**大数据隐私保护技术及其应用**

马小梅

软件工程2105 202126010530

**摘要**

随着互联网技术的发展,人们进入了“大数据”时代，数据的收集、存储和分析已经成为各行各业的重要手段。然而，这也引发了一系列的隐私保护问题，不法分子会搜集用户信息,实行违法犯罪活动。因此,社会需要重视隐私保护，保障人们的隐私安全。本文分析了隐私保护现状,并给出保护隐私的相关技术，希望为大数据时代下的隐私保护技术的发展提供积极的促进作用。

**关键词:** 大数据；用户信息；隐私保护；关键技术

**引言**

众所周知,大数据时代的到来为人们的生产与生活带来了便利,如人们通过使用数据库,如人们通过使用数据库、云存储等方式进而及时发现信息价值，从而应用到生活实践里。此外由于互联网的广泛普及，人们运用智能手机、iPad等移动电子设备，可以随时随地在社交网络上发布地址定位信息、照片、日记等个人隐私信息。这些信息被网络数据服务商等机构不断地跟踪和记录，挖掘其潜在的巨大价值，寻求数据背后隐藏的巨大经济利益。如何保护好个人的信息一直是网络用户、专家学者乃至国家备受关注的问题。总之大数据在带来便捷的同时也埋下来信息安全隐患。因此,基于大数据的隐私保护技术进行深入研究和探索,切实提升大数据隐私保护效果也就显得极为必要。

**一、大数据安全问题**

随着互联网的广泛普及，人们进入了大数据时代，信息安全也处于新的节点。大数据安全问题涉及的主要特点和挑战包括大数据遭受异常流量攻击、信息泄露风险、数据生命周期安全、基础设施安全、个人隐私安全和存储管理风险。在大数据时代背景下，由于各种移动终端设备的普及程度很高，人们可以使用电脑、手机等设备随时随地搜索信息，这也增加了信息泄露的风险。例如，智能手机中的各种软件app存在一些漏洞，不法分子可以利用这些漏洞入侵用户的设备，盗取用户个人信息。

此外，大数据平台本身也面临许多安全问题。这些问题包括缺乏安全规划、性能低下、配置不当以及审计功能不足等。因此，数据安全和个人信息安全都面临着漏洞、泄露和恶意行为等风险。为了解决这些问题，需要新的安全方案和方式来提升数据安全治理能力。

**二、大数据个人隐私保护技术**

**（一）数据库的个人隐私保护**

数据库的个人隐私保护是一个关键问题，特别是在数字化和大数据时代，数据量规模化的增长,数据的采集、传输、存储等环节都会存在数据泄露的安全隐患。个人信息的收集、存储和使用越来越普遍，各个行业需要使用数据库来保存大量用户的个人信息。保护数据库中的个人隐私，需要满足两个核心要求：数据保密性和数据可用性。数据保密性意味着潜在的攻击者无法逆推出准确的敏感信息，对于一些关键信息无法获取。而数据可用性则要求处理后的数据仍然保持某些统计特性或可分辨性，使其在某些业务场景中是有用的。这两个指标在一定程度上是矛盾的，因此如何平衡和调节它们成为了一个重要的问题，需要明确哪些数据字段需要加强保密，哪些字段可以暴露更多信息。从访问技术来讲：访问技术主要是通过对用户进行授权的方式来保证信息安全，同时访问技术还能记录用户的行为，一遍追踪术用户的不合理行为，其中，访问技术也需要根据实际情况灵活应用。从数据加密来讲，数据加密主要以密文的形式存储信息，以便防止不法分子读取用户信息。因此用户信息需要进行加密。此外内部人员在查看加密信息时也需要有不同的权限，进而预防信息泄露。

然而，数据库的安全问题仍然是一个值得关注的话题。确保数据库中的数据安全、隐私保护和可靠性是每个组织和个人的重要任务。因此，了解并采取一些实用的方法来提高数据库的安全性是非常必要的。

**（二）云存储环境下的个人隐私保护**

在云计算技术日益发展并广泛应用的今天，越来越多的组织和个人选择将其数据存储和处理转移到云平台上。然而，这种便捷性同时也带来了数据安全和隐私保护的问题。因此,发展基于云计算的数据隐私保护技术是当前研究热点之一。针对云环境下的数据安全问题，企业和组织需要采取一系列措施和技术进行防护。云存储环境下的个人隐私保护技术旨在确保用户的数据在云端存储和处理过程中保持安全和私密。以下是一些关键的个人隐私保护技术：

首先，数据加密是最基本的保护措施之一。主要是传输加密：使用安全协议（如SSL/TLS）对数据在从用户设备传输到云存储服务提供商的过程中进行加密，防止数据在传输过程中被截获。存储加密：对存储在云中的数据进行加密，包括服务器端加密和客户端加密。即使数据被盗，也无法直接读取其内容。通过将数据转化为密文，只有拥有正确密钥的人才能解密并访问数据。这种方法可以有效防止未经授权的人员获取敏感信息。此外，基于属性的加密方案也可以用于解决云环境下的信任问题，克服了不可信第三方、安全性和性能开销的难题。

其次，访问控制和身份验证也是保护个人隐私的重要手段。通过设置权限和密码保护等方式，可以确保只有被授权的用户才能访问和使用存储在云中的数据。华为云等服务提供商会在用户资料中提供个人数据清单，明确说明相关个人数据处理的业务场景、目的、个人数据范围及处理方式等信息，帮助用户了解并评估相关业务的安全性。

最后，采用一些先进的技术也可以提高云存储中的数据安全性。例如，同态加密允许在密文上进行计算，而无需解密数据，这有助于保护数据的隐私性。差异隐私技术则可以向数据添加足够的噪声，以保护数据集中的私有信息，同时仍然保留足够的信息以供使用。

这些个人隐私保护技术在云存储环境下协同工作，为用户提供了一种可靠的方式来保护他们的个人信息免受未经授权的访问、滥用和泄露。然而，随着技术和威胁环境的发展，这些技术也需要不断更新和改进，以应对新的隐私挑战。

**三、大数据隐私保护关键技术**

**（一）加密技术**

大数据隐私保护技术旨在在处理和分析大规模数据集的同时，确保个人隐私得到有效的保护。大数据隐私保护中的加密技术是保护数据安全性和隐私性的重要手段。它旨在将一个信息（或称明文）经过加密钥匙和加密函数转换，变成无意义的密文，而接收方则将此密文经过解密函数、解密钥匙还原成明文。数据加密可以分为硬件加密和软件加密两种途径。硬件实现网络数据加密的方法有链路层加密、节点加密和端对端加密。

在大数据的环境下，数据量巨大，涉及的个体多且复杂，因此对于数据的加密需求更为严格。同态加密就是在这样的背景下诞生的，它可以在密文上进行计算，而无需解密数据，从而在不暴露原始数据的情况下完成计算。此外，为了解决大数据集中的隐私保护问题，差分隐私技术也得到了广泛的应用。通过向数据添加足够的噪声，来保护数据集中的私有信息，同时仍然保留足够的信息以供使用。这些加密技术在大数据环境下相互配合，可以提供多层次、全方位的数据隐私保护。然而，由于大数据的规模和复杂性，实施加密时也需要考虑性能、可扩展性和管理复杂性等因素。因此，选择和设计适合大数据特性的加密方案并结合信息技术的现状进行发展和完善至关重要。

**（二）数据脱敏**

数据脱敏是一种处理敏感信息的技术，旨在在保持数据的使用价值的同时，去除直接识别个人身份的信息，以保护隐私。在大数据的环境下，这项技术的应用尤为重要。常见的需要脱敏处理的敏感数据包括姓名、身份证号码、地址、电话号码、银行账号、邮箱地址、所属城市、邮编、密码类（如账户查询密码、取款密码、登录密码等）、组织机构名称、营业执照号码、银行帐号、交易日期、交易金额等。

数据脱敏的方法是在保留一定的数据可用性、统计性等基础上，通过失真等变换实现降低数据敏感度。例如，可以对某些字符进行替换或加密，将敏感数据替换为虚构的、无关联的值，如用“\*”或“匿名”替换姓名和地址，或者对数值型数据进行微小的偏移，使其与原始值接近但不完全相同，从而实现对数据的变形处理。还有差分隐私脱敏：在数据发布过程中引入随机噪声，使得单个个体的贡献无法被精确识别，同时保持数据集的整体统计特性；时间序列脱敏：对时间序列数据进行处理，如滑动窗口、时间偏移等，以防止基于时间模式的识别；图像和音频脱敏：对图像中的面部、车牌等敏感部分进行模糊、像素化或马赛克处理。对音频中的语音进行变声或噪声添加，以保护说话人的身份。这样，就可以在开发、测试和其它非生产环境以及外包环境中安全地使用脱敏后的真实数据集。

值得一提的是，腾讯安全玄武实验室披露了一项关于大模型隐私保护的安全脱敏与还原（Hide and Seek, HaS）技术，有望帮助大模型产品使用者从本地终端侧防范隐私数据泄露。此项技术的发展进一步增强了数据脱敏技术在大数据隐私保护方面的应用潜力。数据脱敏技术的选择取决于具体的应用场景、数据类型和隐私保护需求。实施数据脱敏时，需要确保脱敏后的数据既能满足业务需求，又能有效保护个人隐私。同时，也需要定期审查和更新脱敏策略，以应对新的威胁和挑战。

**（三）其他技术**

在大数据隐私保护中，除了数据加密和脱敏等技术外，工作流隐私保护、隐私风险评估和管理以及用户透明度和控制也是至关重要的方面。以下是对这些关键技术的阐述：

在工作流隐私保护方面，采用的主要策略是在数据处理的整个过程中确保隐私性。工作流隐私保护关注的是在整个数据处理生命周期中，从数据收集、传输、存储、处理到销毁，如何确保隐私得到持续的保护。这包括设计和实施隐私保护的工作流程，例如，可以通过数据脱敏、匿名化等技术，将敏感信息从数据集中移除或替换，以降低其识别性和敏感性，同时保持数据的完整性和可用性。此外，还可以采用基于数据分离、数据干扰、安全多方计算、硬件增强和访问模式隐藏等方向的隐私保护技术。

对于隐私风险评估和管理，目标是主动发现并解决大数据收集、数据分析，以及大数据发布过程中存在的隐私风险。风险评估是通过分析数据处理活动、数据类型、数据使用目的、潜在的数据泄露途径等因素，确定隐私风险的级别和影响，然后采取相应的措施来管理和降低这些风险。风险管理则涉及制定和执行相应的策略和措施，以降低已识别的风险至可接受的水平。这可能包括改进数据处理流程、增强访问控制、实施审计和监控机制、提供员工培训等。

在用户透明度和控制方面，旨在让用户对他们的数据如何被收集、存储、处理和使用有充分的了解，并且有能力对这些过程进行有效的控制，让用户的隐私权益得到充分的尊重和保护。这需要提供清晰、易于理解的隐私政策和用户协议，明确告知用户数据的收集目的、存储期限、分享对象以及用户可以行使的选择权（如访问、更正、删除数据等）。用户控制则是赋予用户对其个人信息的决定权，让他们能够选择是否提供数据、同意特定的数据使用方式以及调整隐私设置。实现用户透明度和控制通常需要开发用户友好的界面和工具，使用户能够轻松查看、管理他们的数据和隐私设置。

综合运用以上关键技术，可以构建一个全面的大数据隐私保护框架，确保在大数据的各个环节中，个人隐私得到充分的尊重和保护。同时，随着技术和法规环境的变化，这些措施也需要不断更新和优化，以应对新的隐私挑战和要求。

**结语**

总结而言，大数据隐私保护技术是当前数据安全领域的重要研究方向。随着数据量的爆炸性增长和数据应用的广泛深入，个人隐私的保护面临着前所未有的挑战，因此加强大数据隐私保护显得尤为重要。通过不断创新和应用诸如数据脱敏、加密、差分隐私、同态加密、多方安全计算等关键技术，我们能够在利用大数据价值的同时，有效地维护个体的隐私权益。这些技术的应用不仅限于理论层面，它们已在医疗健康、金融服务、电子商务、社交媒体等多个领域展现出实际效果，帮助企业和组织在合规的前提下进行数据分析和业务优化，更好地管理和利用大数据资源同时确保用户数据的安全性和隐私性，提高业务效率和竞争力。

然而，大数据隐私保护是一个持续发展的领域，新的威胁和挑战不断涌现，需要我们持续研发更为先进和适应性强的保护技术。此外，政策法规的制定和执行、公众隐私意识的提升以及跨学科的合作也是推动大数据隐私保护工作不可或缺的部分。在未来，我们期待看到更加成熟的大数据隐私保护体系，它将深度融合技术创新、法规约束和社会伦理，实现大数据价值的最大化利用与个人隐私的坚实保障之间的平衡，共同构建一个既智能又尊重隐私的数字化社会。因此，对大数据隐私保护技术及其应用的研究和发展，不仅是技术问题，更是社会责任和法律义务的重要体现。

**参考文献**

[1]傅常顺.大数据隐私保护技术的应用[J].电子技术与软件工程, 2016(18):2.

[2] 曾琴.信息科技·计算机软件及计算机应用·《数字技术与应用》.2018，(7)：311

[3] 高瑞，李俊，杨睿超.大数据安全和隐私保护技术架构研究[J].信息系统工程，2018,(10): 78

[4]丁知平林昆.大数据背景下的隐私保护关键技术的研究[J].电脑编程技巧与维护, 2022(2):72-74.

[5] 张娟萍,张海亮.大数据与数据隐私保护相关问题研究[J].软件, 2023, 44(2):81-84.