第5章

# 大数据安全

**学习目标**

- 了解大数据安全需求。
- 了解大数据安全与传统数据安全的区别。
- 了解大数据面临的安全问题。
- 了解大数据安全典型案例。
- 掌握大数据安全常见框架及技术。
- 理解大数据保护的原则与对策。
- 了解大数据安全相关法律法规。

本章围绕大数据安全主题，介绍大数据安全基础、安全问题、典型案例、大数据安全技术、大数据保护的原则与对策以及大数据安全相关法律法规等，旨在帮助读者树立大数据安全意识，掌握大数据技术应用中的安全对策。

## 5.1 大数据安全基础


视频讲解

### 5.1.1 大数据安全需求

数据安全从本质上讲就是保护信息系统或信息网络中的数据资源免受各种类型的威胁、干扰和破坏，即保证数据的安全性。关于大数据安全，根据 2020 年 3 月 1 日正式实施的国家标准《信息安全技术　大数据安全管理指南》(GB/T 37973—2019)中的规定，大数据安全需求包括保密性、完整性、可用性和其他需求 4 个方面。

**1. 保密性**

大数据环境下的保密性需求应考虑以下几个方面：

(1) 数据传输的保密性，使用不同的安全协议保障数据采集、分发等操作中的传输保密要求。

(2) 数据存储的保密性，例如，使用访问控制、加密机制等。

(3) 加密数据的运算，例如，使用同态加密等算法。

（4）数据汇聚时敏感性保护，例如，通过数据隔离等机制确保汇聚大量数据时不暴露敏感信息。

（5）个人信息保护，例如，通过数据匿名化使得个人信息主体无法被识别。

（6）密钥的安全，应建立适合大数据环境的密钥管理系统。

**2. 完整性**

大数据环境下的完整性需求应考虑以下几个方面：

（1）数据来源验证，应确保数据来自已认证的数据源。

（2）数据传输完整性，应确保大数据活动中的数据传输安全。

（3）数据计算可靠性，应确保只对数据执行了期望的计算。

（4）数据存储完整性，应确保分布式存储的数据及其副本的完整性。

（5）数据可审计，应建立数据的细粒度审计机制。

**3. 可用性**

大数据环境下的可用性需求应考虑以下方面：

（1）大数据平台抗攻击能力。

（2）基于大数据的安全分析能力，如安全情报分析、数据驱动的误用检测、安全事件检测等。

（3）大数据平台的容灾能力。

**4. 其他需求**

大数据安全除了考虑信息系统的保密性、完整性和可用性，还应针对大数据的组织特点和大数据活动的其他方面分析安全需求，包括但不限于：

（1）与法律法规、国家战略、标准的合规性。

（2）可能产生的社会和公共安全影响，与文化的包容性。

（3）跨组织的数据共享。

（4）跨境数据流动。

（5）知识产权保护及数据价值保护。

### 5.1.2 大数据安全与传统数据安全的区别

与传统数据安全所面临的威胁不同，在大数据时代，数据通过交流、共享等方式，使其价值得到更大程度的体现和提升，数据的动态利用逐渐走向常态化、多元化，这都使大数据安全表现出与传统数据安全相对不同的特征，具体表现在以下几个方面。

**1. 数据安全保护难度大**

在网络空间中，数据越多，受到的关注度也就越高。大数据因为其巨大的数据量，从而更容易成为网络攻击的目标。一方面，大数据不仅数据量巨大，而且可能包含大量敏感数据，因此对攻击者具有较大的吸引力。一旦其安全屏障被攻破，攻击者就能一次性获得较大的收益。另一方面，分布式的系统部署、开放的网络环境、复杂的数据应用和众多的用户访问，使得大数据在保密性、完整性和可用性等方面面临更大的挑战。

**2. 隐私泄露风险加大**

传统上的数据是企业的资产，只在企业内部环境中使用，没有太大的流动性，因此数据中的个人隐私泄露风险较小。但在大数据时代，由于大数据系统中普遍保存有大量的个人

信息，在发生数据滥用、网络攻击等安全事件时，常会伴随着个人信息的泄露。而且，随着数据挖掘、机器学习以及人工智能等技术的广泛应用，基于大数据的分析能力也越来越强大。个人的日常生活轨迹、购物消费等都会在各式各样的数据系统中留下“数字脚印”，这些信息单独来看没有太大的价值，但一旦将多源数据进行综合分析，就很容易挖掘出更多的个人信息，还原和预测个人生活的轨迹，加剧个人隐私的泄露。哈佛大学的研究显示，只要知道一个人的年龄、性别和邮编，就可以在公开数据库中识别出此人87%的身份数据。因此，在大数据时代，既要注意防止因数据丢失而直接导致的个人信息泄露，也要注意防止因挖掘分析而间接导致的个人信息泄露，要实现这种综合保护需求难度极大。

更重要的是，在大数据时代，无法禁止外部人员对公开、半公开的信息进行挖掘。而各个社交网站均不同程度地开放其所产生的实时数据，这些都加大了个人隐私泄露的风险。

**3. 跨境数据流动带来监管困难**

数据跨境流动是大数据的一个特殊属性。在当前这个信息爆炸的时代，大量的数据不仅在国内流动，还会进行跨境的流动，例如，跨境购物、跨境汇款、机票酒店预订、签证办理等均会产生跨境数据流动。当前，我国已成为数字贸易大国和数据资源大国，拥有超大规模市场、海量数据资源和丰富的应用场景等优势。作为连接全球数字经济的纽带，数据跨境流动虽然为全球化带来了新动能，但也给国家安全、数据主权、隐私保护等带来诸多挑战。例如，发生跨境数据安全事件后，无法进行有效的追溯和审计；一旦敏感数据流向境外，国家安全、社会稳定、人民生活等必然受到影响。

因此，大数据安全已经作为非传统安全因素，重新定义了大国博弈的空间。如何对出境数据进行安全评估、促进数据跨境安全有序自由流动是大数据安全亟须解决的重要课题。

**4. 大数据技术成为新型网络攻击手段**

大数据为企业带来商业价值的同时，也可能被攻击者利用来攻击企业，给企业造成损失。例如，攻击者会收集企业的各种信息，并将收集到的信息进行关联分析进而对企业进行精准画像，从而为攻击企业提供数据支持。

**5. 大数据技术成为高级可持续威胁攻击的载体**

攻击者利用数据挖掘和分析等大数据技术可以发起高级可持续威胁（Advanced Persistent Threat，APT）攻击。相较于基于威胁特征进行实时匹配监测的传统安全监测机制，APT作为一个实施过程，并不具备能够被实时监测出来的明显特征。因此通过将APT代码嵌入隐藏在大数据中，利用大数据技术发起僵尸网络攻击，能够控制海量傀儡机发起更大规模的网络攻击，从而严重威胁网络信息安全。

## 5.2 大数据面临的安全问题

在大数据时代，由于各种大数据系统发展变化快、应用场景广，大数据面临着诸多的安全问题，包括技术层面的安全威胁和国家社会层面的安全挑战。

### 5.2.1 大数据在技术层面的安全威胁

大数据安全威胁渗透在数据生产、采集、处理和共享等大数据产业链的各个环节，风险成因复杂交织，主要包括以下几个方面。

**1. 大数据基础设施安全威胁**

为提供对大数据应用的支持，需要创建支持大数据环境的基础设施。大数据基础设施包括存储设备、运算设备、一体机和其他基础软件等。例如，需要高速的网络来收集各种数据源，需要大规模的存储设备对海量数据进行存储，还需要各种服务器和计算设备对数据进行分析和应用，这些基础设施均具有虚拟化和分布式性质。这些基础设施在给用户带来各种大数据新应用的同时，也会遭受各种安全威胁。

针对大数据基础设施常见的安全威胁包括非授权访问、信息泄露或丢失、网络数据传输过程中的数据完整性破坏、拒绝服务攻击、网络病毒传播以及利用漏洞入侵系统等。

**2. 传统安全措施难以适配**

大数据的海量、多源、异构等特征，导致其与传统封闭环境下的数据应用安全环境有很大的区别。

一方面，大数据的整体技术架构要相对更加复杂。大数据应用一般采用底层复杂、开放的分布式计算和存储架构，为其提供海量数据的分布式存储和高效计算服务。这种技术架构使得大数据应用的系统边界变得模糊，传统基于边界的安全保护措施难以发挥作用。例如，在大数据系统中，数据一般都是分布式存储的，数据可能动态分散在很多个不同的存储设备，甚至不同的物理地点进行存储，这就导致难以准确划定传统意义上的每个数据集的“边界”，传统的基于网关模式的防护手段也就失去了安全防护效果。

另一方面，大数据系统表现为系统的系统（System of System），其分布式计算安全问题也相对更加突出。在分布式计算环境下，计算涉及的软件和硬件较多，任何一点出现故障或遭受攻击都可能导致整体安全出现问题。攻击者可以从防护能力最弱的节点着手进行突破，通过破坏计算节点、篡改传输数据和渗透攻击，最终达到破坏或控制整个分布式系统的目的。这都将使传统基于单点的认证鉴别、访问控制和安全审计的手段面临巨大的挑战。

此外，传统的安全检测技术能够将大量的日志数据集中到一起，进行整体性安全分析，以试图从中发现安全事件。然而这些安全检测技术往往存在误报过多的问题。随着大数据系统的建设，日志数据规模增大，数据的种类将更加丰富，过多的误判会造成安全检测系统失效，降低安全检测能力。因此，在大数据环境下，大数据安全审计检测方面也面临着巨大的挑战。随着大数据技术的应用，为了保证大数据安全，需要进一步提高安全检测技术能力，提升安全检测技术在大数据时代的适用性。

**3. 平台安全机制严重不足**

现有的大数据应用多采用开源的大数据管理平台和技术，如基于 Hadoop 生态架构的 HBase/Hive、Cassandra/Spark、MongoDB 等。这些平台和技术在设计之初考虑的是在可信的内部网络中使用，对大数据应用用户的身份鉴别、授权访问以及安全审计等安全功能需求考虑较少。近年来，随着技术的更新和发展，这些软件通过调用外部安全组件、修补安全补丁的方式逐步增加了一些安全措施，如调用外部 Kerberos 身份鉴别组件、扩展访问控制管理能力、允许使用存储加密以及增加安全审计功能等。但大部分大数据软件仍然是围绕大容量、高速率的数据处理功能开发，缺乏原生的安全特性，在整体安全规划方面考虑不足，甚至没有良好的安全实现。

同时，在大数据系统建设过程中，现有的基础软件和应用多采用第三方的开源组件。这些开源系统本身功能复杂、模块众多、复杂性很高，因此对使用人员的技术要求较高，稍有不

慎，就有可能导致系统崩溃或数据丢失。在开源软件的开发和维护过程中，由于管理松散、开发人员混杂，软件在发布前几乎没有经过权威和严格的安全测试，使得这些软件大都缺乏有效的漏洞管理和恶意后门防范能力。如 2017 年 6 月，Hadoop 的发行版本被发现存在安全漏洞，由于该软件没有对输入进行严格的认证，攻击者可以利用该漏洞攻击系统，并获得最高管理员权限。

物联网技术的快速发展，使得当前设备连接和数据规模都达到了前所未有的程度，不仅手机、计算机、电视机等传统信息化设备已连入网络，汽车、家用电器、工厂设备、基础设施等也逐步成为互联网的终端。在这些新终端的安全防护上，现有的安全防护体系尚不成熟，有效的安全手段还不多，急需研发和应用更好的安全保护机制。

**4. 应用访问控制愈加困难**

大数据应用的特点之一就是数据类型复杂、应用范围广泛，它通常要为来自不同组织或部门、不同身份与目的的用户提供服务。因而随着大数据应用的发展，其在应用访问控制方面也面临着巨大的挑战。

首先是用户身份鉴别。大数据只有经过开放和流动，才能创造出更大的价值。目前，政府部门、央企及其他重要单位的数据正在逐步开放，或开放给组织内部不同部门使用，或开放给不同政府部门和上级监管部门使用，或开放给定向企业和社会公众使用。数据的开放和共享意味着会有更多的用户可以访问数据。大量的用户以及复杂的共享应用环境，导致大数据系统需要更准确地识别和鉴别用户身份，而传统基于集中数据存储的用户身份鉴别难以满足安全需求。

其次是用户访问控制。目前常见的用户访问控制是基于用户身份或角色进行的，而在大数据应用场景中，由于存在大量未知的用户和数据，预先设置角色及权限十分困难。即使可以事先对用户权限分类，但用户角色众多，难以精细化和细粒度地控制每个角色的实际权限，从而导致无法准确地为每个用户指定其可以访问的数据范围。

最后是用户数据安全审计和追踪溯源。针对大数据量时的细粒度数据审计能力不足，用户访问控制策略需要创新。当前常见的操作系统审计、网络审计、日志审计等软件在审计粒度上较粗，不能完全满足复杂大数据应用场景下审计多种数据源日志的需求，难以达到良好的溯源效果。

**5. 大数据滥用和误用风险巨大**

计算机网络技术和人工智能的发展，为大数据自动收集以及智能动态分析提供了方便。但是，大数据技术被滥用也会带来安全风险。一方面，大数据本身的安全防护存在漏洞，对大数据的安全控制力度仍然不够，API 访问权限控制以及密钥生成、存储和管理方面的不足都可能造成数据泄露；另一方面，攻击者也在利用大数据技术进行攻击，例如，黑客能够利用大数据技术最大限度地收集更多用户的敏感信息。

大数据的质量、准确性以及对其的使用也存在风险。例如，从社交媒体获取个人信息的准确性，基本的个人资料（例如年龄、婚姻状况、教育或者就业情况等）通常都是未经认证的，分析结果可信度不高。此外，从公众渠道收集到的信息，可能与需求相关度较小。这些数据的价值密度较低，如果对其进行分析和使用，可能产生无效的结果，从而导致错误的决策。

### 5.2.2 大数据在国家社会层面的安全挑战

大数据作为一种社会资源，在给互联网领域带来技术变革的同时，也对全球的政治、经济、军事、文化、生态等产生了巨大影响。大数据已经成为衡量综合国力的重要标准，事关国家主权和安全，必须加以高度重视。

**1. 国家安全深受大数据影响**

大数据意味着海量数据，也意味着更复杂、更敏感的数据，特别是关系国家安全和利益的数据，例如，国防建设数据、军事数据、外交数据等，极易成为网络攻击的目标。

信息技术与经济社会的交汇融合引发了数据的迅猛增长，数据已成为国家基础性战略资源。同时，大数据的应用范围越来越广泛，国家的政治、经济、军事、文化等各个领域都离不开数据和数字基础设施。各类大数据平台承载着海量的数据资源，其中不乏大量敏感资源和重要数据，必然会成为包括黑客在内的各类敌对势力对一个国家进行网络攻击的重要目标。实际上，各类数据已经成为一些不法分子和敌对势力用来策划、实施、推动各种违法犯罪活动的理想工具，对国家安全和社会稳定造成了巨大的威胁。上升到国家战略层面，涉及国计民生的关键信息基础设施的大数据资源一旦遭到破坏，就会在政治、经济、军事等各领域造成巨大的损失。

面对汹涌的数据洪流，站在国家安全的角度来思考和研究大数据安全，已经成为一个紧迫而现实的挑战。大数据全球化、开放化的特点，使国家的“信息边疆”不断拓展和延伸。大数据安全和国家安全息息相关，没有大数据安全，就没有真正意义上的国家安全。

**2. 社会治理面临大数据挑战**

从我国信息化发展的现实情况看，“不敢共享开放”“不会共享开放”的情况依然较为普遍。相关人员担心数据共享开放会引起信息安全问题，担心数据泄密和失控。因此，加强大数据环境下的网络安全问题研究和基于大数据的网络安全技术研究，建立健全大数据安全保障体系，切实保证数据安全，才能确保大数据“敢共享开放”和“会共享开放”，才能真正促进社会发展。

此外，社会治理创新是我国应对社会转型、化解社会矛盾、协调利益关系所面临的一项重大战略任务。针对目前社会治理领域普遍存在的一些问题，大数据技术通过对海量数据的快速收集与挖掘、及时研判与共享，成为支持社会治理科学决策和准确预判的有力手段，为转型期的社会治理带来了新机遇。而现实问题是，在大数据时代，可以说每个人都是数据的制造者、传递者和消费者，大量现实问题在虚拟的网络环境中讨论和传播，其中不乏大量的误导、篡改及谣传的信息。一方面，这些虚假、错误的信息进入社会治理的数据集后，将会误导基于大数据的科学决策，影响社会治理重点和效果；另一方面，若虚假、错误的信息不被及时发现和处理，极有可能带来恶劣的负面效果，甚至导致社会群体性事件。因此，如何甄别大数据中虚假和错误的信息给社会治理带来了巨大挑战。

## 5.3 大数据安全典型案例

近年来，关于大数据安全的事件层出不穷，下面对国内外大数据安全典型案例进行简单的介绍。

### 5.3.1 Facebook 数据门事件

2019 年 7 月 24 日，美国联邦贸易委员会宣布对美国社交网络公司 Facebook 开出高达 56 亿美元（约合 385 亿元人民币）的罚单，理由是 Facebook 在商业活动中存在“滥用用户隐私数据”的行为，这也是美国政府对科技公司开出的最高金额罚单。

美国联邦贸易委员会表示，经过调查后发现，Facebook 公司的数据政策“欺骗了数千万使用 Facebook 面部识别工具的用户”，并且私自将收集到的用户手机号码等隐私信息用于商业广告活动——这些行为违反了美国法律中有关反欺诈的相关规定。

事件的起因是一家名为剑桥分析（Cambridge Analytica）的数据分析公司收集了 5000 万美国选民的 Facebook 个人资料，并利用这些资料构建了一个强大的软件程序来预测和影响投票箱中的选择结果。据称，该数据分析公司与特朗普赢得大选、英国成功脱欧的竞选团队均有合作。这是 Facebook 这家科技公司有史以来最大的数据泄露和滥用事件之一。Facebook 和剑桥分析公司的标志如图 5-1 所示。




图 5-1 Facebook 和剑桥分析公司的标志

长期以来，Facebook 开放 API 让外部第三方公司在 Facebook 平台上提供心理测验或者小游戏，此举增加了 Facebook 此类社交平台的内容丰富度，也通过用户参与分享进一步强化了用户之间社交关系的黏度，但也为第三方公司违规收集客户信息提供了平台。本次事件的主体剑桥分析公司成立于 2013 年，由美国共和党的亿万富豪罗伯特·默瑟出资，并由曾经担任特朗普顾问的史蒂夫·班农主导成立。这家公司成立的目的非常明确，就是要通过海量数据分析研究，进一步操纵社会舆论风向，借此影响 2014 年的美国国会大选，以及 2016 年的美国总统选举。该公司与当时任职于剑桥大学心理测量中心的亚历山大·科根进行合作。他开发了一款名为“这是你的数字化生活”（this is your digital life）的心理测试程序，用户会通过参与这样的心理测试，得出一个有趣的结果。用户可以选择将测试结果分享在自己的 Facebook 页面上，进而吸引其他 Facebook 好友注意，甚至跟进参与。

通过 Facebook 发布的这款测试程序总共有 27 万人参与，并在参与过程中提供了个人相关数据资料，而通过其所延伸连接的社交网络，剑桥分析公司得以进一步扩大得到了多达 5000 万用户的数据资料，而后基于这些资料，构建心理与行为分析模型，其中包括性格倾向、生活满意度、政治观点，甚至个人喜好等明确可用的资讯指标结果。整起事件最可怕之处不只在于数据泄露和滥用的风险问题，而在于身负为 20 亿用户数据资料把关责任的 Facebook 从头到尾都知道这件事。而剑桥分析公司至少参与了各国超过 200 场竞选，其中包括尼日利亚、肯尼亚、马来西亚、捷克、墨西哥、印度和阿根廷等。在这些国家的选举中，剑桥分析公司使用大量的个人数据来构建信息分析图，以确定选民的政治倾向和宗教信仰、性

取向、肤色等，这些分析结果被用于改变选民的选举倾向，最终影响选举的结果。

### 5.3.2 手机 App 过度采集个人信息

个人信息买卖已经形成一条大规模的产业链，这条产业链结构完整、分工细化，个人信息被明码标价。个人信息泄露的一个主要途径就是经营者未经本人同意暗自收集个人信息，然后将这些信息泄露、出售给他人。

手机 App 最常见的方式是在自身功能不必要的情况下获取用户隐私权限和采集个人信息。许多手机 App 在安装时，会弹出要求用户授权各类信息权限的条款，包括通讯录、短信、摄像头、麦克风、地理位置等，甚至像天气预报、手电筒这类功能单一的手机 App，在安装协议中也提出要读取通讯录。在这种情形下，大多数用户只能接受这些“霸王条款”，选择提供个人信息，否则就无法使用该手机 App。还有一些手机 App 干脆采用“隐形套路”，在安装时未经用户同意，就自动选择授权使用相关的个人信息，例如，某些 App 在用户协议中将“同意”选项设置为较小字体，且已经预先勾选，导致部分消费者在未知情况下进行了授权。于是，用户的个人身份信息、账户信息、常用设备信息、位置信息等，都可能被手机 App 采集。某款 App 请求定位手机权限的界面如图 5-2 所示。




图 5-2 某款 App 请求定位手机权限的界面

此外，在微信朋友圈广泛传播的各种测试小程序也可能在窃取用户个人信息。在小程序授权登录测试页面时，微信、QQ 号、姓名、生日、手机号等很多个人信息都会被测试程序的后台获得，这些信息很可能被用作商业用途，给使用者的切身利益造成损失。同时，不法分子还设计了更加隐蔽的个人信息获取方式，例如，制作多种测试小程序在微信朋友圈进行分发：有的测试小程序负责收集参与测试用户的个人喜好，有些测试小程序负责收集用户的收入水平，有些测试小程序负责收集用户的朋友关系。虽然用户参与某个测试只是提供了部分个人信息，但是当用户长期以来参与了多个测试以后，不法分子就可以获得该用户较为全面的个人信息。

### 5.3.3 Oracle 云数据泄露事件

2025 年 3 月 20 日，一位名为“rose87168”的黑客发帖宣称，已成功入侵至少两个 Oracle Cloud 客户的登录系统，并窃取了包括加密密码、Java 密钥库（Java Keystore，JKS）、轻型目录访问协议（Lightweight Directory Access Protocol，LDAP）记录、安全证书等敏感数据在内的约 600 万条记录。随后，rose87168 开始在暗网兜售数据，并对受影响的企业发起敲诈勒索。按照该黑客的说法，此次数据泄露涉及 14 万 Oracle Cloud 客户，涉及银行、食品、教育、金融、医疗等各领域，名单中不仅有美国企业，还有包括超过 1000 个使用“.cn”域名的中国企业/个人在内的各国用户，其中不乏多家知名上市公司。毫无疑问，这是一起全球性的严重 IT 事故。而事故发生的根本原因是 Oracle 老旧系统存在的安全隐患：本次事故中，涉事服务器运行的是已停产的 Oracle Fusion Middleware 11g，该系统已超过 10 年未更新。而被攻击者利用的 CVE-2021-35587 漏洞早在 2022 年就已被美国网络安全和基础设施安

全局(Cybersecurity and Infrastructure Security Agency,CISA)列为已知可利用漏洞,但Oracle并未及时修复遗留系统。

数据泄露事件发生后,Oracle先是公开声明"Oracle Cloud未发生任何泄露",将责任推卸至"已停用8年的旧版系统",并称泄露数据来自未迁移至新云平台的旧环境。对此,网络安全专家凯文·比蒙特指出:Oracle试图通过"非常具体的措辞"逃避责任,例如,将问题限定在"Oracle Cloud"而非整体服务,但客户数据实际仍由Oracle管理。

随后Oracle又被曝出通过删除http://Archive.org等平台的历史页面记录,试图消除旧系统架构未更新的证据。此外,Oracle还拒绝公开事件相关漏洞的技术细节和修复进展。Oracle要求Archive.org删除网站时光机留存的页面记录如图5-3所示。




图5-3 Oracle要求Archive.org删除网站时光机留存的页面记录

研究人员通过技术手段恢复被删除的页面,证实Oracle部分核心系统超过10年未更新,且这些系统仍存储着用户敏感数据。

事件彻底曝光后,Oracle又试图将攻击归因于第三方组件(如存在缺陷的Java服务),以转移对自身安全管理失职的指责。然而安全公司CloudSEK的后续分析指出,黑客利用的漏洞在Oracle自身产品Oracle Fusion Middleware中,与第三方无关。

Oracle的应对显然引起了各方不满,安全专家丽莎·福特等人批评Oracle的"逃避文化"不可接受,硅谷评论则直指Oracle"贩卖技术垃圾"。针对Oracle的行为,佛罗里达州、得克萨斯州等地相继对Oracle发起集体诉讼,指控其泄露原告方的敏感数据。诉状称,尽管在隐私政策中Oracle承诺"毫不拖延"地向客户报告所有数据泄露事件,但被告并未履行其数据披露义务。

### 5.3.4 免费Wi-Fi带来数据安全风险

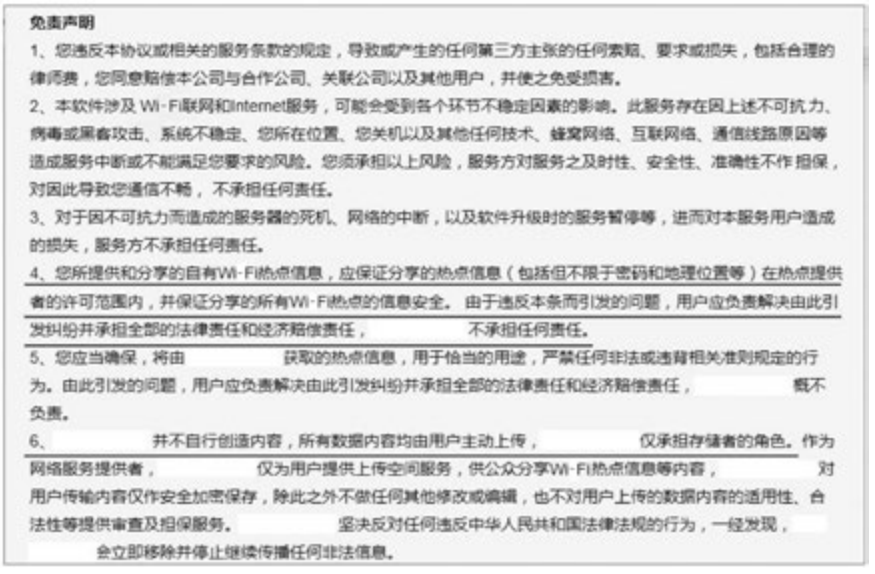

视频讲解

随着移动互联网的发展,上网似乎成为公众生活中难以分割的一部分。面对随时随地的上网需求,免费的Wi-Fi网络已经成为机场、车站、宾馆、饭店以及各类商铺的标准配置,而一些蹭网App也为用户提供了免费接入他人Wi-Fi网络的功能。大家无论走到哪里,总是喜欢先询问商家或通过App寻找是否有免费的Wi-Fi接入。但在公共场所免费Wi-Fi接入的背后,其实存在着不小的信息数据安全风险。具体包括以下几种情况。

**1. 蹭网App入侵他人Wi-Fi网络**

在旺盛的Wi-Fi接入需求下,如"Wi-Fi万能钥匙""雷达Wi-Fi"等手机应用程序受到手机用户的青睐。这些App打着"人人为我,我为人人"的旗号,让用户无需密码即可连接他

人的无线网络，尤其是家庭私有无线网络，极大地方便了用户的 Wi-Fi 接入。其实现原理是基于 Wi-Fi 密码共享原则，即当用户连接到某一 Wi-Fi 网络后，App 会将该 Wi-Fi 的密码自动上传至 App 数据库进行全国共享，使任何一个素未谋面的陌生人只要使用该 App 即可连接到此 Wi-Fi 网络。而这一切，往往是在 Wi-Fi 主人不知情的情况下进行的，即一旦有手机安装了蹭网 App 的朋友和客户曾连接你的 Wi-Fi 网络，那就意味着你的 Wi-Fi 网络密码已经进行了全国共享。某蹭网 App 用户协议中的免责声明部分如图 5-4 所示。


免责声明

1、您违反本协议或相关的服务条款的规定，导致或产生的任何第三方主张的任何索赔、要求或损失，包括合理的律师费，您同意赔偿本公司与合作公司、关联公司以及其他用户，并使之免受损害。

2、本软件涉及 Wi-Fi联网和Internet服务，可能会受到各个环节不稳定因素的影响。此服务存在因上述不可抗力、病毒或黑客攻击、系统不稳定、您所在位置、您关机以及其他任何技术、蜂窝网络、互联网络、通信线路原因等造成服务中断或不能满足您要求的风险。您须承担以上风险，服务方对服务之及时性、安全性、准确性不作担保，对因此导致您通信不畅，不承担任何责任。

3、对于因不可抗力而造成的服务器的死机、网络的中断，以及软件升级时的服务暂停等，进而对本服务用户造成的损失，服务方不承担任何责任。

4、您所提供和分享的自有Wi-Fi热点信息，应保证分享的热点信息（包括但不限于密码和地理位置等）在热点提供者的许可范围内，并保证分享的所有Wi-Fi热点的信息安全。由于违反本条而引发的问题，用户应负责解决由此引发纠纷并承担全部的法律责任和经济赔偿责任，不承担任何责任。

5、您应当确保，将由获取的热点信息，用于恰当的用途，严禁任何非法或违背相关准则规定的行为。由此引发的问题，用户应负责解决由此引发纠纷并承担全部的法律责任和经济赔偿责任，概不负责。

6、并不自行创造内容，所有数据内容均由用户主动上传，仅承担存储者的角色。作为网络服务提供者，仅为用户提供上传空间服务，供公众分享Wi-Fi热点信息等内容，对用户传输内容仅作安全加密保存，除此之外不做任何其他修改或编辑，也不对用户上传的数据内容的适用性、合法性等提供审查及担保服务。坚决反对任何违反中华人民共和国法律法规的行为，一经发现，会立即移除并停止继续传播任何非法信息。

图 5-4　某蹭网 App 用户协议中的免责声明

从图 5-4 中的免责声明可以明显看出该 App 会对用户接入 Wi-Fi 的密码自动进行上传共享。这种行为不仅涉嫌入侵他人 Wi-Fi 网络，还有可能导致个人信息被窃取，因为网主和蹭网者连接在同一个局域网，双方都有泄露网购账号、住址、电话号码、支付密码等重要个人信息的风险，很容易造成财产损失。如果不法分子通过蹭网软件或硬件设备连接网主网络，通过技术手段以其身份发布一些违法内容，网主很可能还需要承担相应的责任。

由于蹭网 App 是基于密码共享而非密码破解，因此要杜绝自己的 Wi-Fi 被蹭网，靠提升密码的复杂度是行不通的，因此对于家庭私有无线网络而言，最有效的方法是设置白名单，从访问控制的角度限制陌生终端的接入。

**2. 免费 Wi-Fi 窃取用户信息**

不法分子会利用用户免费接入 Wi-Fi 网络的需求进行 Wi-Fi 钓鱼。作为一种网络攻击手段，Wi-Fi 钓鱼通过创建虚假的无线网络来欺骗用户连接，进而窃取敏感信息。

Wi-Fi 钓鱼从技术实现上并不复杂，只需要一台计算机、一套无线网络设备和一个网络报文分析软件，就能搭建出一个 Wi-Fi 钓鱼网络。不法分子会在公共场所设置免费钓鱼 Wi-Fi 信号，当用户的手机或笔记本连接上这些免费网络后，通过数据的传输，黑客就能轻松盗取用户手机、笔记本里的照片、电话号码、各种账号密码等个人隐私信息。

2024 年 4 月，澳大利亚一名男子因涉嫌通过伪造免费的 Wi-Fi 网络，窃取毫无戒心的用户的个人数据而被指控。该男子在多个地点，包括珀斯、墨尔本和阿德莱德的机场，以及

多趟航班上，使用设备创建了合法 Wi-Fi 网络的副本。当用户尝试连接到这些网络时，他们会被引导到一个假网页，并被要求使用电子邮件或社交媒体账号登录。警方称，这些登录信息随后被保存到该男子的设备上，并被用于访问更多的个人信息，包括受害者的在线通信内容、存储的图片和视频或银行账户信息。

除不法分子的 Wi-Fi 钓鱼外，许多商家也在 Wi-Fi 上做文章，他们通过 Wi-Fi 后台记录的上网者手机号码等联系信息，更加有针对性地进行广告短信等的投放，以达到精准营销、招揽客户的目的。因此，许多顾客在使用了某 Wi-Fi 之后会收到大量的广告信息，甚至自己的手机号码也会被当作信息进行多次买卖。

**3. Wi-Fi 探针窃取用户信息**

当用户手机 Wi-Fi 功能处于开启状态（不需要连接任何 Wi-Fi 热点）时，会向周围发出 Probe Request 帧以寻找 Wi-Fi。一些公司会利用 Wi-Fi 的这一连接特征在公共场所（诸如商场、超市、写字楼等地）部署“Wi-Fi 探针”。当探针盒子发现 Probe Request 信号后，就可以自动识别出手机的 MAC 地址，将其转换成 IMEI 号，然后再转换成手机号码，在用户不知情的情况下搜集个人信息，并通过发送定向广告等进行有针对性的营销推广。

2019 年中央电视台“3·15”晚会对企业通过“探针盒子”窃取用户信息的灰色产业链进行了曝光。但直到目前，依然有此类“探针盒子”在售，并表示此类盒子可以采集周边用户的手机号码，用于“精准营销”。两款在售的典型“探针盒子”如图 5-5 所示。




图 5-5 两款在售的典型“探针盒子”

针对免费 Wi-Fi 和 Wi-Fi 探针窃取用户信息的行为，专家建议互联网用户在登录任何公共 Wi-Fi 网络时应始终保持警惕，在外出时尽量关闭手机或其他电子设备的 Wi-Fi，以防止设备自动连接到热点。

## 5.4 大数据安全框架与技术

在大数据环境下，数据随着业务和应用的需要，突破了传统信息系统的边界，在不同的载体之间进行流通、融合，对应的数据安全体系建设也必须抛弃传统的单点防御模式，转而

采用全局治理的体系化建设思路。大数据安全建设是一个非常复杂的体系，既要考虑数据生命周期的视角，还要考虑业务与数据流转的视角；既要保护数据计算环境和实体，还要精准控制数据的访问使用。面对复杂的安全需求，显然需要构建一套成熟的安全框架，并基于框架通过具体的安全技术去实现。

视频讲解

### 5.4.1 大数据安全框架

当前主流的大数据安全框架包括我国国家标准定义的数据安全能力成熟度模型、Gartner 公司提出的数据安全治理框架以及 Microsoft 公司的隐私、保密和合规性框架，下面分别对其进行介绍。

**1. 数据安全能力成熟度模型（Data Security capability Maturity Model，DSMM）**

为评估和度量组织机构保障大数据安全的能力，2020 年 3 月 1 日国家标准《信息安全技术　数据安全能力成熟度模型》（GB/T 37988—2019）正式实施。该标准给出了一个将数据按照其生存周期从不同的安全能力维度进行综合考量，划分数据安全能力成熟度等级的三维模型。具体如图 5-6 所示。

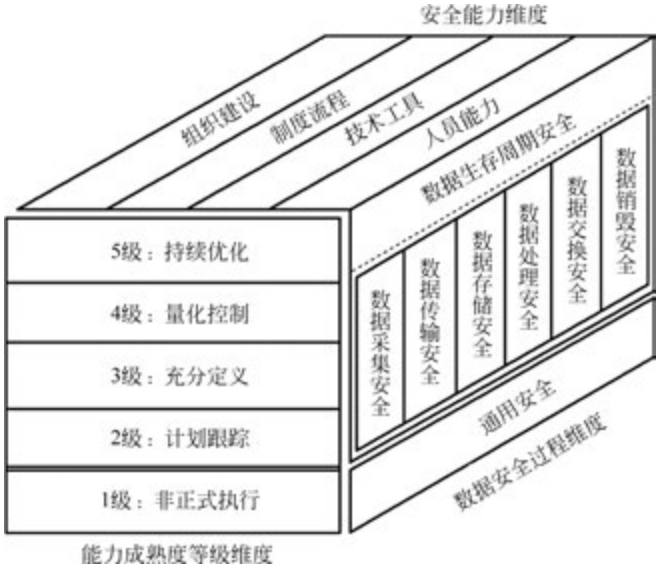



图 5-6　数据安全能力成熟度模型

DSMM 模型由数据安全过程维度、安全能力维度和能力成熟度等级维度共 3 个维度组成。

1）数据安全过程维度

数据安全过程包括数据生存周期安全过程和通用安全过程两部分，共有 30 个过程域（Process Area，PA），每个 PA 由一些基本实践（Base Practice，BP）组成，只有满足过程域中所有的 BP，该 PA 才算符合要求。

数据生存周期安全过程具体包括数据采集安全、数据传输安全、数据存储安全、数据处理安全、数据交换安全和数据销毁安全共 6 个阶段。数据生存周期如图 5-7 所示。

组织内部系统中新产生数据以及从外部系统中收集数据的阶段称为数据采集阶段。数据采集安全阶段包括数据分类分级、数据采集安全管理、数据源鉴别及记录以及数据质量管理 4 个 PA。

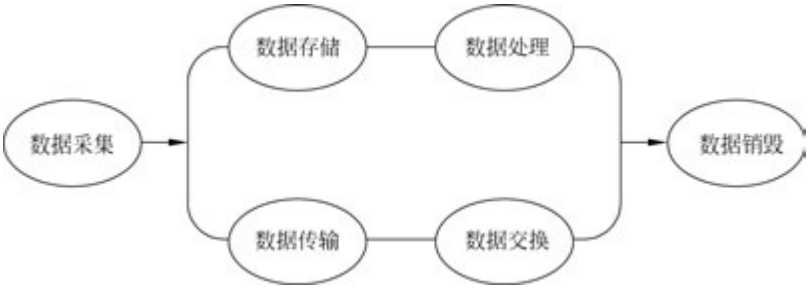



图 5-7 数据生存周期

数据从一个实体传输到另一个实体的阶段称为数据传输阶段。数据传输安全阶段包括数据传输加密、网络可用性管理 2 个 PA。

数据以任何数字格式进行存储的阶段称为数据存储阶段。数据存储安全阶段包括存储介质安全、逻辑存储安全、数据备份和恢复 3 个 PA。

组织在内部对数据进行计算、分析、可视化等操作的阶段称为数据处理阶段。数据处理安全阶段包括数据脱敏、数据分析安全、数据正当使用、数据处理环境安全、数据导入导出安全 5 个 PA。

组织与组织或组织与个人进行数据交换的阶段称为数据交换阶段。数据交换安全阶段包括数据共享安全、数据发布安全、数据接口安全 3 个 PA。

对数据及数据存储介质通过相应的操作手段，使数据彻底删除且无法通过任何手段恢复的过程称为数据销毁阶段。数据销毁安全阶段包括数据销毁处置、存储介质销毁处置 2 个 PA。

需要注意的是，特定的数据所经历的生存周期由实际的业务所决定，可包括完整的 6 个阶段或其中的几个阶段。

通用安全过程包括数据安全策略规划、组织和人员管理、合规管理、数据资产管理、数据供应链安全、元数据管理、终端数据安全、监控与审计、鉴别与访问控制、需求分析、安全事件应急共 11 个 PA。

2）安全能力维度

安全能力维度通过对组织各数据安全过程应具备安全能力的量化，进而评估每项安全过程的实现能力。安全能力分为组织建设、制度流程、技术工具和人员能力 4 个方面。

组织建设从承担数据安全工作的组织应具备的组织建设能力角度，根据数据安全组织架构对组织业务的适用性，数据安全组织承担的工作职责的明确性，数据安全组织运作、沟通协调的有效性等方面进行能力等级区分。

制度流程从组织在数据安全制度流程的建设以及执行情况角度，根据数据生存周期关键控制节点授权审批流程的明确性，相关流程制度的制定、发布、修订的规范性，制度流程实施的一致性和有效性等方面进行能力等级区分。

技术工具从组织用于开展数据安全工作的安全技术、应用系统和工具出发，根据数据安全技术在数据全生存周期过程中的利用情况，应对数据全生存周期安全风险的能力，利用技术工具对数据安全工作的自动化支持能力，对数据安全制度流程固化执行的实现能力等方面进行能力等级区分。

人员能力从组织承担数据安全工作的人员应具备的能力出发，根据数据安全人员所具备的数据安全技能是否能够满足实现安全目标的能力要求（对数据相关业务的理解程度以及数据安全专业能力），数据安全人员的数据安全意识以及对关键数据安全岗位员工数据安全能力的培养等方面进行能力等级区分。

3）能力成熟度等级维度

组织的数据安全能力成熟度划分为 5 个等级。

1 级：非正式执行。随机、无序、被动地执行安全过程，依赖于个人经验，无法复制。

2 级：计划跟踪。在业务系统级别主动地实现了安全过程的计划与执行，但没有形成体系化。

3 级：充分定义。在组织级别实现了安全过程的规范执行。

4 级：量化控制。建立了量化目标，安全过程可度量。

5 级：持续优化。根据组织的整体目标，不断改进和优化安全过程。

在具体分级中，将组织在每个数据安全 PA 的能力成熟度划分为 5 级，针对每个等级下组织应具备的能力要求，从 4 个安全能力维度提出具体的 BP。

其中 3 级要求应包含全部 4 个安全能力，其他等级要求可不包含完整的 4 个数据安全关键能力，当然并非每个安全 PA 的能力成熟度等级都包含完整的 4 个数据安全关键能力。

对于每个数据安全 PA，高等级的能力要求应包括所有低等级能力要求。

**2. Gartner 数据安全治理框架**

作为一家全球知名的信息技术咨询公司，Gartner 在 2018 年下半年发布了数据安全治理（Data Security Governance，DSG）框架。DSG 的理念是数据安全治理是从决策层到技术层，从管理制度到工具支撑，自上而下贯穿整个组织架构的完整链条。它强调根据企业的业务战略、监管合规等诉求，制定数据安全治理战略。数据安全治理战略服务于企业数据战略，而数据战略向上服务于业务战略，由此可以确保数据安全治理过程符合企业业务战略，而不是简单地为安全而安全、为合规而合规的孤立行为。由此，从数据安全治理战略出发，进行威胁与风险分析并实施管控措施，支撑企业业务战略落地。Gartner 数据安全治理框架如图 5-8 所示。

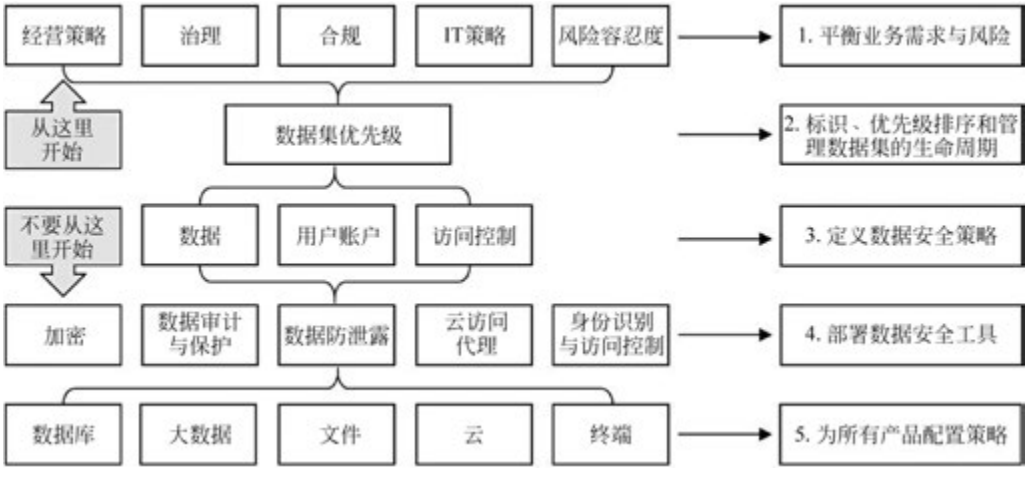



图 5-8　Gartner 数据安全治理框架

DSG 针对数据安全治理给出了一个自上而下的整体框架，分为 5 个步骤：平衡业务需求与风险，标识、优先级排序和管理数据集的生命周期，定义数据安全策略，部署数据安全工具，为所有产品配置策略。平衡业务需求与风险需要平衡经营策略、治理、合规、IT 策略和风险容忍度 5 个维度之间的关系。标识、优先级排序和管理数据集的生命周期需要先明确选择需要治理的对象，优先选择重要数据进行数据安全治理工作，可以通过数据梳理、分类分级等手段，完成数据资产盘点和筛选，在纷杂的数据中理清头绪。定义数据安全策略需要针对数据、用户账户和访问控制给出明确的安全策略。部署数据安全工具根据数据不同生命周期，采取不同的数据安全技术和工具。为所有产品配置安全策略实现数据安全策略的集中管理、同步下发，策略执行对象应包括关系数据库、大数据类型、文档文件、云端数据以及终端等类型。

Gartner 强调，数据安全治理需要从上到下，不能跨越前面的步骤，即应从数据治理阶段介入，而不是被动地等待技术环节的管控。

**3. Microsoft 隐私、保密和合规性框架**

Microsoft 开发了一个针对隐私、保密和合规性的数据治理（Data Governance for Privacy、Confidentiality and Compliance，DGPC）框架，主要用于帮助组织更好地进行数据安全风险控制。该框架围绕人员、流程和技术 3 个核心能力领域进行组织。

人员主要涉及组织架构和人员配置，该框架把数据安全相关组织分为战略层、战术层和操作层，并明确组织中数据安全相关角色职责、资源配置和操作指南。即首先要建立一个由组织内的个人组成的 DGPC 团队，团队中的每个人承担明确的角色和职责，同时要为团队中的人提供足够的资源。该团队可以是一个虚拟组织，团队中的成员需要完成行为原则、政策和流程的定义，安全策略的配置和数据管理的监督等工作。

团队组建完成后，接下来需要定义流程。首先通过查阅组织需要满足的法律法规、标准政策、公司发展战略，充分了解组织在数据安全治理和隐私保护方面需要满足的要求；其次，制定专门用于数据隐私保护的原则和要求；最后，梳理数据流向，探查数据安全、隐私合规的风险，分析风险并采取适当的控制措施。

技术主要涉及技术相关工具，框架提供一种基于安全风险差距分析的技术方法，使用“风险/差距分析矩阵”表格的方式，分析特定数据流并识别存在于流程中的风险问题。该表格由 3 个要素组成，其中数据生命周期为第一维度；关键技术领域和通用控制行为为第二维度。DGPC 风险/差距分析矩阵具体如图 5-9 所示。

信息生命周期分为收集、更新、处理、删除、传输、存储 6 个阶段。

4 个关键技术领域组成一个参考框架，用于判断风险的可接受程度。其中安全基础设施是安全保护的底线，用于减少软硬件设备的内外部入侵危害；身份和访问控制则是为了保护个人信息免受未经授权的访问，需要建立认证机制、数据访问控制机制和用户账户管理机制，实现全流程跟踪管理用户访问过程；数据存放于数据库、文档管理系统、文件服务器之中，并且在生命周期中流动，需要使用信息保护技术对数据分类分级，采用加密等多种手段持续保护；审计和报告是采用监控、自动化审计等方式验证数据访问控制的有效性，发现信息保护流程中的风险点，识别可疑或不合规行为。

通用控制行为需要在信息生命周期的每个阶段同时满足 4 条隐私和保密原则，分别是：在整个信息生命周期内遵守政策规定；减少数据滥用造成的数据机密性风险；减少数据丢

**风险/差距分析矩阵**

| | 安全基础设施 | 身份和访问控制 | 信息保护 | 审计和报告 | 通用控制行为 |
|---|---|---|---|---|---|
| 收集 | | | | | |
| 更新 | | | | | |
| 处理 | | | | | |
| 删除 | | | | | |
| 传输 | | | | | |
| 存储 | | | | | |

图 5-9　DGPC 风险/差距分析矩阵

失造成的数据机密性风险；记录适用的控制措施并验证其有效性。


视频讲解

## 5.4.2　大数据安全技术

大数据安全涉及的主要技术包括身份认证、访问控制、数据加密和日志审计等。

**1. 身份认证**

身份认证是在网络中确认用户身份的有效方法。作为信息安全领域的一种重要手段，能保护信息系统中的数据、服务不被未授权的用户访问。身份认证的目的是确保通信实体身份的真实性和合法性，常用的身份认证方式有以下 5 种。

1）口令认证

基于口令的认证方式是较常用的一种技术。在最初阶段，用户首先在系统中注册自己的用户名和登录口令。系统将用户名和口令存储在内部数据库中，这个口令一般是长期有效的，因此也称为静态口令。为克服静态口令带来的安全隐患，动态口令认证逐渐成为口令认证的主流技术。

2）消息认证

消息认证就是认证消息的完整性，当接收方收到发送方的报文时，接收方能够认证收到的报文是真实的和未被篡改的。它包含两层含义：一是消息源认证，认证信息的发送者是真正的而不是冒充的，即身份认证；二是认证信息在传送过程中未被篡改、重放或延迟等，即消息完整性认证。

3）基于生物特征的认证

生物特征识别技术是指通过计算机与光学、声学、生物传感器和生物统计学原理等技术手段密切结合，利用人体固有的生理特性和行为特征来进行个人身份的鉴定。常用的生物特征包括指纹、人脸、虹膜、声纹等。

4）多因子认证

多因子认证是指用户要通过两种及两种以上的认证机制才能得到授权。例如，用户要插入银行卡、输入登录口令、进行指纹比对，3 种认证均通过后才能获得授权，多因子认证方式可以有效提高安全性。

5）Kerberos 认证体系

大数据环境下，常使用 Kerberos 认证体系为系统提供身份认证技术。Kerberos 是一种计算机网络认证协议，它允许某实体在非安全的网络通信环境下，向另一个实体以一种安全的方式证明自己的身份，可以避免网络实体的身份被仿冒。

**2. 访问控制**

访问控制是数据安全的一个基本组成部分，它规定了哪些人可以访问和使用大数据中海量的信息与资源。制定访问控制策略，可以确保用户身份的真实性，并且确定其相应权限。

访问控制通过认证多种登录凭据以识别用户身份，确认用户身份后，就会赋予其相应级别的访问权限。

访问控制主要有以下 4 种类型，通常会根据组织具体的安全和合规要求，选择对应的方法。

1）自主访问控制

采用自主访问控制，受保护系统、数据或资源的所有者或管理员可以自行设置相关策略，指定有访问权限的成员。

2）强制访问控制

强制访问控制是非自主模型，对象的所有权和能力是由系统而非用户来控制的，由系统根据事先设定的访问控制策略来管理，主要应用在政府或军事等安全性要求较高的环境中。

3）基于角色的访问控制

基于角色的访问控制根据定义的业务功能而非个人用户的身份来授予访问权限。这种方法的目标是为用户提供适当的访问权限，使其只能访问对其在组织内的角色而言有必要的数据。该访问控制是基于角色分配、授权和权限的复杂组合。

4）基于属性的访问控制

通过使用由共同工作的属性组成的策略来授予访问权限。使用属性作为构建块来定义访问控制规则和访问请求，具体通过可扩展访问控制标记语言（eXtensible Access Control Markup Language，XACML）实现。

**3. 数据加密**

数据加密技术是保障数据安全的核心技术之一，主要实现数据的加密和认证功能。常用的加密算法包括对称加密算法（如 DES、AES、SM4 等）、非对称加密算法（如 RSA、ElGamal、SM2 等）和单向哈希函数（如 MD5、SHA 等）。根据加密数据的不同，可以将数据分为静态数据和动态数据。

1）静态数据

静态数据指即使机器关闭后依旧可以存储的数据，包括硬盘、闪存盘、CD、DVD，甚至一些老式软盘或磁带中的数据。

2）动态数据

动态数据指网络通信中的数据，例如，在互联网、Wi-Fi、USB 线、手机与基站之间，以及卫星与地面接收器之间传输的数据。

**4. 日志审计**

审计是追踪集群中用户和服务行为的机制，是安全问题中的一个关键部分。如果没有

审计，那么任何人都可能察觉不到安全被破坏。审计功能对发生的事情均会详细记录，以完善安全模型，具体经常分为 3 类。

1）主动审计

主动审计经常与其他检测告警机制一起使用。例如，若某用户试图访问集群中的资源并被拒绝，主动审计能够生成一封告警邮件发送给安全管理员，通知其该事件的发生。

2）被动审计

被动审计是指那些不会产生告警信息的审计，其通常是业务的最低要求，因此，指定的审计员和安全管理员可以通过查询审计寻找某些事件。例如，集群中存在一个安全漏洞，安全管理员能够通过查询审计日志找到漏洞被利用期间访问的那些数据。

3）安全合规

如果数据涉及敏感信息，如个人身份信息、信用卡号、银行账户等财务信息，则需要在审计后，才能满足内部合规性或法律合规性。

## 5.5 大数据保护的原则与对策

### 5.5.1 大数据保护的基本原则

保护大数据，应该既要实现对数据的保护，又不能影响数据的自由流通和合理应用。即一方面要积极制定数据保护规则，保障与数据相关的权利；另一方面要积极构建数据平台，促进数据的自由流通和合理有效应用。大数据保护涉及的基本原则主要包括数据主权原则、数据保护原则、数据自由流通原则和数据安全原则，下面分别对其进行介绍。

**1. 数据主权原则**

数据主权原则是指一个国家独立自主地对本国数据进行占有、管理、控制、利用和保护的权力。该原则是大数据保护的首要原则。在大数据时代，数据是关系到个人安全、社会安全和国家安全的重要战略资源，无论是在经济发展和国家建设方面，还是在社会稳定方面，世界各国对数据资源的依赖越来越大，国家之间竞争和博弈的主战场也从传统领域逐渐转向大数据领域，因此必须确保数据的主权。

数据主权原则对内体现为一个国家对其政权管辖地域内任何数据的生成、传播、处理、分析、利用和交易等拥有最高权力；对外表现为一个国家有权决定以何种程序、何种方式参加到国际数据活动中，并有权采取必要措施保护数据权益免受其他国家的侵害。

**2. 数据保护原则**

数据保护原则的主旨是确认数据为独立的法律关系客体，奠定构建数据规则的制度基础。在数据保护原则之下，数据的法律性质和法律地位得以明确，从而使数据作为一种独立利益而受到法律的确认和保护。

具体来说，数据保护原则包含两个方面的含义：第一，数据不是人类的“共同财产”，数据的权属关系应该受到法律的保护，法律须确认权利人对数据的权利；第二，数据本身及其流通过程应该受到法律的保护，规范合理的数据流通不但能够确保数据的合理使用，还能够促进数据的再生和再利用。

**3. 数据自由流通原则**

所谓数据自由流通原则，是指法律应该确保数据作为独立的客体能够在市场上自由流

通，而不对其进行不必要的限制。这一原则的含义主要体现在两个方面：一是促进数据的自由流通，数据作为一种独立的生产要素，只有充分流通起来，才能够促进社会生产力的发展；二是反对数据垄断，对于那些利用数据技术优势阻碍数据自由流通的行为，应该予以坚决抵制。

为了确保数据的共享，要积极贯彻落实数据自由流通原则，只有这样才能在全球范围内消除数字鸿沟，建立国际数据共享的新秩序。由于各个国家、地区在信息技术发展方面存在严重的不平衡，这就使得数据的获取和使用存在严重的地区差异，进而对数据在全球范围的自由共享产生影响。因此，为实现数据共享，要坚持数据自由流通原则，加强政府对数据共享的宏观控制能力，在数据共享的发展战略上保持适度超前的政策管理，建立促进数据共享的政策法规制度，加强信息技术的共享。

**4. 数据安全原则**

数据安全原则是指通过法律机制来保障数据的安全，以免数据面临遗失、不法接触、毁坏、利用、变更或泄露的危险。从安全形态上看，数据安全包括数据存储安全和数据传输安全；从内容上看，数据安全可分为信息网络的硬件、软件的安全，数据系统的安全和数据系统中数据的安全。

具体来说，数据安全原则包括以下几方面的含义：第一，保障数据的真实性和完整性，既要加强对静态存储数据的安全保护，使其不被非授权访问、篡改和伪造，也要加强对数据传输过程的安全保护，使其不被中途篡改、不发生丢失和缺损等；第二，保障数据的安全使用，数据及其使用必须确保其保密性，禁止任何机构和个人的非授权访问；第三，以合理的安全措施保障数据系统具有可用性，可以为获得合法授权的使用者提供服务。

### 5.5.2 大数据保护的对策

大数据时代，应从以下几个方面加强数据安全与隐私保护。

第一，从国家法制层面进行管控。目前国家出台了诸多涉及数据安全和隐私保护的法律法规，来实现对数据安全和隐私的保护。当然，从国家法律层面来讲，为顺应大数据时代发展趋势，还需要进一步细化和完善对个人信息安全的立法，出台相应的细化标准与措施。

第二，从企业端源头进行遏制。企业是收集、存储、使用、传播个人数据的主体，因此要从企业端进行遏制、规范。除了要遵循国家法律法规的约束之外，企业应积极采取措施加强和完善对个人数据的保护，不能过度收集个人数据，避免因个人数据的不当使用和泄露而对多方造成损失。

第三，增强个人数据安全意识，学习应用数据安全技术。生活在大数据环境中的每一个人，都应该主动学习数据安全相关技术知识，了解大数据时代可能会存在的关于个人隐私泄露等风险，学会如何去保护自己的隐私数据。同时还应加强安全意识，例如，保护密码等敏感信息，不在社交平台上发布个人定位信息，不要连接公共 Wi-Fi 进行支付等重要操作等。

## 5.6 大数据安全相关法律法规

### 5.6.1 我国大数据安全相关法律法规

我国作为数据量最大、数据类型最丰富的国家之一，大数据安全面临的形势日益严峻。

为应对层出不穷的大数据安全问题，国家层面从顶层制度设计出发，不断完善相关的法律法规和规范性文件等。

**1.《中华人民共和国网络安全法》**

《中华人民共和国网络安全法》于 2016 年 11 月 7 日由中华人民共和国主席签署主席令予以公布，自 2017 年 6 月 1 日起施行，共包含 7 章 79 条。

1）制定目的

《中华人民共和国网络安全法》的第一条指出，制定本法的目的是保障网络安全，维护网络空间主权和国家安全、社会公共利益，保护公民、法人和其他组织的合法权益，促进经济社会信息化健康发展。

2）主要内容

《中华人民共和国网络安全法》内容主要涵盖关键信息基础设施保护、网络数据和用户信息保护、网络安全应急与监测等领域，与网络空间国内形势、行业发展和社会民生关系紧密的主要包括以下内容。

（1）确立了网络空间主权原则，将网络安全顶层设计法治化。

（2）对关键信息基础设施实行重点保护，将关键信息基础设施安全保护制度确立为国家网络空间基本制度。

（3）加强个人信息保护要求，加大对网络诈骗等不法行为的打击力度。

（4）重要数据强制本地存储制度和境外数据传输审查评估制度。

**2.《中华人民共和国数据安全法》**

《中华人民共和国数据安全法》于 2021 年 6 月 10 日由中华人民共和国主席签署主席令予以公布，自 2021 年 9 月 1 日起施行，共包含 7 章 55 条。

1）制定目的

《中华人民共和国数据安全法》第一条指出，制定本法的目的是规范数据处理活动，保障数据安全，促进数据开发利用，保护个人、组织的合法权益，维护国家主权、安全和发展利益。

2）主要内容

（1）确立了数据安全有关的几个重要制度。

（2）规定了有关部门、单位和个人在数据安全保护方面的义务和责任。

（3）规定了在数据安全保护方面违法的法律责任。

**3.《中华人民共和国个人信息保护法》**

《中华人民共和国个人信息保护法》于 2021 年 8 月 20 日由中华人民共和国主席签署主席令予以公布，自 2021 年 11 月 1 日起施行，共包含 8 章 74 条。

1）制定目的

《中华人民共和国个人信息保护法》的第一条指出，制定本法的目的是保护个人信息权益，规范个人信息处理活动，促进个人信息合理利用。

2）主要内容

（1）明确了个人信息的范畴。具体体现在第四条。

（2）规定了个人信息处理规则。

（3）规定了个人信息跨境提供的规则。

（4）规定了个人在个人信息处理活动中的权利。

(5) 规定了个人信息处理者的义务。

(6) 规定了履行个人信息保护职责的部门及其职责。

(7) 规定了在个人信息保护方面违法的法律责任。

《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》合称为我国数据安全法律的“三驾马车”,这 3 部法律构成了我国数据安全领域完整的基础性法律体系,为国家、地方、行业、个人的数据安全保护提供了最重要的法律参考依据。其他由各部委、组织、地方等出台的条例、办法等均是依据这 3 部重要法律对数据安全管理体系进行的进一步丰富和细化。

**4.《网络安全审查办法》**

《网络安全审查办法》于 2021 年 11 月 16 日经国家互联网信息办公室 2021 年第 20 次室务会议审议通过,自 2022 年 2 月 15 日起施行,共包含 23 条。

1) 制定目的

《网络安全审查办法》的第一条指出,制定本办法的目的是确保关键信息基础设施供应链安全,保障网络安全和数据安全,维护国家安全。制定依据为《中华人民共和国国家安全法》《中华人民共和国网络安全法》《中华人民共和国数据安全法》《关键信息基础设施安全保护条例》。

2) 主要内容

(1) 明确了进行网络安全审查的条件与对象。《网络安全审查办法》进一步明确了网络平台运营者需向网络安全审查办公室申报网络安全审查的情形。

(2) 明确了网络安全审查重点评估的因素。

**5.《网络数据安全管理条例》**

《网络数据安全管理条例》于 2024 年 8 月 30 日经国务院第 40 次常务会议审议通过,2024 年 9 月 24 日,国务院总理签署第 790 号国务院令予以公布,自 2025 年 1 月 1 日起施行,共包含 9 章 64 条。

1) 制定目的

《网络数据安全管理条例》的第一条指出,制定本条例的目的是规范网络数据处理活动,保障网络数据安全,促进网络数据依法合理有效利用,保护个人、组织的合法权益,维护国家安全和公共利益。制定依据为《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等。

2) 主要内容

(1) 明确了个人信息保护原则。

(2) 明确了针对重要数据安全保护各部门的职责分工及保护制度,具体体现在第二十九条至第三十三条。

(3) 规定了网络数据跨境安全管理原则。

(4) 规定了网络平台服务提供者应履行的义务。

(5) 规定了有关主管部门的监督管理职责。

(6) 规定了在网络数据安全管理方面违法的法律责任。

以上法律法规的具体章节条目均可通过访问“中国政府网”进行查询。

### 5.6.2 国外大数据安全相关法律法规

世界各国在具体的大数据安全保护实践中分别出台了适合自己国家国情的法律法规。在此简单对欧盟和美国的情况进行介绍。

**1. 欧盟《通用数据保护条例》**

《通用数据保护条例》于 2016 年 4 月 14 日由欧洲联盟议会颁布，自 2018 年 5 月 25 日起在欧盟全体成员国正式生效，共包含 10 章 99 条，被广泛认为是欧盟有史以来最为严格的网络数据管理法规。其前身是欧盟在 1995 年制定的《计算机数据保护法》。

1）制定目的

该条例旨在确立个人数据处理中的自然人保护和个人数据自由流通的规范，保护自然人的基本权利和自由，尤其是个人数据保护的权利。

2）主要内容

（1）规定了个人信息处理必须遵守的原则。具体包括以下内容：

- 合法性、公平性和透明性。个人数据的处理必须以合法、公平和透明的方式进行。数据控制者需要明确说明数据处理的目的，并以易于理解的语言向数据主体提供相关信息。
- 目的限制原则。个人数据只能为特定、明确和合法的目的而收集，不得进行与原目的不相符的进一步处理。
- 数据最小化原则。个人数据的收集和处理应限于实现目的所需的最小范围。数据控制者应确保只收集和保留必要的数据。
- 准确性原则。个人数据必须准确，并在必要时保持最新。数据控制者应采取合理措施，确保不准确的数据被及时删除或更正。
- 存储限制原则。个人数据的保存时间不应超过实现处理目的所需的时间。一旦不再需要，数据应被删除或匿名化。
- 完整性和保密性原则。个人数据应以确保其安全的方式进行处理，包括防止未经授权或非法处理，以及防止意外丢失、破坏或损坏。
- 问责原则。数据控制者应对其数据处理活动负责，并能够证明其遵守 GDPR 的原则。

（2）明确了企业的合规义务和责任。具体包括以下内容：

- 进行数据保护影响评估。对于可能对个人权利和自由构成高风险的数据处理活动，企业必须在处理前进行数据保护影响评估，评估潜在的风险并制定缓解措施。
- 任命数据保护官。在某些情况下，如大规模处理敏感数据或定期系统性监控个人，企业必须任命数据保护官。数据保护官负责监督 GDPR 的合规性，并作为企业与监管机构之间的联络人。
- 维护数据处理活动记录。企业需要记录其数据处理活动，包括处理的目的、数据主体类别、数据接收者、数据保留期限等。这些记录有助于企业证明其合规性，并在需要时提供给监管机构。
- 及时报告数据泄露。在发生可能危及个人权利和自由的数据泄露时，企业必须在 72 小时内向监管机构报告，并在某些情况下通知受影响的个人。

- 管理用户同意和权利。企业需要以清晰、易懂的方式获得用户对数据处理的明确同意，并提供撤回同意的机制。此外，企业还必须响应用户行使其 GDPR 权利的请求，如访问、纠正、删除个人数据等。

（3）完善了相关监管机制，加强了处罚力度。例如，违反某些条款后，将施加最高 20 000 000 欧元的行政罚款，如果是企业的话，最高可处相当于其上一年全球总营业额 4% 的金额的罚款，两者取其高的一项进行罚款。

欧盟各成员国监管机构的检查权、执法权、处罚权以及司法诉讼权等内容，以及欧盟层面的监管机构等，都在条例中得以明确界定。

**2. 美国数据安全保护相关法律法规**

美国由于其联邦制政体的特殊性，到目前为止并没有一个统一的联邦隐私法，但各州根据自身的需要和情况，制定了不同的隐私保护法规。根据适用的协议条款，一般来说这些法律法规会被纳入“适用的美国州数据保护法”或“美国州隐私法”范围。其中比较著名的州立相关法律有《加利福尼亚消费者隐私法案》（California Consumer Privacy Act，CCPA）。该法案于 2018 年 6 月 28 日颁布，2020 年 1 月 1 日生效，是加利福尼亚州首部专门针对消费者隐私保护的法律，也是美国第一部内容全面的隐私法。该法案规定了消费者个人数据处理的基本要求、责任和义务，以及消费者隐私保护、数据泄露通知等方面的内容。

在联邦政府层面，2022 年 6 月 3 日，美国参议院和众议院发布了《美国数据隐私和保护法》（American Data Privacy and Protection Act，ADPPA）草案。该法案是第一个获得两党、两院支持的美国联邦全面隐私保护提案。ADPPA 对敏感数据的定义比在州级立法中的定义广泛得多，且 ADPPA 明确优先于所有州隐私法。

2024 年 4 月 5 日，美国两党、两院发布了《美国隐私权法案》（The American Privacy Rights Act，APRA）草案，这是自 2022 年 ADPPA 之后，美国在联邦层面数据隐私立法的再次尝试。不同于遥遥无期的 ADPPA 法案，APRA 法案由美国参议院商业、科学和交通委员会和美国众议院能源和商业委员会联手推出，旨在为美国人设立明确的全国数据隐私权利和保护措施，消除了现有的各州综合数据隐私法的混杂局面，并建立了强有力的执法机制来追究违规者的责任。

# 小结

本章探讨了大数据安全主题，结合 Facebook 数据门事件、手机 App 过度采集个人信息等大数据安全典型案例，从大数据安全框架和大数据安全技术、大数据保护的原则与对策、国内外大数据安全相关的法律法规等角度进行了介绍，有助于读者在应用大数据技术时树立大数据安全意识。

# 习题

1. 数据安全的本质是什么？
2. 大数据安全与传统数据安全的区别主要体现在哪几个方面？
3. 大数据在技术层面面临的安全威胁有哪些？

4. 数据生存周期安全过程具体包括哪几个阶段？

5. 在 DSMM 的能力成熟度等级维度中，组织的数据安全能力成熟度被划分为哪几个等级？

6. 大数据保护主要涉及哪些基本原则？

7. 请列举我国在大数据安全方面制定的主要法律法规。

8. 大数据时代，可以从哪几个方面加强数据安全与隐私保护？