy re-encryption)结合[47,48]、无安全信道的带关键字搜索公钥加密(SCF-PEKS, secure channel free public key encryption with keyword search)[49]及支持多关键字搜索并能够对返回结果进行排序的加密搜索方案[50]等。但目前的方案还不能解决多域环境中实现高效的多关键字检索的问题。

MIT CSAIL 实验室开发了数据库软件 CryptDB，允许用户查询加密的 SQL 数据库，且能在不解密存储信息的情况下返回结果。系统通过将请求数据的软件和存储加密数据的数据库之间放置代理服务器来保证对加密数据的分析。在某些情况下，代理需要去除不同的加密层来更好地分析数据。CryptDB 首次解决了实用性问题，它将数据嵌套进多个加密层，每层都使用不同的密钥，允许对加密数据进行简单操作，计算时间方面只增加 15%~26%左右[51]。谷歌使用它在搜索大量数据集的 BigQuery 服务中提供加密查询。CryptDB 本质上是通过对已有加解密算法的嵌套，实现对加密数据的细粒度、高效查询与处理。然而，CryptDB 目前仍然是简单的嵌套，并未达到加密算法的深度融合，且查询种类较少、一次查询仍需执行多个加解密算法，因此亟待研究加解密算法间的深度融合。

2.3 多媒体数据的隐私保护

信息隐藏是将消息嵌入如数字图像、音频、视频等媒体数据的一项技术。信息隐藏根据应用的不同包含多个分支，如数字隐写、数字水印、数字指纹、可逆隐藏等，在隐蔽通信、版权保护、数据溯源、完整性认证等领域有重要应用。

2.3.1 可逆隐藏

可逆隐藏[52]能够在提取消息后无损地重构载体，主要用于军事、医疗和司法等领域的敏感图像完整性认证或标注。

目前，主流的可逆隐藏方法通过 2 个步骤实现。第一步，生成一条熵尽量小的序列作为载体；第二步，在保持可逆的条件下，以尽量小的失真在载体序列中嵌入消息。主要的嵌入策略是修改载体序列的直方图，包括直方图平移[53]和扩差[54]等方法，这 2 种方法被认为是次优嵌入方法。

近年来，可逆隐藏被扩展到密文域[55]，用于加密图像的外包存储，这种技术在公有云环境中可以方便云服务器在不得到内容的情况下管理图像。可逆隐藏

也被考虑用于视频监控中的隐私信息保护[56]，但受到容量限制，其保护内容有限。因此，扩展现有可逆隐藏算法，并将其用于图像、视频、音频等多媒体内容的隐私保护中是可逆隐藏的发展趋势之一。

2.3.2　加密域多媒体安全检索与水印溯源

在大规模多媒体分发应用中，为实现对特定数据的溯源，Hartung[57]在服务端嵌入用于版权追踪的数字指纹，并在加密后分发给不同用户以实现溯源。该方案具有较高的安全性，但会大幅增加带宽消耗和服务端的计算量。为此，Parnes 和 Liu 提出了广播加密方案和部分指纹方案以改进上述不足[58,59]。Kundur 等提出了联合式指纹嵌入与解密方案和 JFD 概念[60]。基于 JFD 思想的密文安全指纹嵌入研究工作还有 Lemma 和 Sadeghi 等近期提出的方法[61,62]。JFD 方案主要的问题集中在 2 个方面：1）广播传输的多媒体加密数据对于密码攻击的安全性、加密效果和计算效率存在问题；2）未解密的密文数据所代表的指纹性能不够理想。目前，面向大规模客户端指纹的相关研究[63~65]，主要思路集中在基于部分加密和查表法上。近年，启动的欧盟第 6 次框架计划项目（No.034238, SPEED project—signal processing in the encrypted domain）中，专门研究了加密域中的信号处理，其中包括了大规模多媒体加密域指纹和溯源技术[66]。因此，如何实现高效的加密域指纹检索与安全水印方法是大数据环境下多媒体隐私保护的重要研究方向。

2.4　网络通信的隐私保护

2.4.1　隐写和隐写分析

隐写术作为密码术的补充，可以实现隐蔽通信以保护通信行为隐私。

隐写术已从“非自适应”过渡到了“自适应”阶段，即优先修改失真小（难检测）的区域[67]。针对自适应需求，Filler 等提出了最小化失真隐写编码，称为 STC[68]。STC 主要针对“加性失真”的编码，即假设修改各像素造成的失真是独立的。自 STC 编码以后，隐写术的研究重点主要集中在如何设计合理的失真函数上[69,70]。目前，隐写术的最新趋势是研究非加性失真隐写[71,72]。

隐写分析主要研究如何检测隐写行为是否发生。目前，隐写分析技术的主流思路是：利用载体中相邻元素的相关性提取特征，然后利用机器学习训练分类器区分载体和载密对象[73,74]，最新进展包括针对自适应隐写设计的“自适应隐写分析方法”[75,76]。其未来发展趋势是与大数据分析技术结合，通过分析用户的行为、个性等方式来识别隐写者。

大数据分析与隐写分析的融合将迫使隐写术设计理念发生改变。未来的安全隐写术必须关注通信行为是否完整，载体表达的语义与当前场景是否一致，与使用者的历史行为习惯是否一致等因素。这对隐写者提出了很高的要求，将使隐写工具的易用性大打折扣。因此，未来的隐写工具不能是单一的隐写软件，而应该是一个完备的智能隐写系统，可以感知用户的历史行为习惯、当前的场景，从而配置隐写策略。

2.4.2　匿名通信网络

1981 年，Chaum[77]首先提出了 Mix 技术和匿名通信的概念。所谓 Mix 技术是通过多层加密和多个中间节点路由来实现消息匿名。每个中间节点成批地收取消息，并随机排序，只有本机才能解开上一层加密，然后将消息传递给路由链中的下一个节点。因此，任何人都无法从源头到目的地完整地追踪消息及路由过程，包括转发消息的中间节点。继 Mix 之后，大量匿名通信系统被提出。

在众多匿名通信中，由于 Tor 匿名通信系统[78,79]具有配置简单、性能优良、前向安全、拥塞控制、可变出口策略及端到端的完整性检测等特点，使它成为目前互联网上应用最为广泛的匿名通信系统之一。Tor 网络实现了对数据发送方和接收方的匿名以及双方通信关系的隐藏[80]。然而，Tor 匿名通信系统存在多种安全缺陷[81~84]，如匿名链路的首节点和尾节点可分别与消息发送者和消息接收者直接通信，这使攻击者容易控制匿名通信系统，从而破坏系统的匿名性；另外，节点加入过程缺乏安全性验证；中继节点发生故障时将导致该通信链路中断或 Tor 用户绕行，密钥协商过程易受中间人攻击等。

2016 年，Chaum 等首次提出了一种称作“隐私完整”（PrivaTegrity）的新加密方案[85,86]，该方案指出：隐私完整可提供完全私密的匿名通信。通过“隐私完整”，Chaum 等引入一种全新的混合网络 cMix。在他所设定的 cMix 中，安装了该应用的智能手机将通过与每台服务器共享的一系列密钥与“隐私完整”的 9 台服务器进行通信。Chaum 声称所提出的“隐私完整”的设置比 Tor 更安全，而且由于整个 cMix 过程只需要简单的乘法和除法操作，相对于采用公钥计算的“混合网络”，cMix 更加高效。此外，

不法分子可利用匿名通信系统来进行匿名通信，为其在网络中实施不法行为增加掩护，阻碍审查机构对其的监控与追踪。“隐私完整”除了安全和效率的优势外，还在 9 台服务器架构中为“管理者”提供了不接触任何密钥的第 10 台服务器，它通过联合全部 9 台服务器，重构出消息完整路由路径，并除去用来相乘以加密的随机数，从而实现后台解密，为匿名网络犯罪的追踪提供了可能。

3 隐私计算及其研究范畴

3.1 隐私

隐私的概念在不同国家、宗教、文化和法律背景下，涵盖的范围差别会很大。OECD 对隐私的定义为：“任何与已知个人或可识别的个人相关的信息”[87]。美国注册会计师协会(AICPA)和加拿大特许会计师协会(CICA)在公认隐私原则(GAPP)标准中定义隐私为“个人或机构关于收集、使用、保留、披露和处置个人信息的权利和义务”[88]。

为了对隐私计算的研究范畴给出一个合理的限定，隐私计算中的隐私可界定为：隐私是指个体的敏感信息，群体或组织的敏感信息可以表示为个体的公共敏感信息。因此，可以将信息分为公开信息、秘密信息、隐私信息 3 类。对组织而言，信息包括公开信息和秘密信息；对个人而言，信息包括公开信息和隐私信息 2 类。

隐私信息可描述为 n 维变量 $x=(x_1,x_2,\text{L},x_n)$，每个分量 x_i 表示一类隐私信息，如年龄或住址，其取值范围可定义为一个集合或者一个连续范围区间。为描述方便，本文只讨论隐私分量为离散值的情况。在此集合上定义隐私量化测度 $I(X_i)$，称之为隐私度量。不同分量的取值集合组成了整个隐私信息的取值集合。集合的运算以及定义在集合上的测度运算可应用于隐私的计算模型。

隐私信息的度量可以采用 Shannon 信息熵。假定分量 x_i 的取值集合 X_i 包含 J 个可能取值 $x_{ij}, j=1,2,\text{L},J$，每个取值的概率为 $P(x_{ij})$，则 X_i 包含的平均隐私度量为

$$I(X_i)=-\sum_j P(x_{ij})\log P(x_{ij}) \tag{1}$$

类似地，2 个以上分量的联合隐私信息可以用联合概率分布描述。当 2 个隐私分量 X_i、X_j 不是独立的时候，它们之间的相互关联可以由互信息来表示

$$I(X_i;X_j)=\sum_{i,j}P(x_i,x_j)\log\frac{P(x_i\,|\,x_j)}{P(x_i)} \tag{2}$$

实际上隐私信息的关联分析就是针对互不独立的隐私分量获取目标隐私分量信息的过程。从若干个相关隐私分量 $X_{j1},X_{j2},\text{L},X_{jr}$ 能够获得的目标隐私分量 X_i 的最大信息量为 $I(X_i;X_{j1},X_{j2},\text{L},X_{jr})$。若 $I(X_i;X_{j1},X_{j2},\text{L},X_{jr})=I(X_i)$，则可以通过关联分析，获得目标隐私分量 X_i 的全部信息。

3.2 隐私计算的定义

含有隐私的信息会在网络中传播、在各类信息服务系统中存储、处理（编辑、融合、发布和转发）。在这一过程中隐私感知、隐私保护、隐私分析都依赖于对隐私信息的定量化描述、隐私信息处理过程中形式化描述、隐私度量演化的公理化描述体系。为此，本文提出隐私计算的定义。

隐私计算是面向隐私信息全生命周期保护的计算理论和方法，是隐私信息的所有权、管理权和使用权分离时隐私度量、隐私泄漏代价、隐私保护与隐私分析复杂性的可计算模型与公理化系统。具体是指在处理视频、音频、图像、图形、文字、数值、泛在网络行为性信息流等信息时，对所涉及的隐私信息进行描述、度量、评价和融合等操作，形成一套符号化、公式化且具有量化评价标准的隐私计算理论、算法及应用技术，支持多系统融合的隐私信息保护。隐私计算涵盖了信息搜集者、发布者和使用者在信息产生、感知、发布、传播、存储、处理、使用、销毁等全生命周期过程的所有计算操作，并包含支持海量用户、高并发、高效能隐私保护的系统设计理论与架构。隐私计算是泛在网络空间隐私信息保护的重要理论基础。

隐私计算涉及 6 个因素 (X,S,R,C,F,S)。X 为隐私信息集合，其概率分布的定义对于隐私度量紧密相关；S 为信息所有者集合；R 为信息接收者集合（信息接收者拥有其知识背景、兴趣点、主观感受和理解力等）；C 为隐私泄露收益损失比；F 为信息利用时的约束条件集合（包括含时间、空间、所用设备等环境条件）；S 为对隐私信息操作的集合，隐私感知、隐私保护、隐私分析、隐私信息的交换和二次传播、隐私信息更新等都可定义为对隐私信息集合的特定操作，将其抽象为符号化的描述，根据取值集合上定义的隐私度量可以定义隐私运算的规则，形成隐私计算的公理化体系。隐私计

算模型的核心是刻画隐私度量 I、隐私保护复杂性代价 E、隐私保护效果 G 以及隐私泄露收益损失比 C 4 个量之间的关系。

$$I=f(X,S,R,F,S) \tag{3}$$

$$E=h(X,S,R,F,S) \tag{4}$$

$$G=g(X,S,R,F,S) \tag{5}$$

$$C=e(I,E,G) \tag{6}$$

3.3　隐私计算的研究范畴

隐私信息的全生命周期如图 1 所示。

1) 隐私信息产生。个体在日常生活、使用互联网服务等会产生图片、位置、兴趣爱好、电话号码等各类文本、图像、语音、视频等隐私信息。这些信息可能被会主动或者被动的收集。

2) 隐私感知。从包含隐私的信息中构建隐私变量集合，或从变量集合中确定变量的取值或取值范围，产生隐私元数据，对隐私进行标记和编码，确定隐私变量的概率分布，从而对隐私变量中隐私度量的大小进行计算，为实施隐私保护提供支撑。概率分布的定义既涵盖客观定义，也考虑主体的主观因素，并且可能根据时空变化而变化，从而使隐私计算模型可以具备对主体、时间、空间三维演化的刻画能力。

3) 隐私保护。根据隐私感知得到的数据及其标记，选用相应隐私保护方法，包括密码学方法、信息隐藏方法和数据扰乱方法。密码学方法主要需要研究构造适用于隐私保护、与传统数据加解密不同的密钥管理机制、同态密码方案以及混淆方法等；信息隐藏/隐写的方法则可以用来保护元数据，将元数据以变化的形态来传输，对应的还原控制参数应该与信息本身分割存储和传输；数据处理方法则是去除不同隐私数据间的关联性、添加数据扰动、通过数据匿名化实现隐私保护（如 k-匿名，l-多样性，t-邻近性等）、防止聚类分析、众包计算、深度学习等大数据分析方法。此外，对得到的数据需判定和评价是否需要全标记、标记是否合理及所选用的隐私保护算法是否满足相应的保护需求，即需给出隐私保护算法的评价标准理论和方法。

4) 隐私发布。这个环节是隐私信息在公众网络中传播的隐私计算机制。隐私发布可以采用基于限制发布的隐私保护技术，例如，选择性地发布原始数据、不发布或者发布精度较低的敏感数据。

5) 隐私信息存储。该环节主要研究隐私保护之后的数据高效存储，使数据如何分类、组织、快速检索、判断不同方案的隐私保护信息的同源去重、同源同系统/同源不同系统的一致性维护。研究内容包括同质隐私信息去冗技术、支持隐私保护的重复数据删除技术、隐私感知的混合数据分割存取技术、隐私信息完整性校验机制等。此外，该环节还应考虑大数据存储的高效加密保护技术，以适应海量用户、高并发、多业务流、海量密钥随机交叉的调度应用。

6) 隐私信息的融合处理。在隐私数据的融合处理环节中，由于不同系统在隐私界定、度量方法、隐私保护需求等方面都存在差异，而且随着时间场景的变更，人们对隐私认知也在不断的变化，此外，隐私信息可能被进行二次转发、局部处理、隐私分割、延伸授权等，因此需设计一套协议和封装描述方法，可根据不同的隐私属性、场景、隐私信息等级来自适应地选择不同的隐私保护措施，充分发挥现有隐私保护技术（如数据加密、模糊、混淆等）的各自优势。

7) 隐私交换。在隐私数据交换环节，可能采用的隐私保护方式包括在不同信息系统的交换边界构造一个安全系统进行隐私保护方案的转换、基于隐私代理的跨网跨系统控制参数或约束条件的交换、隐私泄露的追踪溯源等方式。需要研究新型的代理重加密、防密钥泄漏、跨系统交换的访问控制以及追责等机制。此外，针对具有不同隐私保护能力的信息系统间交互隐私信息的场景，需考虑从低保护级别到高保护级别，是否需要提升隐私保护等级；以及从高到低是否需降低隐私保护等级等。

8) 隐私分析。隐私分析是隐私保护的逆过程。从施加隐私保护方案的数据中提取隐私信息取值

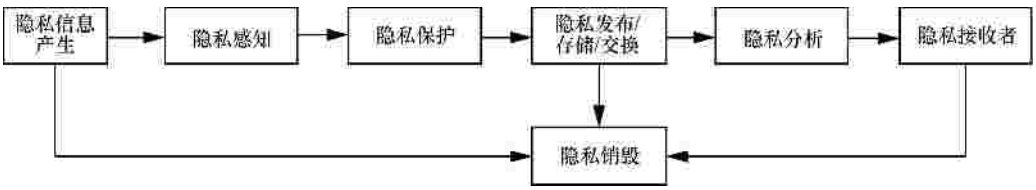

图 1　隐私信息的全生命周期

或确定其取值范围的过程。隐私分析实际上还受到隐私接收者知识背景、其所处环境和使用设备的影响，需要建立考虑这些因素的隐私分析计算模型。

9) 隐私销毁。在不再需要隐私信息，或隐私信息所有者希望终止隐私信息传播时，需要将隐私数据永远不可逆删除或销毁。为此，欧盟和美国已立法分别赋予用户“被遗忘权”和“橡皮”法律。从技术角度而言，实现这一权力需要研究可信删除，或称为确定性删除技术，以确保隐私信息的所有者、管理者和使用者都不可再恢复该信息。同时确保隐私保护的信息不能被隐私分析提取，并建立一套体系或机制，可通知关联系统，一旦数据被销毁，释放相应的存储空间。在当前泛在网络空间环境中，隐私信息的销毁难度非常大，极具挑战性。

4 隐私计算研究的发展趋势

4.1 隐私计算模型的建立与完善

目前，研究者提出的隐私保护技术仅是隐私计算模型的部分内容，但不足以涵盖隐私计算模型。为此，需要分析归纳泛在网络和大数据环境下信息服务演化规律，提炼隐私保护的需求。研究基于多维度的隐私定义、刻画及演化理论，构建隐私计算模型，利用信息论、博弈论、优化理论、计算复杂性理论等工具，给出隐私的量化定义，建立一整套隐私信息处理过程中隐私变换的描述和计算规则，揭示隐私度量、隐私泄露收益损失比、隐私保护与分析复杂性代价以及隐私保护效果之间的内在联系，为隐私保护技术提供一套科学的、体系化的理论工具。

4.2 隐私保护场景适应的密码理论与技术

针对不同的隐私保护场景，需要研究属性密码、同态密码、可重构轻量级密码、可搜索加密等前沿密码体制；针对明文大数据分析导致隐私泄露、大规模数据密态化影响分析效率的矛盾，需要研究密文数据查询计算、密文机器学习算法等密态数据处理技术，为隐私保护提供密码理论和技术支撑。

4.3 隐私控制机制与抗大数据分析的隐私保护技术

隐私信息在不同系统和不同用户间传播，不同系统和用户隐私保护能力差异不可避免，隐私也随时空、主体的不同不断演化，隐私保护强度与隐私保护对象也应具有场景适应的演化能力。应该加强研究大规模数据匿名和混淆等脱敏技术、抗大数据关联分析的数据动态发布和隐私保护效果评估技术，实现抗大数据分析的隐私保护。

4.4 基于信息隐藏的多媒体数据隐私保护

针对多媒体数据，研究转换保护对象的理论，抗机器智能理解的多媒体数据隐私保护方法、多媒体大数据环境下的隐写及分析方法、开发多媒体隐私保护的实用工具。

4.5 支持高并发的隐私保护架构设计理论

大数据巨规模、多样化、高增速等特征以及大数据应用的迅猛发展对隐私保护服务请求的用户容量、并发程度和能效优化提出了极高的要求。研究支持高并发的数据和隐私保护服务系统架构设计理论，设计“服务状态跨层跟随”的高性能数据保护和隐私保护服务计算架构，达到组件化、高效的服务队列处理、业务状态调度和数据管理，支持多算法/多密钥/多数据流的随机交叉处理，将是实现支持海量用户、能效优化的大数据隐私保护处理平台的一种有效的技术途径。

4.6 隐私保护的三维模型

社交网络、云计算、大数据、大搜索等新技术与服务模式的持续发展对隐私保护带来了新的挑战。用户在不同时间、地点需要采用各种不同的隐私保护方案，实现对来自不同网络和信息系统的海量数据的动态随机访问、更新、融合、迁移、删除等操作，以满足其个性化的服务需求。为此，亟待在如图 2 所示的隐私保护的三维模型的基础上研究不同算法或方案的融合机制和实施方案。

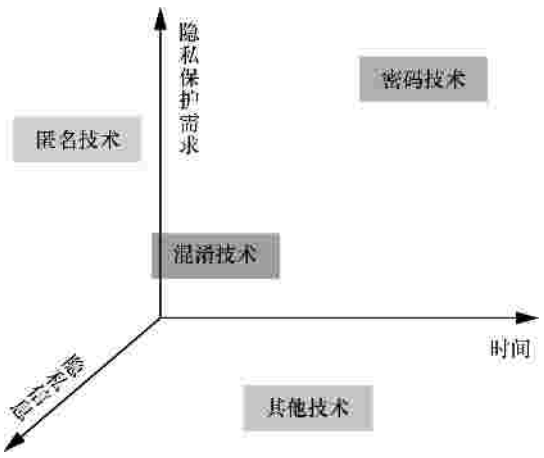

图 2 隐私保护的三维模型

5 结束语

在科技改变生活、网络引领未来的新时代，世界信息技术革命日新月异，信息化和经济全球化相互促进，互联网已经融入社会生活方方面面，深刻

地改变了人们的生产和生活方式。随着移动互联网、云计算和大数据技术的发展，数据的收集、共享、发布与分析会导致用户隐私信息的泄露，给用户带来巨大的损失。现有的各类隐私保护机制和方法并没有实现对隐私进行科学、系统和定量化地刻画，使隐私保护方法的研究缺乏理论指导，呈现碎片化的趋势。个人隐私会同时存在不同信息系统中，客观上这些不同信息系统虽然实现了不同等级的安全保护，但某一信息系统的泄露将导致其他信息系统的隐私保护的失效，因此如何确保存有用户隐私信息的所有信息系统达到同等安全级别成为一个难题。本文在对隐私保护研究现状进行分析的基础上，首次提出了隐私计算的概念，给出了隐私计算的范畴、内涵和框架，并对隐私计算研究的发展趋势进行了展望，期待能促进未来建立科学、系统的隐私保护理论与技术体系。

参考文献：

[1] CULNAN M J, ARMSTRONG P K. Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation[J]. Organization Science, 1999, 10(1): 104-115.

[2] DINEV T, HART P. Privacy concerns and internet use-a model of trade-off factors[C]//Academy of Management. c2003:1-6.

[3] LI H, SARATHY R, XU H. Understanding situational online information disclosure as a privacy calculus[J]. Journal of Computer Information Systems, 2010, 51(1): 62-71.

[4] KEHR F, KOWATSCH T, WENTZEL D, et al. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus[J]. Information Systems Jour l, 2015, 25(6): 607-635.

[5] MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. *l*-diversity: privacy beyond *k*-anonymity[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1): 3.

[6] AGRAWAL D, AGGARWAL C C. On the design and quantification of privacy preserving data mining algorithms[C]//The 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. ACM, c2001: 247-255.

[7] LIU K, KARGUPTA H, RYAN J. Random projection-based multiplicative data perturbation for privacy preserving distrib data mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(1): 92-106.

[8] OLIVEIRA S R M, ZAIANE O R . Privacy preserving clustering by data transformation[C]//The 18th Brazilian Symposium on Databases. c2003:304-318.

[9] OLIVEIRA S R M, ZAIANE O R . Privacy preserving clustering by object similarity-based representation and dimensionality reduction transformation[C]//The Workshop on Privacy and Security Aspects of Data Mining. c2004:21-30.

[10] OLIVEIRA S R M, ZAIANE O R. Privacy preserving frequent itemset mining[C]//The IEEE International Conference on Privacy, Security and Data Mining-Volume 14, Australian Computer Society. c2002: 43-54.

[11] OLIVEIRA S R M. Protecting sensitive knowledge by data sanitization[C]//IEEE. c2003: 613-616.

[12] SAYGIN Y, VERYKIOS V S, ELMAGARMID A K. Privacy preserving association rule mining[C]//Research Issues in Data Engineering: Engineering E-Commerce/E-Business Systems. c2002: 151-158.

[13] CHANG L W, MOSKOWITZ I S. An integrated framework for database privacy protection[M]. Springer US, 2002.

[14] SWEENEY L. *k*-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge Based Systems, 2002, 10(5): 557-570.

[15] LI N H, LI T C, VENKATASUBRAMANIAN S. *t*-closeness: privacy beyond *k*-anonymity and *l*-diversity[C]//IEEE 23rd International Conference on Data Engineering. Istanbul, c2007:106-115.

[16] ZHANG Q, KOUDAS N, SRIVASTAVA D, et al. Aggregate query answering on anonymized tables[C]//IEEE 23rd International Conference on Data Engineering. Istanbul, c2007:116-125.

[17] FANG Y, ASHRAFI M, NG S. Privacy beyond single sensitive attribute[C]//22nd International Conference. DEXA, c2011:187-201.

[18] WANG K, FUNG B C M. Anonymizing sequential releases[C]//KDD. c2006:414-423.

[19] WONG R C, LI J Y, FU A W, et al. (*a*, *k*)-anonymity: an enhanced *k*-anonymity model for privacy preserving data publishing[C]//The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, ACM, c2006:754-759.

[20] TRUTA T M, BINDU V. Privacy protection: *p*-sensitive *k*-anonymity property[C]//The Workshop on Privacy Data Management (PDM). c2006:94.

[21] LI J X, TAO Y F, XIAO X K. Preservation of proximity privacy in publishing numerical sensitive data[C]//The 2008 ACM SIGMOD International Conference on Management of Data. New York, ACM, c2008:473-486.

[22] LI N H, LI T C, VENKATASUBRAMANIAN S. Closeness: a new privacy measure for data publishing[J]. IEEE Trans Knowl Data Eng, 2010, 22:953-956.

[23] CASAS-ROMA J, HERRERA-JOANCOMARTI J, TORRA V. A summary of *k*-degree anonymous methods for privacy-preserving on networks[J]. Advanced Research in Data Privacy, 2015,567:231-250.

[24] YAO C, WANG X S, JAJODIA S. Checking for *k*-anonymity violation by views[C]//The 31st Conference on Very Large Data Bases (VLDB). c2005:910-921.

[25] WANG K, FUNG B C M. Anonymizing sequential releases[C]//The 12th ACM SIGKDD Conference. ACM, New York. c2006.

[26] BYUN J W, SOHN Y, BERTINO E, et al. Secure anonymization for incremental datasets[C]//The VLDB Workshop on Secure Data Management (SDM). c2006.

[27] FUNG B C M, WANG K, FU A W C, et al. Anonymity for continuous data publishing[C]//The 11th International Conference on Extending Database Technology (EDBT). ACM, New York, c2008:264-275.

[28] JIANG W, CLIFTON C. A secure distributed framework for achieving *k*-anonymity[C]//Very Large Data Bases. c2006:316-333.

[29] GOIYCZKA S, XIONG L, FUNG B C M. *m*-privacy for collaborative

data publishing[C]//International Conference on Collaborative Computing: Networking, Applications and Worksharing. IEEE, c2011:1-10.

[30] GAL T S, CHEN Z Y, GANGOPADHYAY A. A privacy protection model for patient data with multiple sensitive attributes[J]. Int J Inf Secur Priv, 2008,2: 28-44.

[31] DAS D, BHATTACHARYYA D K. Decomposition+: improving l-diversity for multiple sensitive attributes[C]//Advances in Computer Science and Information Technology, Computer Science and Engineering. c2012:403-412.

[32] AGRAWAL R, SRIKANT R. Privaey-preserving data mining[C]// ACM SIGMOD Reeord. c2000:439-450.

[33] AGGARWAL C C, YU P S. A general survey of privacy-preserving data mining models and algorithms[M]. Privacy-Preserving Data Mining. Springer US, 2008:11-52.

[34] LI L, KANTARCIOGLU M, THURAISINGHAM B. The applicability of the perturbation based privacy preserving data mini g for real-world data[J]. Data & Knowledge Engineering, 2008, 65(1):5-21.

[35] WITTEN I H, FRANK E. Data mining: practical machine learning tools and techniques[M]. Morgan Kaufmann, 2005.

[36] CLIFTON C, KANTARCIOGLU M, VAIDYA J, et al. Tools for privacy preserving distributed data mining[J]. ACM Sigkdd Explorations Newsletter, 2002, 4(2): 28-34.

[37] JAGANNATHAN G, PILLAIPAKKAMNATT K, WRIGHT R N. A new privacy-preserving distributed *k*-clustering algorithm[C]//SDM. c2006:494-498.

[38] RIVEST R L, ADLEMAN L, DERTOUZOS M L. On data banks and privacy homomorphisms[J]. Foundations of Secure Computation, 1978:169-179.

[39] GENTRY C. Fully homomorphic encryption using ideal lattices[J]. The Annual Acm Symposium on Theory of Computing, 2009, : 169-178.

[40] ATALLAH M J, PANTAZOPOULOS K N, RICE J R, et al. Secure outsourcing of scientific computations[J]. Advances in Computers, 2002, 54(01):215-272.

[41] ATALLAH M J, LI J. Secure outsourcing of sequence comparisons[J]. International Journal of Information Security, 2005, 4(4):277-287.

[42] GENNARO R, GENTRY C, PARNO B. Non-interactive verifiable computing: outsourcing computation to untrusted workers[J]. Lecture Notes in Computer Science, 2010, 6223:465-482.

[43] CHAUM D, PEDERSEN T P. Wallet databases with observers[J]. Lecture Notes in Computer Science, 1994, 740:89-105.

[44] CURTMOLA R, GARAY J, KAMARA S, et al. Searchable symmetric encryption: improved definitions and efficient constructions[C]//The 13th ACM Conference on Computer and Communications Security. ACM, c2006: 79-88.

[45] BONEH D, DI CRESCENZO G, OSTROVSKY R, et al. Public key encryption with keyword search[C]//Advances in Cryptology- Eurocrypt 2004. Springer Berlin Heidelberg, c2004:506-522.

[46] DAN B, WATERS B. Conjunctive, subset, and range queries on encrypted data[C]//The Theory of Cryptography Conference. c2006: 535-554.

[47] YAU W C, PHAN R C W, HENG S H, et al. Proxy re-encryption with keyword search: new definitions and algorithms[M]//Security Technology, Disaster Recovery and Business Continuity. Spr ger Berlin Heidelberg, 2010: 149-160.

[48] SHAO J, CAO Z, LIANG X, et al. Proxy re-encryption with keyword search[J]. Information Sciences, 2010, 180(13): 2576-2587.

[49] FANG L, SUSILO W, GE C, et al. A secure channel free public key encryption with keyword search scheme without random oracle[M]. Cryptology and Network Security. Springer Berlin Heidelberg, 2009: 248-258.

[50] CAO N, WANG C, LI M, et al. Privacy-preserving multi-keyword ranked search over encrypted cloud data[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(1): 222-233.

[51] POPA R A, REDFIELD C, ZELDOVICH N, et al. CryptDB: protecting confidentiality with encrypted query processing[C]//The Twenty-Third ACM Symposium on Operating Systems Principles. ACM, c2011: 85-100.

[52] SACHNEV V, KIM H J, NAM J, et al. Reversible watermarking algorithm using sorting and prediction[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2009, 19(7): 989-999.

[53] LUO L, CHEN Z, CHEN M, et al. Reversible image watermarking using interpolation technique[J]. Information Forensics and Security, IEEE Transactions on, 2010, 5(1): 187-193.

[54] TIAN J. Reversible data embedding using a difference expansion[J]. IEEE Trans Circuits Syst Video Techn, 2003, 13(8): 890-896.

[55] MA K L, ZHANG W, ZHAO X, et al. Reversible data hiding in encrypted images by reserving room before encryption[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(3): 553-562.

[56] PARUCHURI J K, CHEUNG S C S, HAIL M W. Video data hiding for managing privacy information in surveillance syste J]. EURASIP Journal on Information Security, 2009: 7.

[57] HARTUNG F, GIROD B. Digital watermarking of MPEG-2 coded video in the bitstream domain[C]//The 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97). c1997:2621.

[58] PARVIAINEN R, PARNES P. Large scale distributed watermarking of multicast media through encryption[M]. Communications and Multimedia Security Issues of the New Century. Springer US, 2 9-158.

[59] ZHAO H V, LIU K J. Fingerprint multicast in secure video streaming[J]. IEEE Transactions on Image Processing, 2006, 15(1): 12-29.

[60] KUNDUR D, KARTHIK K. Video fingerprinting and encryption principles for digital rights management[J]. Proceedings of the IEEE, 2004, 92(6): 918-932.

[61] ADELSBACH A, HUBER U, SADEGHI A R. Finger casting-joint fingerprinting and decryption of broadcast messages[M]. Transactions on Data Hiding and Multimedia Security Ⅱ. Springer Berlin Heidelberg, 2007: 1-34.

[62] CELIK M U, LEMMA A N, KATZENBEISSER, et al. Secure embedding of spread spectrum watermarks using look-up-tables[C]//Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference. IEEE, c2007:153-156.

[63] LEMMA A, KATZENBEISSER S, CELIK M, et al. Secure watermark embedding through partial encryption[M]. Digital Watermarking. Springer Berlin Heidelberg, 2006: 433-445.

[64] CELIK M U, LEMMA A N, KATZENBEISSER S, et al. Lookup-table-based secure client-side embedding for spread-spectrum watermarks[J]. IEEE Transactions on Information Forensics & Security, 2008, 3(3):475-487.

[65] ADELSBACH A, HUBER U, SADEGHI A R. Fingercasting–joint

fingerprinting and decryption of broadcast messages[M]. Transactions on Data Hiding and Multimedia Security II. Springer Be lin Heidelberg, 2007:1-34.

[66] http:www.speedproject.eu/[EB/OL].

[67] KER A D, BAS P, BÖHME R, et al. Moving steganography and steganalysis from the laboratory into the real world[C]//ACM Workshop on Information Hiding & Multimedia Security. c2013:45-58.

[68] FILLER T, JUDAS J, FRIDRICH J. Minimizing additive distortion in steganography using syndrome-trellis codes[J]. IEEE Transactions on Information Forensics & Security, 2011, 6(3):920-935.

[69] HOLUB V, FRIDRICH J. Digital image steganography using universal distortion[C]//The first ACM Workshop on Information Hiding and Multimedia Security. ACM, c2013: 59-68.

[70] FRIDRICH J J, KODOVSKÝ J. Multivariate Gaussian model for designing additive distortion for steganography[C]//ICASSP. c2013: 2949-2953.

[71] LI B, WANG M, LI X, et al. A strategy of clustering modification directions in spatial image steganography[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(9): 1905-1917.

[72] DENEMARK, T, FRIDRICH J. Improving steganographic security by synchronizing the selection channel[C]//ACM Workshop on Information Hiding and Multimedia Security. ACM, c2015:5-14.

[73] XIONG G, PING X, ZHANG T, et al. Image textural features for steganalysis of spatial domain steganography[J]. Journal of Electronic Imaging, 2012, 21(3): 033015-1-033015-15.

[74] KODOVSKÝ J, FRIDRICH J, HOLUB V. Ensemble classi?ers for steganalysis of digital media[J]. IEEE Trans On Information Forensics and Security, 2012, 7(2): 432-444.

[75] TANG W, LI H, LUO W, et al. Adaptive steganalysis against WOW embedding algorithm[C]//The 2nd ACM Workshop on Information Hiding and Multimedia Security. ACM, c2014: 91-96.

[76] DENEMARK T, SEDIGHI V, HOLUB V, et al. Selection- channel-aware rich model for steganalysis of digital images[C]// Information Forensics and Security (WIFS), 2014 IEEE Internat nal Workshop. IEEE, c2014: 48-53.

[77] CHAUM D. Untraceable electronic mail, return addresses and digital pseudonyms[J]. Communications of the ACM, 1981, 24(2):84-90.

[78] DINGLEDINE R, MATHEWSON N, SYVERSON P. Tor: the second-generation onion router[J]. Journal of the Franklin Institute, 2004, 239(2): 135-139.

[79] The Tor project[EB/OL]. https://www:torproject:org/,2003.

[80] ZHOU Y, YANG Q, YANG B, et al. A tor anonymous communication system with security enhancements[J]. Journal of Computer Research and Development, 2014, 51(7):1538-1546

[81] MURDOCH S J，DANEZIS G. Low-cost traffic analysis of Tor[J]. IEEE Symposium on Security and Privacy, 2005, 47(3): 183-195.

[82] BRIAN N L, MICHEAL K R, WANG C. Timing attacks in low-latency mix systems: extended, abstract[C]//Financial Cryptography. Berlin: Springer, c2004: 251-265.

[83] FEAMSTER N, DINGLEDINE R. Location diversity in anonymity networks[C]//The Workshop on Privacy in the Electronic Society. ACM, c2004:66-76.

[84] JANSEN R, TSCHORSCH F, JOHNSON A, et al. The sniper attack: anonymously deanonymizing and disabling the tor network[C]// Network and Distributed System Security Symposium. c2014.

[85] CHAUM D, JAVANI F, KATE A, et al. cMix: anonymization by high-performance scalable mixing[C]// 25th USENIX Security Symposium. c2016.

[86] http://www.aqniu.com/news-views/13063.html[EB/OL].

[87] http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm[EB/OL].

[88] http://www.aicpa.org/interestareas/informationtechnology/reso s/privacy/generallyacceptedprivacyprinciples/pages/default.aspx[EB/OL].

作者简介：

李凤华（1966-），男，湖北浠水人，博士，中国科学院信息工程研究所副总工、研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。

李晖（1968-），男，河南灵宝人，博士，西安电子科技大学教授、博士生导师，主要研究方向为密码学、无线网络安全、云计算安全、信息论与编码理论。

贾焰（1960-），女，四川成都人，博士，国防科学技术大学教授，主要研究方向为大数据、网络信息安全和社交网络。

俞能海（1964-），男，安徽无为人，博士，中国科学技术大学教授、博士生导师，主要研究方向为图像处理与媒体内容安全、互联网信息检索与数据挖掘。

翁健（1976-），男，广东茂名人，博士，暨南大学教授、博士生导师，主要研究方向为密码学与信息安全。