隐私保护技术研究综述

李晓晔[1] 孙振龙[2] 邓佳宾[3] 宋广军[4]

(齐齐哈尔大学计算中心 齐齐哈尔 161006)[1] (齐齐哈尔大学招生办 齐齐哈尔 161006)[2]

(齐齐哈尔大学网络信息中心 齐齐哈尔 161006)[3] (齐齐哈尔大学计算机与控制工程学院 齐齐哈尔 161006)[4]

摘 要 随着数据发布和数据挖掘的广泛应用及快速发展,如何保护隐私数据以防止敏感信息泄露,已经成为当前的研究热点。文中分别从这两个层面对隐私保护技术进行分析总结,并对已有算法进行深入对比分析,最后指出该领域中可深入研究的两个方向。

关键词 隐私保护,数据发布,数据挖掘

中图法分类号 TP309 **文献标识码** A

Survey of Study on Privacy Preserving

LI Xiao-ye[1] SUN Zhen-long[2] DENG Jia-bin[3] SONG Guang-jun[4]

(Computer Center, Qiqihar University, Qiqihar 161006, China)[1]

(Enrollment Office, Qiqihar University, Qiqihar 161006, China)[2]

(Network Information Center, Qiqihar University, Qiqihar 161006, China)[3]

(Department of Computer and Control Project, Qiqihar University, Qiqihar 161006, China)[4]

Abstract As the extensive applications and fast development of data publishing and data mining, how to protect private data to prevent the disclosure of sensitive information, has already become a current research hotspot. This paper analyzes the privacy preserving technology respectively from the two aspects, and makes further comparison and analysis of existing algorithms. Finally, we point out two directions can be in-depth study in this field.

Keywords Privacy preserving, Data publishing, Data mining

1 引言

随着信息技术的发展,各行各业都积累了大量有用的数据。一方面,基于信息共享、知识决策、科学研究等方面的需要,要求将这些数据进行发布。但如果发布者直接发布原始数据,而不采取适当的数据保护措施,很可能会泄露个人的敏感信息,进而危害数据所有者。因此,为了保证个人敏感信息的安全,应该在发布数据的同时进行隐私保护。另一方面,数据挖掘作为一个强有力的数据分析工具,能够从大量的、不完全的、有噪声的数据中识别和提取隐含的、未知的、新颖的、潜在有用的知识和规则。数据挖掘在科学研究、商务决策、医学研究等领域做出了巨大的贡献,同时也产生了不可避免的隐私泄露问题,受到业界及社会各方面的越来越多的关注。因此,数据挖掘应该在隐私保护的条件下开展。

本文主要从以下两个层面分析总结隐私保护技术:(1)隐私保护的数据发布(Privacy Preserving Data Publishing, PPDP)[1],即在数据发布中,对原始数据进行扰动、加密或匿名等处理,以实现隐私保护。(2)隐私保护的数据挖掘(Privacy Preserving Data Mining, PPDM)[2],即在隐私保护的条件下,分别针对关联规则、分类、聚类等,研究高效的挖掘算法。

2 隐私度量与评估标准

隐私保护技术需要在保护隐私的同时,最大可能地保证数据的有效性,即兼顾对实际应用的价值。通常,可以从以下两个角度对隐私保护进行度量与评估。

2.1 披露风险

披露风险[3]定义为,攻击者根据背景知识,可能从发布的数据集中,所披露隐私的概率。即通过攻击者披露隐私的多少,来侧面反映隐私保护的效果。若发布者最终发布的数据集中,所有敏感数据的披露风险均小于阈值 $\alpha(0\leqslant\alpha\leqslant1)$,则称该数据集的披露风险为 α。现有的隐私度量,可以统一使用披露风险来描述。

例如,Machanavajjhala 等人提出了一种静态数据发布原则 l-Diversity[4],保证了发布数据集的披露风险小于 $1/l$。该原则使得每一个等价类的敏感属性至少有 l 个不同的值,即确保攻击者最多以 $1/l$ 的概率披露某个体的敏感信息。Xiao 等人提出了动态数据发布原则 m-Invariance[5],保证了每个时刻,发布数据集的披露风险小于 $1/m$。该原则使得某时刻发布的数据集中,每一个等价类至少有 m 条记录,且这些记录都有不同敏感属性值。

本文受黑龙江省教育厅科学技术研究项目(12511601),齐齐哈尔大学青年教师科研启动项目(2011k-M04\2011k-M20)资助。

李晓晔(1981—),女,硕士,讲师,主要研究方向为数据挖掘、隐私保护,E-mail:qdpyb@163.com;孙振龙(1981—),男,硕士,讲师,主要研究方向为数据挖掘、网络安全;邓佳宾(1981—),男,硕士,工程师,主要研究方向为信号处理;宋广军(1963—),男,博士,教授,硕士生导师,主要研究方向为网络安全、数据库、数据挖掘。

特殊情况下，发布数据集时不做任何处理，则认为披露风险为1；若所发布数据集的披露风险为0，则认为该数据集实现了完美隐私(Perfect Privacy)[6]。显然，完美隐私实现了最大程度的隐私保护。但是，真正的完美保护并不存在，也只有在具体假设、特定场景下才成立。

2.2 信息缺损

信息缺损表示为，经过隐私保护技术处理后数据的信息丢失，是针对发布数据集质量的一种度量。

最小信息缺损原则，通过比较原始数据和匿名数据的相似度来衡量隐私保护的效果。信息缺损越小，说明发布数据集的有效性越高。但是，这种度量原则需要考虑准标志符中每个属性的每个取值的泛化和隐匿带来的信息缺损，计算代价较高，适用于对单个属性进行度量。

ILoss[7]度量标准，要求检查每条记录准标志符中每个属性的取值泛化带来的信息缺损，进而计算出每条记录泛化后的信息缺损，再根据每条记录的信息缺损，计算整个发布数据集的信息缺损。

3 隐私保护的数据发布

隐私保护的数据发布，是在数据发布中，对原始数据进行扰动、加密或匿名等处理，以实现隐私保护。

3.1 基于数据失真的隐私保护技术

数据失真技术，那通过扰动原始数据来实现隐私保护。数据扰动的基本思想是隐藏真实的原始数据，只呈现出数据的统计学特征。具体地讲，经过扰动之后，攻击者通过发布的失真数据，不能重构出真实的原始数据，即不能发现真实的原始数据。并且，失真后的数据仍然保持某些性质不变，即利用失真数据得出的某些信息等同于从原始数据上得出的信息。

数据交换是一种基本的扰动技术，是在记录之间交换数据的值，保留某些统计学特征而不保留真实数值。另外一种技术是随机化，是对原始数据加入随机噪声从而隐藏真实数值。值得注意的是，任意对数据进行随机化，并不能保证数据和隐私的安全[8]，因为利用概率模型进行分析，可能揭露随机化过程中的众多性质。

倪巍伟等人提出了一种隐私保护数据干扰方法 NETPA[9]，其通过对数据点及邻域点集的分析，借助信息论中熵的理论，对原始数据中数据点的邻域主属性值用其 k 邻域点集内数据点在该属性的均值进行干扰替换，在较好地维持原始数据 k 邻域关系的情况下，达到保护原始数据隐私不泄露的目的。张勇等人提出了一种基于邻域相关性的数据扰动算法 NCDP[10]，其分析每个数据点邻域中与其邻域亲密的所有点以及邻域的平衡性，在不平衡情况下除去亲密集中可能的局部噪声数据点，向每个邻域亲密点进行一定长度的平移，得到扰动后的数据点。扰动后的数据点不仅实现了对原始数值的保护，而且扰动前后数据点的邻域亲密点仍然维持亲密关系，从而保持了邻域的稳定性。

3.2 基于数据加密的隐私保护技术

基于数据加密的隐私保护技术，多用于分布式应用中，是通过密码机制实现他方对原始数据的不可见性以及数据的无损失性，以实现隐私保护。

在分布式环境下，两个或多个参与者想要共同完成一项计算，但希望每个参与者只能看到自己的输入数据和经计算后的最终结果。这个问题可以抽象为无信任第三方参与的SMC问题，即安全多方计算问题，其是密码学的一个分支，是用安全的方式实现分布式协同工作。

Pinkas 针对安全多方计算，在基于数据加密的隐私保护数据挖掘[11]方面，进行了一定的探讨。文中指出安全两方计算的构建要易于安全多方计算，同时，用于计算评价函数的最佳组合电路的规模是影响协议的一个重要方面，而健忘传输协议以及协议的改进程度又是影响计算的一个主要因素。目前，关于安全多方计算的研究主要集中于降低计算开销、优化分布式计算协议等。

3.3 基于限制发布的隐私保护技术

所谓限制发布，是指不发布或者发布敏感度较低的数据，即有选择地发布原始数据，以实现隐私保护。限制发布在隐私披露风险与数据敏感度之间进行折中，即保证隐私披露风险在一定阈值范围之内，有选择地发布敏感数据。

当前此类技术的研究热点，集中于数据匿名化。数据匿名化一般采用两种基本操作，一种是抑制，即不发布某些数据项；另一种是泛化，即对数据进行更概括、抽象的描述。数据匿名化的研究重点，主要是设计更好的匿名化原则，使发布数据既能很好地保护隐私，又具有较大的使用价值。同时，针对特定的匿名化原则，设计更为高效的匿名化算法。

Sweeney 提出了 k-匿名原则[12]，保证所发布数据集中的每一条记录，不能区分于其他 $k-1$ 条记录，这里“不能区分”，只针对非敏感属性项而言。文中称不能相互区分的 k 条记录为一个等价类，一般 k 值越大，对隐私的保护效果越好，但丢失的信息也越多。杨高明等人定义了单敏感值(α,k)匿名模型和多敏感值(α,k)匿名模型[13]，并分别设计了两个聚类算法予以实现。对于既包含连续属性又包含分类属性的数据集，给出了数据集的详细映射与处理方法，使数据集中点的距离可以方便的计算，彻底避免了把数据点距离和信息损失混淆的情况。王波等人提出了一种面向个体的个性化扩展 l-多样性隐私匿名模型[14]，该模型在传统 l-多样性的基础上，定义了扩展的 l-多样性原则，并通过设置敏感属性的保护属性来实现个体与敏感值之间关联关系的个性化保护需求。同时，还提出了一种个性化扩展 l-多样性逆聚类(PELI-clustering)算法来实现该隐私匿名模型。

4 隐私保护的数据挖掘

隐私保护的数据挖掘，是指在隐私保护的条件下，分别针对关联规则、分类、聚类等，研究高效的挖掘算法。

4.1 隐私保护的关联规则挖掘

在隐私保护的关联规则挖掘中，规则隐藏成为其中的研究热点，即尽可能降低敏感规则的支持度或者置信度，以此使得需要保护或隐藏的规则不被挖掘出来。

Oliveira 等人基于数据清洗的思想，提出了一系列隐私保护关联规则挖掘算法。算法的基本思想是，在原始数据库中，找出支持敏感规则的敏感事务，进行项目的删除或增加，以降低敏感规则的支持度或置信度到指定的阈值，从而隐藏敏感规则，并减少数据清洗阶段的影响。其中，SWA 算法[15]的具体操作是，对数据库中的事务进行分组清洗，先找出事务组中的敏感事务并确定其牺牲项，即在敏感规则中出现频率最高的项，再针对每一条敏感规则，根据给定的泄漏阈值计算所需清洗的敏感事务数，而后选择最短的几个事务删除牺牲项，从而使规则的支持度降低，实现敏感规则的隐藏。该算法操作

简单，只需根据启发式规则，修改敏感事务和牺牲项即可，并且可以为每一条敏感规则设置不同的泄漏阈值，从而提供了更为灵活的安全控制。事实上已经证明，在这类算法中应用启发式技术，有选择性地修改数据以达到最佳隐藏效果，是一个 NP 难度问题[16]。

Saygin 等人基于数据阻塞的思想，提出了一种隐私保护关联规则挖掘算法[17]，该算法是以不确定值，例如"?"，替换原始数据库中支持敏感规则的数据，从而将规则的支持度或置信度调整为不确定的区间值。当敏感规则支持度或置信度区间下界，低于最小支持度或置信度阈值时，认为规则得到了隐藏。该算法减小了错误数据产生错误规则的可能性，从而增强了数据的实用性。

Rizvi 基于随机扰乱技术，提出了一种隐私保护关联规则挖掘算法——MASK 算法[18]，主要针对购物篮事务数据集。该数据集的列字段由商品名称组成，每一行代表某一位顾客购买商品的情况，其中"1"表示已购买对应商品，"0"表示未购买对应商品。因此，该数据集中每一行是一个 0 与 1 的字符串，整个数据集是一个 0 与 1 的矩阵。该算法使用概率方法改变原始数据的值，使得数值以概率 P 保持不变，以概率 $1-P$ 取反。若某列的值由"1"取反变为"0"，相当于删除该列值；若由"0"取反变为"1"，则相当于添加噪声项。该算法需要重构项目集的支持度，即估算项目实际的支持度，并非重构项目的实际值，进而发现频繁项目集。该算法由于对 0 与 1 采用相同的保护策略，使得数据集的密度增大，导致程序运行时间增加，降低了算法执行效率。

4.2 隐私保护的分类挖掘

隐私保护的分类挖掘是指在数据挖掘的过程中，建立一个准确的、无隐私泄露的分类模型。

Agrawal 等人提出了一种基于重构技术的隐私保护分类挖掘算法[19]，该算法先在原始数据中添加随机偏移量，以进行随机化混乱，再利用 Bayesian 公式推导出原始数据的密度函数，建立一种准确程度接近真实数据分布的分类标记，以此重构决策树。同时，文中引入了一些量化方法，从置信度以及预测的准确程度上，对算法进行了检验。但是，该算法中推导原始数据的密度函数，使用的迭代法计算量特别大，而且仅限原始数据为数值型数据，且尽可能均匀分布，并不能处理布尔类型和枚举类型的数据，所以此算法仍然有进一步改进的余地。

Chang 等人提出了一种结合吝啬降级法以及决策树分析的隐私保护方法[20]。所谓降级是指，对公开发布的信息进行隐私保护处理的过程。吝啬降级旨在需降级的数据集中，形式化信息的处理形式，信息处理方向为安全环境到数据公开环境。就其实现而言，算法通过产生一个参变量基础集来实现数据的降级，使用一个参数 $\theta(0\leqslant\theta\leqslant1)$ 来取代敏感数据，θ 表示某个属性取得一个可能值的概率。同时，对于降级前后的数值的熵进行计算，使用二者的差值同数据库变化前后的置信度的降低程度进行比较，从而得出其对数据库的修改是否可以接受。而决策树用来对可能存在的信息推导渠道进行分析，以此决定需要对哪些数据进行降级。

Amirbekyan 等人基于 k-最近邻分类算法提出了相应的隐私保护算法[21]，该算法在协议的执行中，利用半可信第三方的参与，同时应用同态加密算法，设计出两方共享结果的安全点积计算协议、两方/多方安全向量和计算协议，并在此基础上提出了安全最大值计算协议，以及安全距离度量协议，进而实现了隐私保护的 k-最近邻分类算法。显然，半可信第三方的参与会降低协议的安全性，但能使算法效率获得较大的提高。

4.3 隐私保护的聚类挖掘

安全地计算数据间的距离是隐私保护的聚类挖掘的关键，当前研究基于距离的隐私保护聚类挖掘方法居多。

Oliveira 等人提出了一种用于聚类的隐私保护算法[22]，该算法是通过对数据进行整体旋转变换(RBT)，来达到对敏感数据的隐藏。算法的具体步骤为，首先，对数据进行统一的规范化处理，得出变换后的新值，且在处理的过程中，可以按照规定的准则，去除一些不影响聚类结果的敏感数据。然后，根据规定的阈值计算变换矩阵，计算准则是，随机选出两组列向量，与变换矩阵相乘后产生结果列向量，结果列向量与原始列向量相减产生差向量的方差，应该大于或等于规定的阈值。最后，通过该准则选出合适的变换矩阵，所有属性列与该变换矩阵运算后得到的数据集，即为最终的结果数据集。由于该算法是基于旋转变换的等距变换，因此在变换前后挖掘结果相同。但是因为旋转角度 θ 的旋转范围是根据所要求的最低隐私保护度确定的，所以当对隐私保护的要求较高时，算法有可能无法取得合适的旋转角度。

隐私保护的分布式聚类有两种常用的模型，一种是 Naive 隐私保护聚类模型，各个站点先加密本地数据后传递给信任第三方，再由信任第三方进行聚类后返回结果。另一种是隐私保护的多次聚类模型，所有站点先对本地数据进行隐私保护聚类并发布结果，再对各个站点发布的结果进行二次处理，最终实现隐私保护分布式聚类[23]。此外，还有一些隐私保护聚类算法，如在任意划分数据环境下的 k-Means 聚类算法[24]，引入随机数以保证安全传输的最大期望聚类算法，一维空间及多维空间的数据等距变换聚类算法，都能够有效地解决隐私保护聚类挖掘中，数据类型适用性不强等缺点。

结束语 本文主要从数据发布和数据挖掘两个层面，分析总结了隐私保护技术，并对其中比较典型的算法进行了介绍和分析。但大多算法是针对特定的应用和数据集，并且在保密性、准确度、效率及可扩展性等方面存在一些不足。同时，如何综合各种数据处理技术，提出具有一定的通用性的隐私保护算法，也将是值得进一步研究的方向。

总之，隐私保护技术具有很大的研究空间，其中还存在不少问题值得深入探讨，以下列举该领域中可深入研究的两个大方向，仅供研究人员参考。

(1)差分隐私保护

2006 年，Dwork 等人首次提出了差分隐私保护(Differential Privacy)[25]，并且开发出一套提供差分隐私保护的应用框架 PINQ(Privacy Integrated Queries)[26]，以便于开发者进行相关的差分隐私保护系统开发。

差分隐私保护采用添加噪声的技术使敏感数据失真，是基于数据失真的隐私保护技术。虽然其基于数据失真技术，但所需加入的噪声量与数据集的大小无关。因此，即使对于大型数据集，也只需添加极少量的噪声，就可以达到高级别的隐私保护。此外，差分隐私保护定义了一个极为严格的攻击模型[27]，并对隐私披露风险给出了定量化的表示和证明。差分隐私保护可以保证，在数据集中删除或添加一条数据，不会影响到查询结果。因此，即使在最坏情况下，攻击者已知除某

条记录之外的所有敏感数据，仍然可以保证不泄露这条记录的敏感信息。差分隐私保护极大地保证了数据的可用性，同时大大降低了隐私泄露的风险。正是由于差分隐私保护技术的诸多优势，使得该方法在国外一经出现，就掀起了一股研究热潮，但在国内还少见相关文献。

(2)社会网络隐私保护

社会网络，又称作社会关系网络或社交网络，作为一种新兴的互联网商业模式，正受到越来越多的关注。社会网络构成一类特殊的图，许多图论的研究方法均可以用于社会网络。1967 年，Milgram 的连锁信实验和六度空间理论，也称为小世界现象，是基于社会网络的一个著名研究。Leskovec 等人利用 MSN Messenger 收集信息，进行六度空间理论的研究，是历史上有关此理论的最大规模的研究[28]。

实际上，社会网络中的大多信息与个人生活密切相关，其中很多内容涉及到个人隐私。研究表明[29]，当前社会网络中的用户，部分缺乏对个人隐私的关注，部分表现出一定程度的关心，但仍然低估了隐私泄露的风险。虽然，互联网中的隐私保护已经受到了各界的关注，但是基于 Web 2.0 和社会网络这个特定环境下，关于隐私的研究仍然处于起步阶段。随着社会网络的进一步发展及其用户群的进一步扩大，基于社会网络的隐私保护将会很快成为研究热点。

参 考 文 献

[1] Fung B, Wang K, Chen R, et al. Privacy-Preserving Data Publishing: A Survey of Recent Developments [J]. ACM Computing Surveys, 2010, 42(4): 1-53

[2] Agrawal R, Srikant R. Privacy-Preserving Data Mining [J]. ACM SIGMOD Record, 2000, 29(2): 439-450

[3] 周水庚，李丰，陶宇飞，等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报，2009：847-861

[4] Machanavajjhala A, Kifer D, Gehrke J. l-Diversity: Privacy beyond k-Anonymity[C]// Proceedings of the 22nd International Conference on Data Engineering. Atlanta, Georgia, USA, 2006: 24-35

[5] Xiao X, Tao Y. m-Invariance: Towards Privacy Preserving Republication of Dynamic Datasets[C]// Proceedings of the ACM SIGMOD Conference on Management of Data. Beijing, China, 2007: 689-700

[6] Machanavajjhala A, Gehrke J. On the Efficiency of Checking Perfect Privacy[C]// Proceedings of the Symposium on Principles of Database Systems. Chicago, Illinois, USA, 2006: 163-172

[7] Xiao X, Tao Y. Personalized Privacy Preservation[C]// Proceedings of ACM SIGMOD Conference on Management of Data. Chicago, 2006: 229-240

[8] Kargupta H, Datta S, Wang Q, et al. On the Privacy Preserving Properties of Random Data Perturbation Techniques[C]// Proceedings of the IEEE International Conference on Data Mining. Melbourne, Florida, 2003: 99-106

[9] 倪巍伟，徐立臻，祟志宏，等. 基于邻域属性熵的隐私保护数据干扰方法[J]. 计算机研究与发展，2009(03)：498-504

[10] 张勇，倪巍伟，祟志宏，等. 基于邻域相关性的面向聚类数据扰动方法[J]. 计算机研究与发展，2011(S3)：79-85

[11] Pinkas B. Cryptographic Techniques for Privacy Preserving Data Mining[J]. ACM SIGKDD Explorations, 2002, 4(2): 1-14

[12] Sweeney L. k-Anonymity: A Model for Protecting Privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570

[13] 杨高明，杨静，张健沛. 聚类的(α, k)-匿名数据发布[J]. 电子学报，2011(08)：1941-1946

[14] 王波，杨静. 一种基于逆聚类的个性化隐私匿名方法[J]. 电子学报，2012(05)：883-890

[15] Oliveira S R M, Zaiane O R. Protecting Sensitive Knowledge by Data Sanitization [C]// Proceedings of the 3rd IEEE International Conference on Data Mining. Melbourne, 2003: 613-616

[16] Atallah M, Bertino E. A Elmagarmid. Disclosure Limitation of Sensitive Rules [C]// Proceedings of the IEEE Knowledge and Data Exchange Workshop. Chicago, 1999: 45-52

[17] Saygin Y, Verykios V S, Clifton C. Using Unknowns to Prevent Discovery of Association Rules [J]. ACM SIGMOD Record, 2001, 30(4): 45-54

[18] Rizvi S J, Haritsa J R. Maintaining Data Privacy in Association Rule Mining[C]// Proceedings of the 28th VLDB Conference. Hong Kong, 2002: 682-693

[19] Agrawal R, Sriknat R. Privacy-Preserving Data Mining [C]// Proceedings of the ACM SIGMOD Conference on Management of Data. Dallas, Texas, 2000: 439-450

[20] Chang L W, Moskowitz I S. Parsimonious Downgrading and Decision Trees Applied to the Inference Problem[C]// Proceedings of Workshop on New Security Paradigms. New York, 1998: 82-89

[21] Amirbekyan A, Estivill-Castro V. Privacy-Preserving k-NN for Small and Large Data Sets[C]// Proceedings of Seventh IEEE International Conference on Data Mining. Omaha, Nebraska, 2007: 699-704

[22] Oliveira S R M, Zaiane O R. Achieving Privacy Preservation when Sharing Data for Clustering [C]// Proceedings of the Workshop on Secure Data Management in a Connected World in Conjunction with VLDB' 2004. Toronto, Ontario, Canada, 2004: 67-82

[23] Jagannathan G, Pillaipakkamnatt K, Wright R N. A New Privacy-Preserving Distributed k-Clustering Algorithm [C]// Proceedings of the 2006 SIAM International Conference on Data Mining. Bethesda, Maryland, 2006: 492-496

[24] Jagannathan G, Wright R N. Privacy Preserving Distributed k-Means Clustering over Arbitrarily Partitioned Data[C]// Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, IL, USA, 2005: 593-599

[25] Dwork C. Differential Privacy[C]// Proceedings of the 33rd International Colloquium on Automata Languages and Programming. Berlin, 2006: 1-12

[26] http://research.microsoft.com/en-us/projects/PINQ/

[27] 李杨，温雯，谢光强. 差分隐私保护研究综述[J]. 计算机应用研究，2012(9)：3201-3211

[28] Leskovec J, Horvitz E. Planetary-scale Views on A Large Instant-Messaging Network[C]// Proceedings of the 17th International Conference on World Wide Web. 2008: 915-924

[29] 罗亦军，刘强，王宇. 社会网络的隐私保护研究综述[J]. 计算机应用研究，2010(10)：3601-3604