

社交网络中的隐私保护研究综述

姚瑞欣, 李晖, 曹进

(西安电子科技大学网络与信息安全学院综合业务网理论与关键技术国家重点实验室, 陕西 西安 710071)

摘要: 随着信息技术的发展, 社交网络逐渐成为人们沟通的主要方式, 而敏感信息暴露在开放的社交网络中所导致的多种隐私信息泄露问题也引起了人们的日益关注。首先, 介绍了社交网络及隐私的相关概念; 其次, 对当前社交网络中的隐私保护所采用的主要方法: 匿名技术和访问控制技术进行了详细的介绍、分析和讨论; 最后, 对现有方法存在的问题和挑战进行了分析, 并给出了一些潜在的研究热点, 为未来研究工作指明方向。

关键词: 社交网络; 隐私保护; 匿名; 访问控制

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00036

Overview of privacy preserving in social network

YAO Rui-xin, LI Hui, CAO Jin

(State Key Laboratory of Integrated Service Network, School of Cyber Engineering, Xidian University, Xi'an 710071, China)

Abstract: Social network has gradually become the main way in communication with the development of information technology and the issue of privacy preserving in social network has also obtained more and more attention due to the interactive information exposed in the open social networks. Firstly, an overview of the concept of social network and privacy was given. Secondly, the current mainly two methods: anonymity mechanism and access control technique employed for privacy preservation in social networks were reviewed and discussed. Finally, the vulnerabilities and challenges of the current mechanism were analyzed and the potential research issues for the future research works were showed.

Key words: social network, privacy preserving, anonymity, access control

1 引言

随着互联网及信息技术的快速发展, 社交网络成为人们交流沟通的重要方式, 如美国的 MySpace、Facebook, 日本的 Mixi, 英国的 Bebo 等, 国内有 QQ、人人网、微信等^[1]。在社交网络中, 人们发布真实的个人信息以达成交友目的, 因此备受欢迎, 但分享日志、上传照片等行为也

使越来越多的信息暴露在社交网络中, 信息的公开化引起了人们对隐私保护的关注。各社交平台也积极采取多种措施以保护隐私: Facebook 为用户提供隐私设置功能, 由用户决定信息可由哪些用户及开发者获取; Twitter^[2]同 Facebook 一样, 为用户提供个性化的隐私保护设置, 此外, Twitter 自 2012 年在 Firefox 浏览器上提供“Do Not Track”隐私保护选项以阻止网站在计算机上留下

收稿日期: 2016-02-17; 修回日期: 2016-03-02。通信作者: 姚瑞欣, nice127@163.com

基金项目: 国家自然科学基金—广东联合基金资助项目 (No.U1401251)

Foundation Item: The National Natural Science Foundation of China—Guangdong Jointly Funded Project (No.U1401251)

Cookie^[3]; QQ 空间有细分的权限管理列表,可分别设置空间、说说、相册的可见范围,也可设置说说等的评论权限,通过这些细粒度访问控制实现隐私保护。

本文对社交网络中的隐私保护机制进行了研究,主要集中在匿名和访问控制两方面。前者是从分析人员的视角,后者是从社交网络成员自身的视角研究的。匿名方面,研究者针对属性暴露问题已取得了一些研究成果,但研究主要集中于预测属性而非保护属性,国外研究关注点主要放在对用户隐私的评估和人性化订制个人隐私策略上,而国内研究人员更关注信任算法、匿名、节点发现算法等^[4-6]的研究。此外,OSN(online social network)用户即使进行了访问控制,攻击者也会从其他途径获取用户的信息^[7]。

在 OSN 访问控制模型的研究方面,传统的基于 Web 的访问控制模型已经不能适用于 OSN 这种复杂环境中的隐私保护需求。国内外多个研究小组均提出了不同的访问控制模型,为 OSN 的快速发展提供兼顾隐私及效率的隐私保护策略,国内主要集中在信任值计算、路径搜索等算法以及决策机制的研究上以实现商业应用及舆情监控的作用,目前,并没有成形的,如传统访问控制、强制访问控制(MAC, mandatory access control)、自主访问控制(DAC, discretionary access control)的 OSN 访问控制模型。

由于当前国内外均无 OSN 匿名及访问控制模型方面系统性的剖析和总结,因此,本文从核心思想、基本技术及研究成果等方面对匿名及访问控制方法进行了研究,并对各自存在的问题和挑战进行了分析并给出了一些未来的研究热点。

2 社交网络及隐私相关概念

2.1 社交网络及其安全问题

社交网络作为 Web 2.0 时代新兴的网络现象并没有一个被广泛认可的定义。Schneider 等^[8]提出的 OSN 涵盖了用户和资源 2 个方面;Boyd^[9]提出的 SNS(social network site)强调用户在 OSN 中的主导地位;Adamic^[10]提出的 SNS(social networking service),着重强调社交网络构建过程和服务性质,这 3 种定义从不同侧面定义了社交

网络。

本文中讨论的 OSN 更接近 Schneider 的定义,但范围扩展为用户、资源及活动,其功能主要包括^[9]:网络功能,即创建、更新公开或半公开的用户信息及好友列表;数据功能,即维护个人信息及提供在线聊天、社区、视频及音乐共享、评论等功能,也包含通过应用程序编程接口(API, application programming interface)开发应用插件等;访问控制功能,即用户自定义其隐私设置以保护个人隐私。

谈及 OSN 中的安全问题,由于其平台的开放性,用户多且用户数据量巨大,用户填写真实信息比例大,因此,有着不同于传统 Web 的安全问题^[1]。

1) 传统的攻击方式:传统攻击在社交网络这一平台同样可以实施,如 XSS 蠕虫^[11],而且结合了 Ajax 等技术之后其破坏力更大。

2) 网络架构攻击:例如, Sybil Attacks^[12]、分布式拒绝服务攻击,指由于社交网络本身结构及模型带来的安全问题。

3) 垃圾信息:主要包含广告信息^[13]和钓鱼网址^[14]。

4) 隐私泄露:一方面是社交网络提供商及第三方应用滥用信息^[15-17];另一方面是信息链接^[18],如未授权第三方从不同途径中整合数据以推断一个用户的身份或者行为。隐私泄露的实际危险在于用户无法完全控制自身信息的传播,需通过访问控制实现用户授权给他的朋友访问自身数据,并对提供商及其他未授权实体隐藏数据^[19,20],也可以对数据进行加密以实现隐私保护。

2.2 隐私及其保护

隐私的概念在很多领域(如哲学、心理学、社会学)已被研究多年,但一直缺乏明确地符合时代发展又经得起实践检验的定义^[21]。在某种意义上,隐私随生活经验而变化,依赖于特殊的情境^[22](如时间、地点、职业、文化、理由),因此,它是灵活的、动态的,很难得出一个通用的隐私概念。

隐私保护主要包含敏感知识的保护和敏感数据的保护^[23]。个人的隐私属性,如工资、家庭住址、工作、疾病史的保护属于敏感知识保护;数

据中的潜在关联规则、序列模式等信息的保护属于敏感数据保护。隐私保护技术的关注点: 一是数据以何种形式发布可防止隐私泄露, 依靠传统的访问控制技术和加密技术即可做到这点; 二是保证发布的数据有足够的可用性, 这要求不能切断隐私数据的访问通道, 也不需对数据进行加解密, 而是破坏数据项与个体间的对应关系, 或者降低这种对应的概率。

社交网络中用户的隐私保护研究主要集中于匿名和访问控制 2 个方面^[24]。匿名在第 3 节中介绍, 访问控制在第 4 节中介绍。

3 匿名方法的研究

近年来, 研究者们针对关系数据库中数据发布的匿名问题进行了深入的研究, 随着社交网络的发展, 其隐私保护问题成为了近年来的研究热点, 关系数据库中的匿名模型及算法被尝试用于社交网络, 本文首先介绍关系数据库中的匿名方法的研究成果, 随后对社交网络中的匿名方法研究进展进行介绍, 最后给出数据经匿名后的可用性度量方法。

3.1 关系数据库中的隐私保护

研究关系数据库中的匿名保护方法, 首先需认识数据的匿名过程, 包括: 1) 明确需保护的数据; 2) 对攻击者具有的背景知识及其攻击方式进行建模; 3) 处理数据; 4) 定义数据可用性。需要匿名保护的数据主要是指属性信息, 在关系数据库中, 属性划分为 3 类: 标识符 (ID, identifier)、准标识符 (QI, quasi-identifier)、隐私属性 (SA, sensitive attributes)。标识符指姓名、身份证号等能够准确标识个人身份的属性; 准标识符可以是一个属性或是多个属性的集合, 指不能明确标识个体但可以与其他表进行链接以标识个体的属性; 隐私属性指能使攻击者将信息与个体相对应的信息。

对属性信息匿名之后, 攻击者能否成功解匿名取决于攻击者的背景知识强度及节点间的结构相似性^[25], 针对攻击者背景知识, Zhou 等^[26]列举了节点属性、节点的度、目标个体的邻居结构、目标个体间的链接关系等背景知识; Liu 等^[27]以目标节点的度作为背景知识来窃取个体隐私;

Backstrom 等^[28]以目标个体的结构信息作为背景知识; Wondracek 等^[29]依据 OSN 中的群成员信息来定位目标个体; Cordella 等^[30]对以图结构为背景知识的目标个体进行精确子图匹配或同构判定; Zheleva 和 Liu 等^[31,32]在其研究中将各种类型的背景知识模型化; 前面的攻击者均采用的是被动攻击, 即攻击者利用网络子图的结构特殊性来将个体或个体间关系对应到节点或边, 也可采用主动攻击, 即攻击者创建新的用户以攻击对象, 如 Zou 等^[33]采用嵌入的子图作为攻击者的背景知识。

在完成了匿名的前两步: 明确需保护的数据, 明确攻击者背景知识及对背景知识建模之后, 接下来对数据进行匿名处理, 在本节中, 本文首先给出了关系数据库中的匿名方法, 其次根据不同的攻击模型对匿名方法进行分类说明, 最后对匿名度量方法进行了介绍。

3.1.1 匿名方法

关系数据库中的匿名方法主要包括以下 5 类。

1) 泛化: 包括全局泛化和局部泛化, 是将某一属性值用更概括的属性值来替代。全局泛化^[34,35]指所有的属性根据泛化树泛化到一个层次; 局部泛化指采用不同局部泛化算法实现匿名化。Xiao 等^[36]提出了一种个性化匿名方法, 指允许用户定义自身隐私属性的泛化层次, 很好地满足了特殊需求。聚类是一种特殊的泛化, 它将表中的 n 条记录划分至 m 个不同聚类, 每个聚类中的点数不少于 k 个。Brand^[37]和 Fuller^[38]都利用聚类实现了 k -anonymity。

2) 抑制: 指用特殊符号代替现有属性以使现有属性值更为模糊的匿名方法, 如将手机号码写作 159****9468 以实现匿名。Meyerson 等^[39]证明, 采用泛化和抑制方法实现 k -anonymity 的最优解是一个 NP (non-deterministic polynomial) 难问题, 大多研究均在寻找近似解。

3) Anatomization: Anatomization^[40]不改变 QI 属性值和隐私属性值, 而是将两者分开至 2 个独立的表中, 这样, 虽然数据不发生改变, 但原有数据挖掘方法将不再适用, 而且 Anatomization 发布连续数据的效果不佳。

4) 数据扰动: 数据扰动是通过添加噪音^[38,41]、数据置换^[42,43]、人工数据合成等方法对原始数据进

行一定的修改,但保留原始数据的统计信息^[44,45]。添加噪音用于数值型隐私数据,是用 $s+r$ 代替原始数据 s , 其中, r 是符合某一分布的随机值; 数据置换是指交换记录的隐私属性值; 人工数据合成依据现有数据构建一个统计模型, 并从模型中抽取样点来构成合成数据以代替原始数据。

5) 贪心方法: 在关系数据库匿名中使用广泛, Zhou 等^[46]提出了实现 k -neighborhood 匿名的贪心算法, 并详述了算法的步骤; Cormode 等^[47]针对无向无权重的单边二分图, 依据 k -anonymity 和贪心算法提出 (k, s) -分组匿名, 该匿名方法可有效抵抗静态攻击及链接攻击。

3.1.2 攻击类型与匿名实现模型

不同的攻击就会对应有不同的匿名模型, 现有的攻击类型如下。

记录链接攻击: 一个准标识符值所对应的表中若干记录构成一个等价组, 若一个体具有某准标识符取值, 通过链接对应到某条记录, 就能推算出该个体具有某属性的概率。Sweeney 等^[48]提出 k -anonymity 模型: 若某一记录具有某一 QI 值, 应至少还有 $k-1$ 条记录具有此 QI 值, 使链接攻击的可能性为 $\frac{1}{k}$ 。Wang 等^[49]提出 (X, Y) -anonymity 模型, X, Y 为 2 个不相交属性集, X 中每一值至少对应 Y 中 k 个不同值, k -anonymity 是其特例。

属性链接攻击: 当一等价组内属性过于单一, 攻击者便可根据记录所处等价组来推测其隐私属性。Machanavajjhala 等^[50]提出 l -diversity 模型, 要求每一 QI-group 至少包含 l 个不同的敏感属性, 利用多样性避免此类攻击。Wong 等^[51]提出了 (a, k) -anonymity 模型, 要求每个 QI 属性至少要与 k 条记录相对应, 从 QI 属性判断隐私属性时置信度不大于 a 。Li 等^[52]提出 t -closeness 模型, 指等价类中敏感特性的分布以及整个表中特性的分布距离不超过阈值 t 。

表链接攻击: 表链接攻击指攻击者推测出记录是否在表中的攻击, Nergiz 等^[53]针对此类问题提出了 δ -presence 模型。

3.1.3 匿名度量与评价

度量信息损耗以验证信息可用性, 是算法执行过程中决策的重要依据, 下面列举 6 种可用的信息损耗度量方法: 基于泛化树的度量^[54]、基于

信息熵的度量^[55]、LM (loss metric)^[56]、CM (classification metric)^[56]、DM (discernibility metric)^[57]以及 AM (ambiguity metric)^[58]。

3.2 社交网络中的隐私保护

社交网络数据的隐私保护匿名步骤与关系数据库相同, 不同之处在于隐私数据: 社交网络节点中的属性信息的保护与关系数据库中的数据保护相同, 但社交网络中节点间具有一定关系, 这些关系信息属于结构信息, 社交网络中的隐私保护研究主要是针对结构信息的保护。结构上的隐私泄露主要分 3 类: 个体身份泄露 (identity disclosure)、连接泄露 (link disclosure)、内容泄露 (content disclosure)。个体身份泄露指某节点被关联到某特定个体; 连接泄露指节点间的边的隐私信息被攻击者攻击; 内容泄露指与节点相关的敏感信息, 如邮件信息被破坏。在本节中, 将依据此分类对社交网络保护模型进行讨论。

3.2.1 匿名方法

相比于关系数据库中常用的匿名方法: 泛化及数据扰动等, 社交网络中常用的匿名方法有聚类及图修改法等。聚类包括边聚类和节点聚类^[31,59]: 边聚类通过完全删除隐私边、随机去掉部分边或将节点划分至不同聚类^[23]并保留聚类间的边和类型来实现; 节点聚类是将节点合理划分为组 (通过最大相似度来度量合理性), 每组不少于 k 个节点, 数据发布时, 仅公布组内节点个数及边密度以保护隐私。

图修改法包括最优图构建法、随机化修改方法、贪心方法。其中, 最优图构建法^[60,61]是假设攻击者已知目标节点的度, 将图构建问题转化为构建已知度数序列的图问题; 随机化修改方法^[59,62]是保持节点不变, 通过随机删除 m 条边, 再随机增加 m 条边, 使攻击者获得的结果中含有噪声, 实现隐私的保护。

3.2.2 信息泄露类型与匿名实现模型

1) 个体身份泄露

个体身份泄露指个体的属性信息、结构位置信息、标签信息等被攻击者获得而被识别, 针对该类隐私数据的匿名模型包括: Hay 等^[59]研究了识别匿名网络中已知个体的方法, 提出了 k -candidate 匿名, 通过随机增加、删除边来应对

此类攻击, 实现图结构查询返回的候选节点数目至少是 k 个; Liu 等^[27]假设攻击者知道某些节点的度的信息, 通过最少加边来实现 k -degree 匿名, 使图中相同度数的节点至少有 k 个; Zhou 等^[46]考虑攻击者知道由目标节点的直接邻域构成的子图, 若结构单一便可识别出目标节点, 提出 k -neighborhood 匿名, 通过泛化节点标签和加边来保证任一节点和其邻居节点构成的子图至少与其他 $k-1$ 个子图同构; Zou 等^[33]假设攻击者的背景知识为一名用户周围的所有子图, 提出 k -automorphism 匿名, 使图中任一子图都与至少 $k-1$ 个子图同构, 并提出 k -match 算法以抵抗任何以图结构为背景知识的攻击。

2) 连接泄露

连接即连接节点之间的边, 边分为隐私的和公开的, 需要对隐私边进行保护。

针对无权图的连接泄露保护研究如下。Zheleva 等^[31]针对从匿名的网络中推测出隐私关系的连接泄露提出了 3 种算法: 去掉所有隐私边、随机化去除部分隐私边、聚类。Ying 等^[62]研究了 2 种随机化技术: ① 随机添加 n 条边后随机删除 n 条边, 并提出一种光谱保护随机化方法来引导在社交图谱中选择增加或删除的边; ② 随机交换边。这 2 种随机化技术, 前者保证边总量不变, 后者保证节点度数不变, 并扩展了 2 种方法, 提出了保留特征值谱的算法。Cheng 等^[63]提出 k -isomorphism 匿名, 使 G 包含 k 个两两同构的非连通子图。Liu 等^[27]针对图形的匿名化问题进行研究, 提出一种简单高效的算法, 以最少的边修改操作实现 k -Degree 匿名。Yang 等^[64]针对 k -automorphism 匿名忽略了边泄露而 k -isomorphism 匿名降低了信息可用性的问题, 提出 AK -secure 隐私保护模型, 并提出了适用于该模型的图匿名算法以降低信息损耗。

对有权图的连接泄露保护包括: ①Liu 等^[65]研究的 2 种算法, 高斯随机算法及贪心扰动算法, 用以保护节点对间的最短路径长度, 并最大限度保留一些边的权值, 从而保证数据的可用性; ②Das 等^[66]提出以线性规划为基础, 适当改变原始图中边的权值并保护网络中的线性属性, 如最短路径、最短生成树; ③Liu 等^[67]基于 k -anonymity

隐私保护模型, 提出 k -anonymity 算法以减少加边和权值改变。

3) 内容泄露

内容泄露是指个人隐私在网络中透露给了他人, 采用传统的数据屏蔽技术或由用户有选择地开启和禁用服务即可解决此类问题。

3.2.3 匿名度量与评价

为了保证所发布的数据的可用性, 某些社交图谱的特性要求在匿名之后有所保留。Zheleva^[31]在其研究中量化了施加匿名操作后, 对所发布的数据带来的影响; Ying 等^[62]分析了随机化对网络中各种特性的影响, 并对边匿名所能达到的程度进行了理论分析; Zou 等^[33]使用匿名前后图修改的边数来量化信息损失, 修改的边数越少, 信息损失越少。

4 访问控制方法的研究

4.1 访问控制模型及分类

访问控制系统一般包括: 主体、客体和访问控制策略, 依次指资源请求发起者、被访问的资源、对访问权限具体定义的语句。

社交网络访问控制模型可依据其 8 个特征^[68]进行分类, 这 8 个特征依次为: 请求者身份, 即模型设定的用户身份证明方式; 资源与好友关系决策权, 即资源与关系的映射或分配证书的决定权; 访问控制权, 即用户是否有权决定访问控制规则; 资源控制权, 即用户对各类资源的访问控制权限; 好友关系管理, 即用户对好友关系的分类细度; 凭证分布, 凭证即用户使用的访问控制规则, 凭证分布即凭证的存放地点, 最好的方式是由客户存储; 转授权凭证, 最好的方式是由用户控制其传递; 透明度, 即用户对目前访问控制状态的信息掌握情况。综合这 8 个特征即可对各访问控制模型的优缺点进行判断。

4.2 传统的访问控制模型

传统访问控制模型最初是为了解决大型主机上的数据授权访问问题, 主要分为 2 类^[69,70]: DAC 和 MAC。随着系统规模的扩大, MAC、DAC 逐渐不能满足实际需求, 面向对象的基于角色的访问控制^[71] (RBAC, role-based access control) 应

运而生。RBAC 将访问许可权分配给一定的角色, 用户饰演不同的角色来获得角色所拥有的访问许可权。DAC、MAC 和 RBAC 均为基于主客体观点的被动访问控制模型, 其授权是静态的; 基于属性的访问控制^[72] (ABAC, attribute-based access control) 由于其可实现对主体、客体以及访问控制策略的细粒度刻画而广受关注, 其核心思想是将主体、客体、权限以及访问请求所处的上下文环境通过属性及属性值对来描述^[73], 这样即可实现属性值的运算, 使策略描述更加精确。

4.3 面向社交网络隐私保护的访问控制模型

1) 基于信任的访问控制模型

在 OSN 访问控制模型研究初期, 根据资源拥有者和请求者间的关系类型及信任级别, 研究出基于信任的访问控制模型, 主要包括: Kruk 等^[74]提出一种访问权委派模型 D_FOAF (distributed-friend of a friend), 考虑了 U2U (user-to-user) 这一关系类型, 未考虑 U2R (user-to-resource) 和 R2R (resource-to-resource) 关系, 不适用于功能愈加复杂的 OSN; Carminati 等^[75]提出的基于规则的访问控制模型, 依据现有关系的类型、深度及信任水平来授予关系真实性证书, 在关系信任值计算中一次仅支持一种关系类型, 在文献[76]中, Carminati 提出半分布式架构的访问控制机制, 由客户端进行访问控制, 在文献[77]中, Carminati 提出了分布式安全框架, 由用户依据关系类型、深度和信任值来控制资源的访问及好友关系, 同时采用密码技术来控制资源分享。

但依据信任来进行访问控制, 访问控制管理不够灵活、扩展性能不佳, 且在信任语义的定义、兼顾效率及隐私的信任值计算和监控方案上仍有较大问题^[78]。

2) 基于语义网技术的访问控制模型

语义网^[79]由万维网之父蒂姆·伯纳斯·李提出的概念, 是 WWW (world wide Web) 的延伸, 指将计算机所能理解的语义信息添加入万维网的文档中, 使万维网成为通用信息交换平台。Carminati 等^[80,81]提出采用基于语义网技术的网页本体语言 (OWL, Web ontology language) 和语义网规则语言 (SWRL, semantic Web rule lan-

guage) 来刻画用户间的信任关系, 并依此进行授权、管理, 提出 OSN 可扩展的细粒度访问控制模型, 并在文献[81]中, 以实验性的社交网络测试所提出的访问控制方案。

Masoumzadeh^[82]使用社交网络系统本体 (SNO, social network system ontology) 来描述信息以表达复杂的用户、数据、用户—数据间关系, 提出了访问控制本体 (ACO, access control ontology) 概念以及更为细粒度的访问控制模型 OSNAC, 并证明了此模型的适用性, 但仍需考虑本体和规则推导的效率以及复杂性问题。

3) 基于关系的访问控制模型

在 OSN 的实际应用中, 人们期望它能够提供更真实的个人信息, 且交友模式建立在现实生活中的社交圈, 便产生了基于关系的访问控制机制。Fong 等^[83]对 Facebook 的访问控制机制做了形式化描述, 构建出基于实际 OSN 的隐私保护模型, 之后, Fong 等又引入了模态逻辑语言, 对模态逻辑语言扩展后将其应用于访问控制模型构建, 形成了语言表达能力更强的基于关系的 OSN 访问控制模型 ReBAC^[84] (relationship-based access control)。之后, Bruns^[85]使用混合逻辑 (hybrid logic) 语言对 ReBAC 进行了改进, 使 ReBAC 表达能力更强, 效率更高。

Fong 等在访问控制模型设计中引入了策略语言, 使策略语言成为又一研究热点。Park 等^[86]从用户活动的角度, 将 OSN 的功能和控制活动相区分, 将核心控制活动分为 4 类: 属性、策略、关系、会话, 构造出一种适用于 OSN 的访问控制框架, 随后以正则表达式为策略语言, 提出基于 U2U 关系的访问控制模型 UURAC (user-to-user relationship-based access control), 之后对其完善, 使其支持 U2R 及 R2R^[87], 随后, Park 等依据社交图谱中的关系路径类型及跳数来制定授权策略, 使策略语言更具灵活性, 并提供简单的策略冲突消解机制^[88]; Pang 等^[89]提出一种基于双线性对密码体系的密码协议, 以在分布式社交网络中实现基于关系的访问控制, 并在实际社交网络数据库中进行模拟, 对其效率进行了评估。

4) 基于属性加密的访问控制模型

Shuai^[90]提出基于属性加密的访问控制机制

Masque, 以实现加密数据的互动分享及灵活的访问控制, 其中, 半可信的 OSN 提供商不触及用户的敏感数据且由用户制定个人的访问控制策略, 但其撤销及重定义以及当用户量增大时如何保证加解密效率有待更深入的研究。

Baden^[91]提出 Persona 方案, 将属性加密和传统公钥加密结合得出可由用户自主决定访问策略的无中心系统, 由于个人数据加密存储, 不要求中间媒介可信。

5 结束语

目前, 社交网络中的隐私保护仍然面临着诸多问题和挑战, 需要进一步展开研究。

1) 大规模社交网络。当前的多种隐私保护技术都仅限于小规模社交网络, 当规模逐渐庞大时, 原有技术会因为算法复杂度过高而被淘汰, 目前, 仅有少数针对大规模社交网站的研究, 例如, Narayanan 等^[92]提出一种基于网络拓扑的大规模社交网络的解匿名技术, 利用目标网络与辅助网络的结构相似性来实现, 但这些研究并不成熟, 需要更深入的探索。

2) 动态社交网络。目前, 大多技术主要针对单次发布的数据, 但对于动态复杂的 OSN, 这些技术并不能提供足够的保护, 这使 OSN 中动态数据发布隐私保护技术的研究成为一个重要课题^[93], 例如, Zou 等^[33]提出 k -match 算法使图形满足 k -automorphism 匿名以抵抗任何以图为背景知识的攻击, 并扩展了 k -match 算法以处理动态网络匿名发布中的问题; 谷勇浩等^[94]分析了 k -匿名算法, 并基于此, 提出基于聚类的动态图发布隐私保护方法; Cheng 等^[63]通过泛化 ID 来处理动态网络多重发布问题, 但文中未能给出增删节点的解决方案; Chen^[95]将差分隐私保护方法应用于 OSN 隐私保护中, 但节点间的高度相关性使算法复杂度随数据规模增长较快; Malik 等^[96]提出了一种基于语义网技术的访问控制机制, 通过这种语义标识机制来自动检测敏感信息, 实现 OSN 文本信息的动态隐私保护。从这些研究来看, 目前的解决方案存在着对社交图谱动态变化适应性不强、算法效率低、未考虑或仅考虑单一的背景知识攻击等缺陷, 需要进一步研究。

3) 策略语言。策略语言制定访问控制策略需遵循的格式及标准, 目前的访问控制策略语言有正则表达式、四值逻辑^[97]、简单策略语言 (SPL, simple policy language)、可扩展的访问控制标记语言^[98] (XACML, eXtensible access control markup language)、语义网技术等, 但目前已有的策略语言或表达能力欠缺, 或规则推导效率低, 或策略冲突检测及消解能力不足, 因此, 需设计出更好的策略语言, 也可利用策略合成或策略整合代数融合多种逻辑语言形成自成体系的策略语言。这方面的研究取得了一定的成果: Burns^[99,100]提出基于四值逻辑的策略语言 PBel; Li 等^[101]基于自动化理论及线性约束, 提出一种策略合成语言 (PCL, policy combining language), 可精确表达及评估多种策略合成算法 (PCA, policy combining algorithm), 且实现了与 XACML 的集成; Ni 等^[102]提出了一种策略整合代数 D-algebra 来分析策略语言决策机制, 是很好的策略语言设计及实现工具。

4) 多方访问控制。在 OSN 中, 一些共享信息涉及到多名用户, 对信息相关者做拥有者、贡献者、相关者以及传播者的身份划分, 其中, 拥有者指该信息存放地点为该用户的空间内; 贡献者指该用户在他人空间中发布某类信息; 相关者指被数据拥有者用 @ 等方式与信息相关联的用户; 传播者指该用户分享了此类信息并且将其存在于本身的空间中。

但多数访问控制模型仅支持资源拥有者指定访问策略, 未考虑到贡献者、相关者和传播者, 这就可能引起共享冲突情况的发生, Hu 等^[103]构造出一种简单易用的可实现多用户共享数据的访问控制模型, 并为 OSN 设计了联合授权策略以及冲突解决机制; Squicciarini 等^[104]针对共享内容, 提出所有权由内容创建者分配给多个用户的所有权共享方式, 但这种处理方法易导致管理混乱; Bennett 等^[105]采用基于关系的访问控制机制, 提出 Alloy Analyzer 以检测共享冲突及潜在的错误设置。可以看出, 现有的多方访问控制解决方案缺少灵活性及细粒度性, 且存在缺少表达力强的授权语言以描述复杂的多方授权问题, 有待进一步研究。

参考文献:

- [1] 刘建伟, 李为宇, 孙钰. 社交网络安全问题及其解决方案[J]. 中国科学技术大学学报, 2011, 41(7):565-575.
LIU J W, LI W Y, SUN Y. Security issues and solutions on social networks[J]. Journal of University of Science & Technology of China, 2011, 41(7):565-575.
- [2] Twitter [EB/OL]. <https://zh.wikipedia.org/wiki/Twitter>.
- [3] BILTON N. Twitter implements do not track privacy option[N]. The New York Times, 2012-05-26.
- [4] BAO J, CHENG J. Group trust algorithm based on social network[J]. Computer Science, 2012, 39(2):38-41.
- [5] WEI W, LI Y, ZHANG W. Study on GSNPP algorithm based privacy-preserving approach in social networks[J]. Computer Science, 2012, 39(3): 104-106.
- [6] WANG X G. Discovering critical nodes in social networks based on cooperative games[J]. Computer Science, 2013, 40(4): 155-161.
- [7] AJAMI R, RAMADAN N, MOHAMED N, et al. Security challenges and approaches in online social networks: a survey[J]. International Journal of Computer Science & Network Security, 2011, 11(20).
- [8] SCHNEIDER F, FELDMANN A, KRISHNAMURTHY B, et al. Understanding online social network usage from a network perspective[C]//IMC. c2009:35-48.
- [9] BOYD D M, ELLISON N B. Social network sites: definition, history, and scholarship[J]. Journal of Computer-mediated Communication, 2010, 38(3):16-31.
- [10] ADAMIC L, ADAR E. How to search a social network[J]. Social Networks, 2005, 27(3):187-203.
- [11] NGUYEN N P, XUAN Y, THAI M T. A novel method for worm containment on dynamic social networks[C]//Military Communications Conference. c2010: 2180-2185.
- [12] WEI W, XU F, TAN C C, et al. Sybil defender: defend against sybil attacks in large social networks[C]//IEEE Infocom. c2012: 1951-1959.
- [13] WEIPPL E, GOLUCH S, KITZLER G, et al. Friend-in-the-middle attacks: exploiting social networking sites for spam[J]. Internet Computing IEEE, 2011, 15(3):28-34.
- [14] AHN G J, SHEHAB M, SQUICCIARINI A. Security and privacy in social networks[J]. Internet Computing, 2011, 15(3): 10-12.
- [15] DEY R, TANG C, ROSS K, et al. Estimating age privacy leakage in online social networks[J]. IEEE Infocom, 2012, 131(5):2836-2840.
- [16] YANG C C. Preserving privacy in social network integration with τ -tolerance[C]// 2011 IEEE International Conference on Intelligence and Security Informatics (ISI). c2011:198-200.
- [17] IRANI D, WEBB S, PU C, et al. Modeling unintended personal-information leakage from multiple online social networks[J]. IEEE Internet Computing, 2011, 15(3):13-19.
- [18] KRISHNAMURTHY B, WILLS C E. Characterizing privacy in online social networks[C]//The first workshop on online social networks. c2008:37-42.
- [19] LUO W, XIE Q, HENGARTNER U. FaceCloak: an architecture for user privacy on social networking sites[C]//International Conference on Computational Science & Engineering. c2009:26-33.
- [20] BEYE M, JECKMANS A J P, ERKIN Z, et al. Privacy in online social networks[J]. International Journal of Computer Applications, 2012, 41(13):5-8.
- [21] SMITH H J, DINEV T, XU H. Information privacy research: an interdisciplinary review[J]. MIS Quarterly, 2011, 35(4):989-1016.
- [22] BANSAL G, ZAHEDI F, GEFEN D. The moderating influence of privacy concern on the efficacy of privacy assurance mechanisms for building trust: a multiple-context investigation[C]//The International Conference on Information Systems(ICIS). Paris, c2008.
- [23] 宋文略. 社交网络数据的隐私保护研究[D]. 南京: 南京大学, 2011.
SONG W L. Social network data privacy protection research[D]. Nanjing: Nanjing University, 2001.
- [24] HUANG Q, ZHU J, SONG B, et al. Game model of user's privacy-preserving in social networks[J]. Computer Science, 2014, 41(10):184-190.
- [25] MARTIN D J, KIFER D, MACHANAVAJJHALA A, et al. Worst-case background knowledge for privacy-preserving data publishing[C]//IEEE 23rd international Conference on Data Engineering(ICDE). c2007:126-135.
- [26] ZHOU B, PEI J, LUK W S. A brief survey on anonymization techniques for privacy preserving publishing of social network data[J]. ACM Sigkdd Explorations Newsletter, 2008, 10(2):12-22.
- [27] LIU K, TERZI E. Towards identity anonymization on graphs[C]//ACM Sigmod. c2008:93-106.
- [28] BACKSTROM L, DWORK C, KLEINBERG J. Wherefore art thou R3579X?: anonymized social networks, hidden patterns, and structural steganography[C]//The 16th International Conference on World Wide Web. c2007: 181-190.
- [29] WONDRAČEK G, HOLZ T, KIRDA E, et al. A practical attack to de-anonymize social network users[C]// 2010 IEEE Symposium on Security and Privacy (SP). c2010:223-238.
- [30] CORDELLA L P, FOGGIA P, SANSONE C, et al. A (sub) graph isomorphism algorithm for matching large graphs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2004, 26(10):1367-1372.
- [31] ZHELEVA E, GETOOR L. Preserving the privacy of sensitive relationships in graph data[C]//Pinkdd. c2007:153-171.
- [32] LIU K, DAS K, GRANDISON T, et al. Privacy-preserving data analysis on graphs and social networks[J]. 2008.
- [33] ZOU L, CHEN L, ZSU M T. K -automorphism: a general framework for privacy preserving network publication[J]. The VLDB Endowment, 2009, 2(1):946-957.

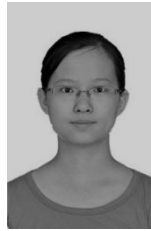
- [34] SAMARATI P, SWEENEY L. Generalizing data to provide anonymity when disclosing information (abstract)[C]//The 17th ACM Sigact-Sigmod-Sigart Symposium on Principles of Database Systems. c1998:188.
- [35] SAMARATI P. Protecting respondents' identities in microdata release[J]. IEEE Transactions on Knowledge & Data Engineering, 2001, 13(6):1010-1027.
- [36] TAO Y, XIAO X. Personalized privacy preservation[C]//The 2006 ACM SIGMOD International Conference on Management of Data. c2010:229-240.
- [37] BRAND R. Microdata protection through noise addition[C]//Inference Control in Statistical Databases From Theory to Practice. c2002:97-116.
- [38] FULLER W A. Masking procedures for microdata disclosure limitation[J]. Journal of Official Statistics, 1993.
- [39] MEYERSON A, WILLIAMS R. On the complexity of optimal k -anonymity[C]// The 23rd ACM Sigmod-Sigcat-Sigart Symposium. c2010:223-228.
- [40] XIAO X, TAO Y. Anatomy: simple and effective privacy preservation[C]//International Conference on Very Large Data Bases. c2006: 139-150.
- [41] KIM J J, WINKLER W E, CENSUS B O T. Masking microdata files[C]//The Survey Research Methods. c1997:114-119.
- [42] REISS S P, POST M J, DALENIUS T. Non-reversible privacy transformations.[C]//The ACM Symposium on Principles of Database Systems. California, c1982:139-146.
- [43] DALENIUS T, REISS S P. Data-swapping: a technique for disclosure control [J]. Journal of Statistical Planning & Inference, 1982, 6(1):73-85.
- [44] DU W, ZHAN Z. Abstract: using randomized response techniques for privacy-preserving data mining [C]//The Ninth ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, Washington DC. c2003:505-510.
- [45] EVFIMIEVSKI A, SRIKANT R, AGRAWAL R, et al. Privacy preserving mining of association rules[J]. Information Systems, 2004, 29(4): 343-364.
- [46] ZHOU B, PEI J. Preserving privacy in social networks against neighborhood attacks[C]//IEEE International Conference on Data Engineering. c2008:506-515.
- [47] CORMODE G, SRIVASTAVA D, YU T, et al. Anonymizing bipartite graph data using safe groupings[J]. VLDB Journal, 2010, 19(1): 115-139.
- [48] SWEENEY L. K -anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2008, 10(5):557-570.
- [49] WANG K, FUNG B C M. Anonymizing sequential releases[C]// IEEE Computer Science. c2006:414-423.
- [50] MACHANAVAJHALA A, KIFER D, GEHRKE J. L -diversity: privacy beyond k -anonymity[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1):24.
- [51] WONG C W, LI J, FU W C, et al. α, k -anonymity: an enhanced k -anonymity model for privacy preserving data publishing[C]// ACM Sigkdd. c2006:754-759.
- [52] LI N, LI T, VENKATASUBRAMANIAN S. T -Closeness: privacy beyond k -anonymity and l -diversity[C]// IEEE 23rd International Conference on Data Engineering (ICDE). c2007:106-115.
- [53] NERGIZ M E, ATZORI M, CLIFTON C. Hiding the presence of individuals from shared databases[C]//The ACM Sigmod International Conference on Management of Data. c2007:665-676.
- [54] AGRAWAL R, SRIKANT R. Privacy preserving data mining[M]// Foundations and Advances in Data Mining. Berlin Heidelberg: Springer, 2000:439-450.
- [55] GIONIS A, MAZZA A, TASSA T. K -anonymization revisited[C]//IEEE 24th International Conference on Data Engineering (ICDE). c2008:744-753.
- [56] IYENGAR V S. Transforming data to satisfy privacy constraints[C]//The eighth ACM Sigkdd International Conference On Knowledge Discovery And Data Mining. c2002: 279-288.
- [57] BAYARDO R J, AGRAWAL R. Data privacy through optimal k -anonymization[C]// The 21st International Conference on Data Engineering (ICDE). c2005: 217-228.
- [58] NERGIZ M E, CLIFTON C. Thoughts on k -anonymization[J]. Data & Knowledge Engineering, 2007, 63(3):622-645.
- [59] HAY M, MIKLAU G, JENSEN D, et al. Resisting structural re-identification in anonymized social networks[J]. The Vldb Endowment, 2008, 1(1):797-823.
- [60] DIESTEL R. Graph theory[J]. Oberwolfach Reports, 2000, 311(1):67-128.
- [61] GROSS J L, YELLEN J. Graph theory and its applications, second edition (discrete mathematics and its applications)[M]. Chapman & Hall/CRC, 2005.
- [62] YING X, WU X. Randomizing social networks: a spectrum preserving approach[C]//The SIAM International Conference on Data Mining. Georgia, c2008:739-750.
- [63] CHENG J, FU W C, LIU J. K -isomorphism: privacy preserving network publication against structural attacks[C]// International Conference on Management of Data. 2010:459-470.
- [64] YANG J, WANG B, YANG X, et al. A secure k -automorphism privacy preserving approach with high data utility in social networks[J]. Security & Communication Networks, 2014, 7(9): 1399-1411.
- [65] LIU L, WANG J, LIU J, et al. Privacy preserving in social networks against sensitive edge disclosure[R]. Technical Report Technical Report CMIDA-HiPSCCS 006-08, 2008.
- [66] DAS S, ÖMER E, ABBADI A E. Anónimos: an LP-based approach for anonymizing weighted social network graphs[J]. IEEE Transac-

- tions on Knowledge & Data Engineering, 2010, 24(4):590 - 604.
- [67] LIU C G, LIU I H, YAO W S, et al. *K*-anonymity against neighborhood attacks in weighted social networks[J]. Security & Communication Networks, 2015, 8(18):3864-3882.
- [68] 韩钰佳. 社交网络访问控制安全研究[D]. 西安: 西安电子科技大学, 2013.
- HAN Y J. Social network access control security research [D]. Xi'an: Xidian university, 2013.
- [69] SNYDER L. Formal models of capability-based protection systems[J]. IEEE Transactions on Computers, 1981, 30(3):172-181.
- [70] 李风华, 苏锐, 史国振, 等. 访问控制模型研究进展及发展趋势[J]. 电子学报, 2012, 40(4):805-813.
- LI F H, SU M, SHI G Z, et al. Research status and development trends of access control model[J]. Acta Electronica Sinica, 2012, 40(4): 805-813.
- [71] FERRAILOLO D F, KUHN D R. Role-based access controls[C]//The 15th NIST-NCSC National Computer Security Conference. c1992:554-563.
- [72] HU V C, KUHN D R, FERRAILOLO D F, et al. Attribute-based access control[J]. Computer, 2015, 48(2):85-88.
- [73] LIN L, HUIAI J, LI X. Attribute-based access control policies composition algebra[J]. Journal of Software, 2009, 20(2):403-414.
- [74] KRUK S R, GRZONKOWSKI S, GZELLA A, et al. D-FOAF: distributed identity management with access rights delegation[C]//Asian Semantic Web Conference. c2006:140-154.
- [75] CARMINATI B, FERRARI E, PEREGO A. Rule-based access control for social networks[M]//On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops. Berlin: Springer, 2006:1734-1744.
- [76] CARMINATI B, FERRARI E, PEREGO A. Enforcing access control in web-based social networks[J]. ACM Transactions on Information & System Security, 2009, 13(1):6:1-38.
- [77] CARMINATI B, FERRARI E, PEREGO A. A decentralized security framework for Web-based social networks[J]. International Journal of Information Security & Privacy, 2008, 2(4):22-53.
- [78] FERRARI E. Access control, privacy and trust in on-line social networks: issues and solutions[M]. Berlin :Springer, 2011.
- [79] LEE T B, HENDLER J, LASSILA O. The semantic Web[J]. Semantic Web Research & Applications, 2001, 284(5):28-37.
- [80] CARMINATI B, FERRARI E, HEATHERLY R, et al. A semantic Web based framework for social network access control[C]//ACM Symposium on Access Control Models & Technologies. c2009: 177-186.
- [81] CARMINATI B, FERRARI E, HEATHERLY R. Semantic Web-based social network access control[J]. Computers & Security, 2011, 30(2/3):108-115.
- [82] MASOUMZADEH A, JOSHI J. OSNAC: An ontology-based access control model for social networking systems[C]//2010 IEEE Second International Conference on Social Computing (Social-Com). c2010:751-759.
- [83] FONG P W L, ANWAR M, ZHAO Z. A privacy preservation model for facebook-style social network systems[C]//The 14th European Conference on Research in Computer Security. c2009:303-320.
- [84] FONG P W L. Relationship-based access control: protection model and policy language[C]//The First ACM Conference on Data & Application Security & Privacy. c2011:191-202.
- [85] BRUNS G, FONG P W L, SIAHAAN I, et al. Relationship-based access control: its expression and enforcement through hybrid logic[C]//ACM Conference on Data and Application Security and Privacy. c2012:117-124.
- [86] PARK J, SANDHU R, CHENG Y. A user-activity-centric framework for access control in online social networks[J]. IEEE Internet Computing, 2011, 15(5):62-65.
- [87] YUAN C, PARK J, SANDHU R. A user-to-user relationship-based access control model for online social networks[C]//International Conference on Data & Applications Security & Privacy. c2012:8-24.
- [88] YUAN C, PARK J, SANDHU R. Relationship-based access control for online social networks: beyond user-to-user relationships[C]//2012 International Conference on Privacy, Security, Risk and Trust (PASSAT), and 2012 International Conference on Social Computing (SocialCom). c2012:646-655.
- [89] PANG J, ZHANG Y. Cryptographic protocols for enforcing relationship-based access control policies[C]//IEEE Computer Software and Applications Conference. c2015.
- [90] SHUAI H, ZHU W T. Masque: access control for interactive sharing of encrypted data in social networks[C]//The 6th international conference on Network and System Security. c2012:503-515.
- [91] BADEN R, BENDER A, SPRING N, et al. Persona: an online social network with user-defined privacy[J]. ACM Sigcomm Computer Communication Review, 2009, 39(4):135-146.
- [92] NARAYANAN A, SHMATIKOV V. De-anonymizing social networks[C]// Eprint Arxiv:0903. c2009:173-187.
- [93] BHAGAT S, CORMODE G, KRISHNAMURTHY B, et al. Prediction promotes privacy in dynamic social networks[J]. WWW, 2010.
- [94] 谷勇浩, 林九川, 郭达. 基于聚类的动态社交网络隐私保护方法[J]. 通信学报, 2015(S1).
- GU Y H, LIN J C, GUO D. Clustering-based dynamic privacy preserving method for social networks[J]. Journal on Communications, 2015(S1).
- [95] CHEN R, FUNG B C M, YU P S, et al. Correlated network data publication via differential privacy[J]. VLDB Journal, 2014, 23(4): 653-676.
- [96] MALIK I D, SÁNCHEZ D, VIEJO A. Privacy-driven Access Control in Social Networks by Means of Automatic Semantic Annotation[J]. Computer Communications, 2016, 76:12-25.
- [97] BELNAP N D. A useful four-valued logic[M]//Modern Uses of

Multiple-Valued Logic. Berlin:Springer,1977:5-37.

- [98] CLIFTON C, KANTARCIOGLU M, VAIDYA J. Defining privacy for data mining[J]. National Science Foundation Workshop on Next Generation Data Mining, 2002, 1(26): 1.
- [99] BRUNS G, HUTH M. Access control via belnap logic: intuitive, expressive, and analyzable policy composition[J]. ACM Transactions on Information & System Security, 2011, 14(1):1165-1182.
- [100]BRUNS G, DANTAS D S, HUTH M. A simple and expressive semantic framework for policy composition in access control[C]// ACM Workshop on Formal Methods in Security Engineering. c2007:12-21.
- [101]LI N, WANG Q, QARDAJI W, et al. Access control policy combining: theory meets practice[C]//ACM Sacmat. c2009:135-144.
- [102]NI Q, BERTINO E, LOBO J. D-algebra for composing access control policy decisions[C]//The 4th International Symposium on Information, Computer, and Communications Security. c2009: 298-309.
- [103]HU H, AHN G J, JORGENSEN J. Multiparty access control for online social networks: model and mechanisms[J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 25(7):1614-1627.
- [104]SQUICCIARINI A C, SHEHAB M, WEDE J. Privacy policies for shared content in social network sites[J]. VLDB Journal, 2010, 19(6):777-796.
- [105]BENNETT P, RAY I, FRANCE R. Analysis of a relationship based access control model[C]// The Eighth International Conference on Computer Science & Software Engineering. c2015: 1-8.

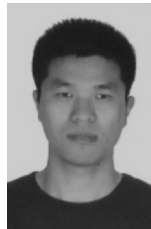
作者简介:



姚瑞欣(1994-),女,山西运城人,西安电子科技大学硕士生,主要研究方向为密码学、隐私保护。



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。



曹进(1985-),男,陕西西安人,博士,西安电子科技大学副教授,主要研究方向为无线网络安全。