关于大数据安全与隐私保护的研究

谢　丹
杭州合众数据技术有限公司　浙江　杭州　310000

摘　要：数字经济发展的当务之急是为数据的供给和利用提供安全合规的环境，提高数据流通效率，实现数据价值的最大化。中央全面深化改革委员会发布的《关于构建数据基础制度更好发挥数据要素作用的意见》表明，隐私计算可以破解数据供给和利用的难题。本文主要对大数据安全与隐私保护进行论述，详情如下。

关键词：大数据；安全；隐私；保护

引言

2008年由美国著名计算机科研人员最初发表的《大数据计算：在商务、科学和社会领域创建革命性突破》中所说大数据真正的用途不是数据本身，而是新用途和新见解。大数据是一种在不同时间和位置内使用常规的计算机程序对数据进行合规化的搜索、采集、分类和管理处理数据的集合。大数据具有海量性、高速性、多样性和易变性。大数据的规模非常庞大且并不断地变化。目前我国使用大数据信息的类型比较多，常用于处理音频、视频和各种网络文件、GPS定位系统、社交工具和社交软件等。处理信息速度快、准确性能好、并能有效的分解分析信息时效性高，这是大数据信息最显著的特性，是区别于传统数据分析系统的最大特点。

1 大数据时代的信息特征

行业通常以Volume，Variety，Value，Velocity和Veracity来概括大数据的特征。Volume指的是数据规模巨大；Variety意思是数据的种类繁多。所以数据可被分为结构化与非结构化。目前非结构化数据逐渐增多，其中地理位置信息、网络日志、音频、视频、图片等为非结构化数据，这要求对数据的处理能力更加精确；Value表明的是价值的密度。价值密度的大小与数据总量的大小成负相关；Velocity是大数据区分于传统数据挖掘的最显著的特征；Veracity指的是数据来自各种、各类信息系统网络以及网络终端的行为或痕迹。体量大，结构多样，时效性强是大数据的特征，因而需采用新型的计算架构与智能算法技术来处理这些大数据。

2 大数据安全与隐私保护

2.1 实施行业准则监管

加强政府对行业自律和个人隐私保护的行政监督。在数据隐私的行业自律保护中，政府应充分发挥自身职能作用，将数据隐私的行业进行有效监督；其次，应注意引导民众安全使用信息、培育信息标准化；最后，政府出台政策促进新兴技术的开发和利用，鼓励互联网企业根据我国市场情况制订符合本国发展的行业标准和大多数商业认可的法律、法规。互联网协会、中国电子商务诚信联盟等行业组织应发挥更积极的作用。在行业自律方面，可制订《中国互联网行业隐私保护自律公约》，如《中国互联网行业自律公约》，专门保护数据隐私权，为整个行业提供隐私声明样本；此外，还可成立专门的数据隐私保护行业组织评估和认证数据隐私政策的实施情况；同时，这些组织要向各行、各业的企业宣传数据隐私保护的意义及作用，让其明白有效地保护数据隐私不仅是保护数据用户，而且更是保护整个社会的上网环境。

2.2 数据要素流通使用的应对策略

管理制度与技术支撑相互保障的数据要素流通使用全流程合规可信体系包括合规可信制度体系、合规可信技术体系以及管理制度与支撑技术协同方案。数据要素可信流通使用制度体系包括事前审查制度、事中监控制度、事后审计制度等。技术体系包括数据交易系统技术、区块链系统技术、跨隐私平台的联邦学习系统技术以及可信执行环境技术等。在数据流通使用的事前审查阶段，制定针对交易主体、交易数据和交易合约的审查制度，应对参与主体和数据采集安全风险；在技术上采用“机器审查+人工核验”方式保证审查流程合规可信，即对于资质信息、数据质量、交易条目等标准信息，如企业法人信息、营业执照、数据规模与量级、禁止交易数据清单等，采用基于机器学习算法进行自动审查与人工抽验方法；对于交易目的、数据来源等主观性较大的数据属性，采用人工核验方法。在数据流通使用的事中监控阶段，针对流通使用涉及的平台系统及软硬件、数据、云、网、端等环节制定安全保障制度，构建交易主体监控管理体系、算法行为监控管理体系和订单履行监控管理体系；在技术上设计基于智能算法支撑的保障体

系，如基于智能识别技术的参与主体身份认证，保证参与主体可信；基于标识技术的数据权限管理方法，实现交付数据访问可控；面向数据用量异常检测的自动感知技术，监控数据合规加工使用；基于区块链技术的数据流通使用过程信息存证，保证数据流通使用全过程可追溯。在数据流通使用的事后审计阶段，制定数据滥用审计制度、数据侵权审计制度、主体失信审计制度，旨在确保数据流通使用全过程合规、争议可裁决、权益可保障；在技术上设计基于区块链存证信息的再审计体系，对数据流通使用全过程进行安全审计；基于数据标识和关联技术的数据追踪体系，对数据二次流通、转卖等侵权行为进行查验取证；融合交易主体信用评估制度体系与区块链可追溯技术，构建数据信用综合评估服务，推动数据流通市场公正可信发展。

2.3 基于分类分级的数据安全防控策略

2.3.1 数据安全级别的动态控制策略

数据安全级别不是一成不变的，会随着时间、业务、规模和数据开放状态等因素动态变化。在资源层面上的动态影响因素主要有3个：业务状态、时效、资源规模：（1）业务状态：在特定事件后，数据安全级别可能发生改变，如行业标准文档的起草、评审到最终发布，文件内容会逐渐公开化；（2）时效：在一定时间后，数据安全级别可能发生改变。互联网数据作为典型案例，具有时效性短的特性；（3）资源规模：当数据资源达到一定规模时，数据安全级别可能发生改变。可想而知，依靠亿万级的平台打的或车辆导航数据，能够绘制出高分辨率的全国交通路线图，甚至精确定位重要的单位机构，此时的受侵害客体或将转变为国家安全。字段层面的动态影响因素主要有两个：是否经过脱敏处理，是否是回填字段、标签字段或统计字段。通常当数据经过脱敏后，其安全级别会随之降低。如果字段并非来源数据，而是经过数据融合、治理得到的回填、打标签或统计信息，这些字段体现了更高的数据价值，也伴随着更高的敏感度。记录层面更多的是根据红名单的规则进行动态调整，字段关系层面可根据构成关系的字段的数据安全级别设计动态控制规则。

2.3.2 基于数据安全级别的数据脱敏策略

当用户能够查看特定资源的数据时，其中的高敏感字段、记录的信息脱敏就至关重要。可依据数据安全级别，在字段、记录、字段关系的层面上进行脱敏。（1）对于字段，当用户的权限级别低于字段的安全级别时，对这些字段的数据进行脱敏处理；（2）对于记录，当用户的权限级别低于敏感记录的安全级别时，将一整行的记录脱敏后展示或不返回；（3）对于字段关系，分以下两种情形：①一个字段由复数个字段组成，对高于用户权限的组成字段进行脱敏；②两个字段构成了推导关系，至少对其中一个字段进行脱敏。

2.4 加强计算机网络安全防范意识

大数据网络时代，提高计算机网络安全不仅要靠技术，还需要我们从自身意识着手，警惕危险、提高计算机网络安全防范意识。从小处着手，个人在使用计算机网络技术时，要注重个人信息的保护，不要将个人信息到处传播，对于一些不良网站或是不可信网站不要点击，不给不法分子任何可乘之机；然后是企业及各单位对员工、学生等人群进行计算机网络安全防范意识的宣传，并向大家演示各种可以泄露个人信息或者被窃取信息的方式，以实例来提醒大家我们的个人信息会从哪些地方被盗取；然后是政府可以加大对网络安全的宣传力度，用各种后果来警醒我们若是被盗取了个人信息会导致些什么样的后果，我们将承受什么样的损失，以此来加大我们对网络安全防范的意识，认识到网络安全防护的重要性；另外，政府还可以根据时代发展特征，以信息安全、网络风险为重点进行各种网络安全宣传，通过不同的宣传方式让网络安全防范意识深入人心，从而更易被大众理解并接收。这样，我们在使用计算机网络技术时，能够下意识地对我们的信息安全进行保护，并通过使用正版软件来保护我们的网络安全，从而提高了我们的网络安全应对和防范能力。

2.5 健全网络安全防范机制

完善计算机网络安全防范机制是更高效的应对网络安全风险问题的保障，对网络安全防范具有积极的意义。因此，在大数据背景下，必须要完善计算机网络安全防范机制，以此保证网络安全防范工作的顺利展开。具体的实施措施有以下几个方面：①完善网络使用管理机制。可以加强对不良网站、非法网站的打击力度，清空这些存在安全风险网站的生存空间；或者对进行信息储存的硬盘或者移动存储设备加强检查或者做好数据备份，从而减少病毒攻击频率或者减少意外泄露的风险，即使意外遗失后还有备份可以使用；还可以通过制定详细的网络条例以保证黑客的操作是违法行为，从而致使他们在进行违法操作时知道有达摩克利斯之剑悬在头顶，慎重而行。②完善安全防护责任机制。大数据背景下，为了保证网络安全，国家或者企业都有设置网络信息安全管理员，以确保网络数据信息和工作内容不会因为病毒而被盗取；制定针对性的网络安全防护条例与责任制度，一方面是帮助我们树立良好的网络安全防护意

识，从源头上保护我们的数据隐私，另一方面则是将安全防护责任扩大落实到个人，以此提高网络安全防护的效果。

2.6 数据全生命周期安全治理

根据上文对数据生命周期的研究，其中涵盖数据采集、传输、存储、处理、使用以及销毁等环节，各环节采取的措施各不相同。在数据采集环节，有必要加强对数据来源的验证，判断数据的合法性，核对数据质量问题，分析数据是否合规，对于采集到的数据项完成级别判定与标签设置等工作；数据传输环节，应保证所有数据不会被恶意篡改或者被不法分子窃取，一般会使用SFTP以及HTTPS等合理的加密协议，避免信息被泄露，或者应用摘要算法保证数据在传输期间的完整。如果数据需要进行导入或者导出操作，此时，还应对数据的提供与接收者、数据来源与去向做好溯源管理；数据存储环节，有必要做好数据的分离存储，特别是对于敏感与非敏感性数据，以及不同等级的数据，需要及时被存储于不同的区域，再使用类似SM4的密码技术完成数据加密存储，最后，定期对数据进行备份测试即可；数据处理环节，应对于数据加工等操作合理授权，只允许特定人员完成以上操作，处理时保证所有数据都能完成脱敏与去标识化处理，防止信息被泄露。如果处理后的数据存在敏感性变化，此时应重置标签；数据使用环节，对于数据的使用需提前完成授权管理，只有符合要求的用户进行数据访问操作，数据共享或者数据开放前应完成脱敏处理，或者依靠隐私计算技术对数据完成不可见的共享；数据销毁环节，如果数据不再使用，此时应及时销毁，但必须要用正规的销毁工具将数据和副本全部销毁，最大程度上保证数据的安全。

结语

总而言之，现阶段各个领域加大了对大数据应用的重视程度，大数据的使用与安全治理问题日益突出。根据数据的生命周期情况，了解数据质量现状，遵循相应治理原则，完善数据安全治理技术架构，加强隐私保护，实现用户身份认证，对不同的数据采取适当的存储与备份、恢复方式，全方位保护数据应用安全。

参考文献

[1]魏来，陈洋.大数据环境下加强个人信息安全与隐私保护的对策建议[J].沈阳工程学院学报（社会科学版），2019，15（02）：167-171.

[2]董淑芬，李志祥.大数据时代信息共享与隐私保护的冲突与平衡[J].南京社会科学，2021（05）：45-52+70.

[3]张聪丛，郜颖颖，赵畅，杜洪涛.开放政府数据共享与使用中的隐私保护问题研究—基于开放政府数据生命周期理论[J].电子政务，2018（09）：24-36.

[4]刘芝兰，刘永贵，刘瑞.后疫情时代全球高校信息安全六大挑战《2021年EDUCAUSE地平线报告—信息安全版》分析（上）[J].中国教育网络，2021（Z1）：40-41.

[5]匡文波，童文杰.个人信息安全与隐私保护的实证研究—基于创新扩散理论的大数据应用视角[J].武汉大学学报（人文科学版），2016，69（06）：104-114.