

Big Data Security and Privacy Protection

Dan GOU Ke WANG Chenxi LI

School of Computer and Software, Jincheng College, Sichuan University, Chengdu, Sichuan, 611731

Abstract

Following the pace of the network era, the source channels of personal data in the network are increasing. More enterprises and individuals than in the past obtain beneficial value by processing and analyzing a large number of data. However, in this process, people's information security is facing dangers, but also brings some troubles to people's daily life. This paper briefly describes the basic knowledge of big data and the crisis accompanied by big data, and analyses the related technologies and measures of how to protect data security and personal information in the era of big data.

Key Words

Big Data Crisis, Data Security, Privacy Protection

DOI:10.18686/jsjxt.v1i2.654

大数据安全和隐私保护

苟丹 王科 李晨曦

四川大学锦城学院计算机与软件学院, 四川成都, 611731

摘要

跟随网络时代的脚步, 个人数据在网络中的来源渠道不断增加, 比过去更多的企业和个人通过处理和分析大量数据来获取有利价值, 但是, 在这过程中人们的信息安全面临着危险, 同时也给人们的日常生活带去了一些困扰。本文简述了大数据的基本知识和大数据伴随的危机, 浅析了在大数据时代下如何保护数据安全和个人信息的相关技术和措施。

关键词

大数据危机; 数据安全; 隐私保护

1. 引言

在大数据时代, 交通、物流等很多行业的模式都在发生改变, 给人们提供了更加便利的生活模式, 例如共享单车的推广、购物 APP 的使用等, 但是, 人们同时也遇到了一些隐私安全问题, [1]大量的公开数据显示, 2011 年到目前为止, 遭受到数据泄露的人大概有 11.2 亿。人们对于个人隐私的认知也不断地发展起来, 数据安全的保护技术逐步发展, 国家对于个人信息的安全措施也进行了修缮, 从而保护个人隐私和数据安全得到了人们的高度重视。

2. 大数据

2.1 大数据的定义和特征

大数据, 不一定是数据量够大才能成为大数据, 而是能够让软件工具捕捉、存储、处置和分析的数据或资料。[2]大数据的 4V 特征包含:

(1) 体量 (volume) 是指可采集的指数增长的海量非结构化的数据和相对较少的结构化数据, 并且数据的存储和计算量也很大;

(2) 多样性 (variety) 是指庞大的数据展示出的数据形式多样性, 不同方式获取的数据可以呈现出多种形式, 比如文本、图片、语音、视频、页面、机器语言等, 也表示为不同的构造, 分为结构性数据和非结构数据;

(3) 价值密度低 (value) 是指通过大量的数据分析可以提取的价值却如同沙里淘金一样稀少, 所以获取的价值密度低;

(4) 速度 (velocity) 是指对于网络时代而言, 数据更新和数据获取的速度非常快, 如果不及时处理数据就会过时, 并且大数据处理数据的速度也非常快。

2.2 大数据的优势

快速发展的网络化和信息化扩大了人们的交流渠道, 彼此间关系愈发的密切, 生活更是越来越便利, 产生的数据不断增加, 大数据的概念随之出世。据报告, 使用 Google 搜索引擎的用户平均每秒有 200 万人, 每天“脸书”的社交网站中共共享的用户信息超过 40 亿, 并且, 医学行业、通信行业、金融业、交通业等各行各业也不停的在产生大量数据。^[3]

2.2.1 促进企业结构的变革

更多的行业会获取的相关数据进行分析, 从中获取有商业价值的信息, 来测量事物的发展走向、获取更大的商机。有关机构通过调查研究从事大数据方向的企业得出, 2017 年中国的数据产业规模多达 4000 多亿元, 增幅在 30% 左右, 大数据的应用为企业提供的利润巨大。相对于原始的长期摸索, 掌握真实、有效的数据, 让一些企业可以结合当前趋势得以进一步的改革和发展。

2.2.2 带来人们生活的便利

医疗行业将大数据结合云计算、人工智能等技术研究出更高效的医疗方案, 对于人们的医疗手段有了进一步的提升, 并且其他领域利用捕获的数据通过剖析, 为用户推送更有效、更高效的信息, 让人们从中体验到了大数据带来的便利。

3. 大数据伴随的危机

大数据带来了好处和价值的同时也带来了一些隐私问题和数据安全问题, 近几年来, 都发生了数据泄露事件, 比如 2018 年 9 月, “脸书”的用户信息再次遭受到了大规模的泄漏, 大约有 5000 万用户的账户可能被入侵、甚至遭受到窃取;^[4]

3.1 网络危机

由于互联网的生活, 个人信息的“透明化”, 发生了

大量的网络数据被盗事件, 从而造成了用户数据的流散, 其中也伴随着财产损失的危险。网络是大量数据产生的平台之一, 是信息采集的一个重要途径, 2018 年 AcFun 视频网站宣称, 遭受到黑客入侵泄露了大约有 800 万的用户数据, 并且黑客将获取的数据放到网络上售卖, 售价为 1 元 800 条, 让人们再次关注到了数据在网络中的安全问题。在虚拟的网络世界, 数据的真实性和可信度是无法准确判断的, 虚假的信息在一定程度上不仅会影响人们恐慌甚至会造成社会的动荡, 2011 年的强盐风波, 一度引起了人民的恐慌。

3.2 应用危机

应用软件的安全性也影响到了数据的安全, 将病毒程序植入到应用中让个人信息受到了要挟, 违法分子利用带有病毒的应用程序获取个人信息和个人财产, 2018 年华夏银行技术处长将病毒编成程序入侵华夏银行的核心系统, 造成的损失高达 700 余万。

3.3 新技术危机

随着新技术不断的产生, 数据的安全问题更加凸出。云计算的出现, 大量数据被存取在云端上, 比如 2011 年在线社区 CSDN, 发生了 600 万用户账号、密码被泄露的事件, 云存储的安全成了新的技术问题。

3.4 企业和用户数据安全问题

企业对于用户隐私的不重视, 2016 年西安城西海伦春天小区接到多个业主的电话投诉, 声称个人信息被开发商泄露, 因此不断的接到各种装修推销电话, 经调查表明是开发商自主将业主的信息泄露的, 买卖用户的个人信息成为了一种新的交易方式。2018 年 3 月, 纽约时报揭露了剑桥分析机构未取得用户许可, 私自使用 Facebook 用户的个人数据。除了企业自身问题, 企业内部对人员管理不当造成了数据安全问题, 2017 年思科因的员工由于操作不当引发了云服务数据泄露。

用户对于保护个人隐私安全的还不够看重, 在大数据时代中, 人们花在网络上的时间明显增加, 留下了很多的个人信息, 由于人们都没有意识到个人信息随意填写的危机, 所以增加了个人隐私暴露的几率。

4. 大数据安全和隐私保护技术

4.1 数据发布匿名技术

数据的匿名发布主要是针对结构化的数据比如学生个人信息表,这种技术是将原始的数据表经过匿名的处理后,然后进行使用,相关人员才能获取其中准确的信息。数据发布匿名技术是非动态的,但是在日常生活中,数据的形式大多是以非结构化存在的,而且形式众多、结构复杂、是连续的、多次的,这是这项技术的难点所在。

4.2 网络社交匿名技术

大多数的数据都源于网络,不法分子在网络中通过一些相关联的节点来推断用户的有效信息,这项技术针对这种手段提出了两种方法,一种是将网络中有关联的节点,进行点和边的处理;一种是匿名的处理信息中比较私密的部分。这项技术需要花时间整理信息属性,然后结合图结构和部分已知信息来推断需要匿名的信息。

4.3 数据水印技术

水印技术就是将一些可标识的信息,满足不影响原始数据的条件时,将水印标志植入到原始数据中,以此来保护数据安全。常见的是在一些图片或者文本中应用水印技术,但是水印是有强弱之分的,对于脆弱的水印是可去除的,并且水印技术多应用在静态数据中的,对于动态的数据处理还存在进步空间。

4.4 访问控制技术

访问控制权限主要在网络方面发生作用,访问控制技术对资源设置可访问权限,这项技术广泛用于角色访问控制,角色作为用户的代理人,用于和权限进行沟通,当用户需要对信息进行读取时,需通过身份和权限验证,获取资源。

4.5 数据加密技术

数据加密技术作为网络安全的基础,是通过加密钥匙和加密函数来对数据进行处理,从而让原始数据得以保护,只有通过正确的解密函数和密钥处理才能读取到正确的数据。由于需要密钥的处理才能获取正确信息,所以这项技术对于密钥的保护成为一个安全问题。

4.6 数据溯源技术

数据溯源技术类似于日志,讲究的是一种溯本追源的手段,依照跟踪路径重新展示数据的历史情况和改变

的过程,从而达到数据历史文档的追溯效果。^[5]但是将这项技术用于数据安全和隐私保护还存在安全隐患,因为数据本身就具有隐私性,记录数据的源头对于用户而言是敏感的。

5.大数据安全和隐私保护措施

因为数据泄露事件不停的发生,让国家关注到了数据安全和隐私保护问题,因此国家在不断完善相关法律条例。不仅是国家,数据安全和隐私保护问题也应该得到人们和企业重视。

5.1 增加个人防范意识

人们需要增强对个人隐私的防范意识,要正确的认识到在网络中,个人信息是不能随意透漏的,要学会辨别网络中的真伪。从自身出发,杜绝非法分子对个人数据进行不正当的操作。个人隐私,最有力的保护是国家法律,完善个人隐私保护法,让人们充分认识到个人隐私泄露会带来的危害。国家广泛的传输保护个人隐私的思想,增强人们对自身的保护意识,从根源上杜绝个人隐私的暴露。

5.2 避免用户隐私泄露

企业要负责任的管理好用户数据,不能因为自身利益随意泄露用户的隐私,企业要使用漏洞较少的软件产品,对于新技术带来的数据安全问题要做好防范措施。企业对于人员的管理要进一步优化,要让企业员工充分认识到个人隐私是受到法律保护的,增加员工的道德责任意识从而避免用户数据的泄露。

6.总结

大数据时代不仅诞生了各种新型的技术,还伴随着许多数据安全和隐私保护问题,在支持技术发展的同时还要保护好个人的信息安全。本文简述了大数据带来好处以及大数据伴随的危机,介绍了五种相关技术对数据进行保护,让人们和企业同时行动起来保护个人隐私,从而避免和防范个人数据泄露事件的出现,才利于数据时代的发展。

参考文献

[1] 姚雨晴. 大数据环境下个人信息保护研究[D]. 华中师范大学,2018.

- [2]徐乐. 大数据时代隐私安全问题研究[D].成都理工大学,2016.
- [3]冯登国,张敏,李昊.大数据安全与隐私保护[J].计算机学报,2014,37(01):246-258.
- [4]李俊慧. Facebook5000 万用户数据泄露 国内巨头需引以为戒[N]. 第一财经日报,2018-10-15(A10).
- [5]明华, 张勇, 符小辉. 数据溯源技术综述[J]. 小型微型计算机系统, 2012, 33(9):1917-1923.

作者介绍

第一作者：苟丹（1997-），女，汉，四川省泸州市，本科，四川大学锦城学院，研究方向：大数据

第二作者（通讯作者）：王科（1985—），男，讲师，硕士研究生，四川大学锦城学院，研究方向：云计算

第三作者：李晨曦（1998-7），男，汉，贵州省贵阳市，本科，四川大学锦城学院，研究方向：大数据技术开发