软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(8):1889−1890 [doi: 10.13328/j.cnki.jos.005097]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

数据开放与隐私管理专题前言*

孟小峰[1], 林东岱[2,3]

[1](中国人民大学 信息学院,北京 100872)

[2](中国科学院 信息工程研究所,北京 100093)

[3](信息安全国家重点实验室(中国科学院 信息工程研究所),北京 100093)

通讯作者: 孟小峰, E-mail: xfmeng@ruc.edu.cn

中文引用格式: 孟小峰,林东岱.数据开放与隐私管理专题前言.软件学报,2016,27(8):1889−1890. http://www.jos.org.cn/1000-9825/5097.htm

随着大数据时代的到来,数据通过开放共享将产生更大的价值,不但可以提高社会运行效率,更能激发巨大的商业价值.例如,社交网络记录了全球数十亿人每天发布的言行,推荐系统可以通过分析人们的言行了解人们对各种产品个性化的偏好并做出准确的广告推荐;GPS 每时每刻记录着全球数十亿使用者的位置,政府部门可以通过分析人们在城市中的轨迹了解道路规划是否合理.根据麦肯锡公司分析,如果教育、交通、商业、电力、石油业天然气、卫生保健以及金融业为 7 大行业能够公开行业内的数据,每年基于数据开放的大数据应用将创造 3 万亿美元的经济社会效益.

然而,数据开放无法避开伴随而来的更加严峻的隐私问题.随着未来数据开放程度的加深,在单一开放数据中不敏感的信息在多个行业同时开放数据后可能变成敏感信息.例如,当市政部门公开包含通信方式的房屋信息,通信公司公开包含通信方式的用户信息以后,个人的住址可能通过这两个公开数据相互参照被泄露.因此,为了正确地进行数据开放,隐私管理需要扩展当前的保护范畴,定义新的隐私度量机制并针对大数据开放场景设计新的保护方法.新的数据公开需求从隐私管理理论和应用角度等多个方面带来了挑战性的问题.

为了反映当前隐私保护的研究现状及最新研究成果,展示数据开放与隐私管理面临的理论和技术上的新挑战,揭示数据开放与隐私管理的研究热点及研究方向,《软件学报》和孟小峰教授和林东岱研究员两位特约编辑共同策划和组织了数据开放与隐私管理专题.本期《软件学报》“数据开放与隐私管理专题”组稿通过专题公开征文获得 16 篇投稿,论文分别在多个方面阐述了数据开放和隐私管理领域具有重要意义的研究成果.本专题的审稿严格按照期刊审稿要求进行,特约编辑先后邀请了 24 余位相关领域的专家参与评审,每篇论文邀请至少 2 位专家进行评审,历经初审、复审、终审各阶段,整个流程历经半年,最终从中遴选出 3 篇高质量的论文入选本专题.这 3 篇论文分别对数据开放中一般性隐私理论,社交网络图数据开放和时空数据开放这两种重要大数据应用场景下的隐私问题进行了探讨,在一定程度上反映了当前国内各研究单位在数据开放和隐私管理领域的主要研究方向.

数据开放中的隐私是什么?它如何度量?这是数据开放场景下隐私管理的基本问题.《隐私保护的信息熵模型及其度量方法》总结了现有隐私管理技术中对隐私的度量方式和面对数据开放这一场景下更加通用的基于香农信息论的信息熵模型.具体来说,这篇文章从基本信息熵模型开始,逐渐扩展到含敌手攻击的隐私保护信息熵模型、带主观感受的信息熵模型和多隐私信源的信息熵模型,为数据开放程度加深后攻击者利用具有多种背景知识和利用多来源数据相互参照的攻击方法下的隐私保护方法的强度和敌手攻击能力提供了量化测评.

社交网络数据作为数据开放的一类重要数据,其隐私信息的安全性是当前隐私管理研究中的热点问题.

* 收稿时间: 2016-06-17

© 中国科学院软件研究所 http://www.jos.org.cn

《一种保持结点可达性的高效社会网络图匿名算法》针对图数据可用性的重要指标可达性查询精度的保护,提出了可达性保持图匿名化(RPA)算法.其基本思想是将结点进行分组并采取贪心策略进行匿名,从而减少匿名过程中的可达性信息损失.为保证 RPA 算法的实用性,针对其执行效率进行优化,首先提出采用可达区间来高效地评估边添加操作所导致的匿名损失;其次,通过构建候选邻居索引,进一步加速 RPA 算法对每个结点的匿名过程.基于真实社会网络数据的实验结果表明了 RPA 算法的高执行效率,同时验证了生成匿名图在可达性查询方面的高精度.

时空数据开放也是隐私管理的重点研究对象.其中,多数据源开放引起的不同数据间相互参照的情况是隐私管理需要解决的重要问题.《时空数据发布中的隐式隐私保护》首次定义并研究了时空数据相互参照而引起的隐私保护问题,为数据开放大场景下的隐私管理提供了一种思路.具体来说,本文提出了基于“发现-消除”的隐私保护框架,基于前缀过滤的嵌套循环算法用于发现隐式收集的时空数据中可能泄露用户隐私的记录,并基于频繁移动对象的假数据添加方法消除这些记录.此外,分别提出了更高效的反先验算法和基于图的假数据添加算法.最后,在若干真实数据集上对提出的算法进行的充分实验证实了这些算法有较高的保护效果和性能.

承蒙各位作者、审稿专家和编辑部等方面的全力支持,本专题得以顺利出版.目前数据开放和隐私管理涉及领域十分广泛,这给审稿人及特约编辑的审稿、选稿带来巨大挑战.由于投稿数量大、主题广泛、时间安排紧张、专题容量有限等原因,本专题仅选择了部分有代表性的研究工作进行刊登,无法全面体现隐私保护领域所有的最新研究工作.部分优秀稿件无法列入发表,敬请谅解.

我们要特别感谢《软件学报》编委会和编辑部,从专题的立项到征稿启示的发布,从审稿专家的邀请到评审意见的汇总,以及最后的定稿、修改、出版,他们都付出了辛勤的汗水.本专题的出版期望能够给广大研究人员带来启发和帮助.在审稿过程中难免出现不尽如人意之处,希望各位作者和读者包容和谅解,希望同行不吝批评指正.最后,衷心感谢各位作者、审稿专家和编辑部的辛勤工作!

孟小峰(1964－),男,河北邯郸人,博士,教授,博士生导师,CCF 会士,主要研究领域为 Web 数据管理,移动数据管理,XML 数据管理,云数据管理.

林东岱(1964－),男,博士,研究员,信息安全国家重点实验室主任,中国密码学会常务理事,中国保密协会理事,主要研究领域为密码学与安全协议,网络与系统安全,伪随机序列,大规模方程求解与代数分析.

© 中国科学院软件研究所 http://www.jos.org.cn