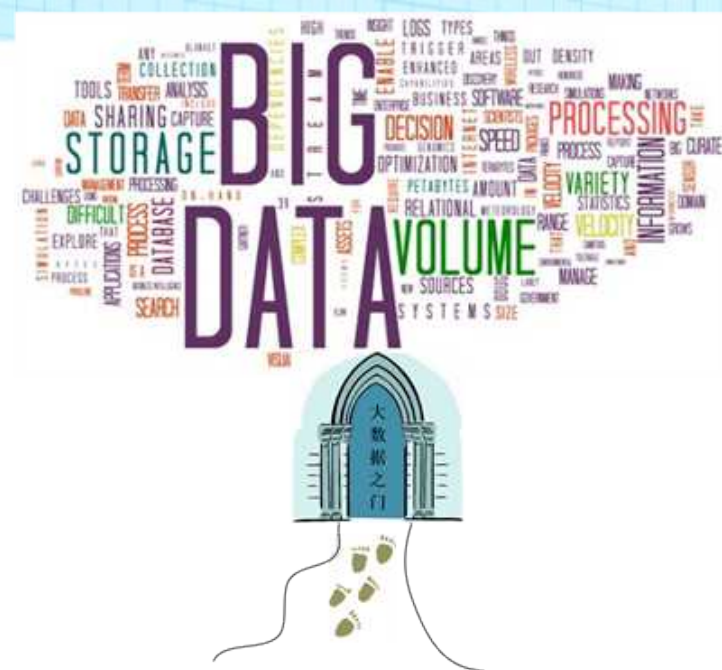




大数据平安及隐私

参考书及补充

重点与难点

- 内容：大数据平安、隐私及其保护的根本概念。
- 重点：平安及隐私保护技术。
- 难点：密码学和协议。

内容提纲

5.1 大数据平安

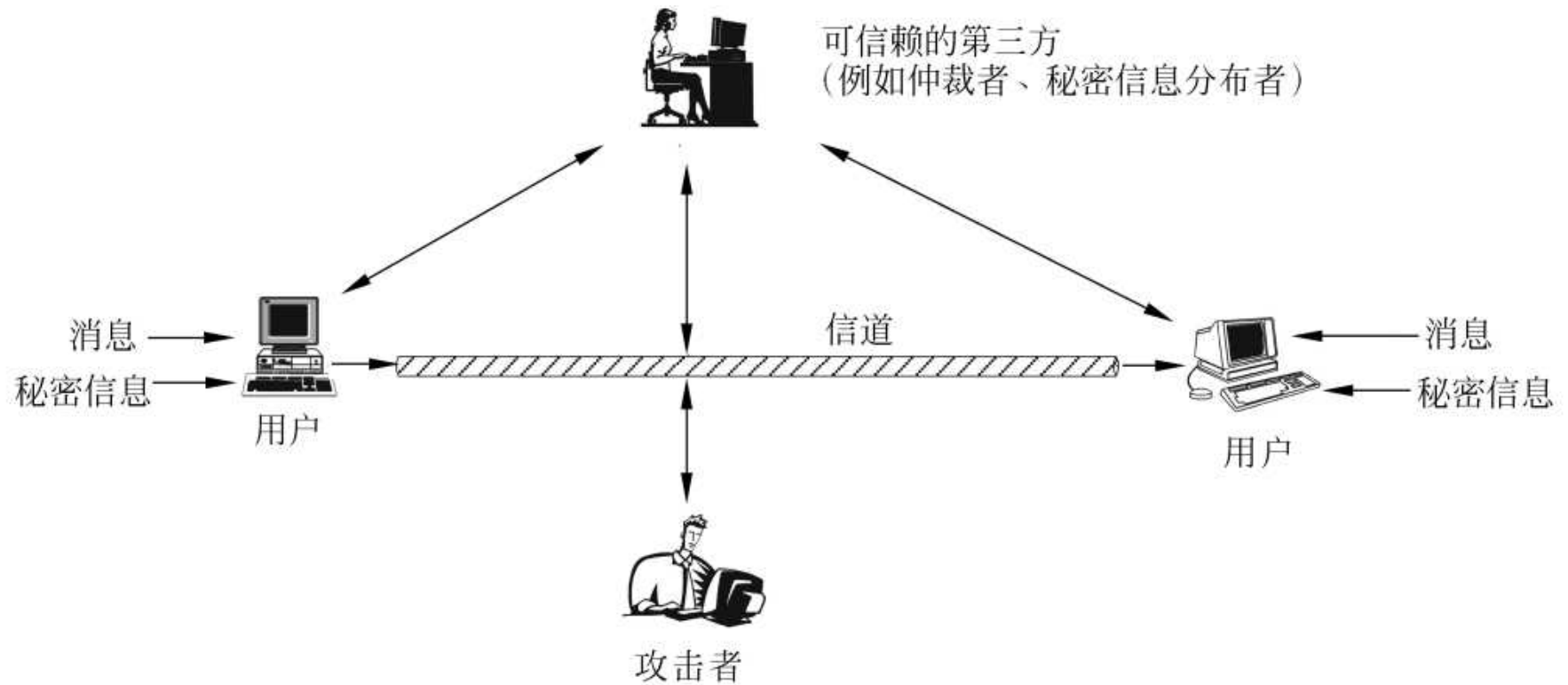
5.2 隐私及其保护

5.3 信息平安技术

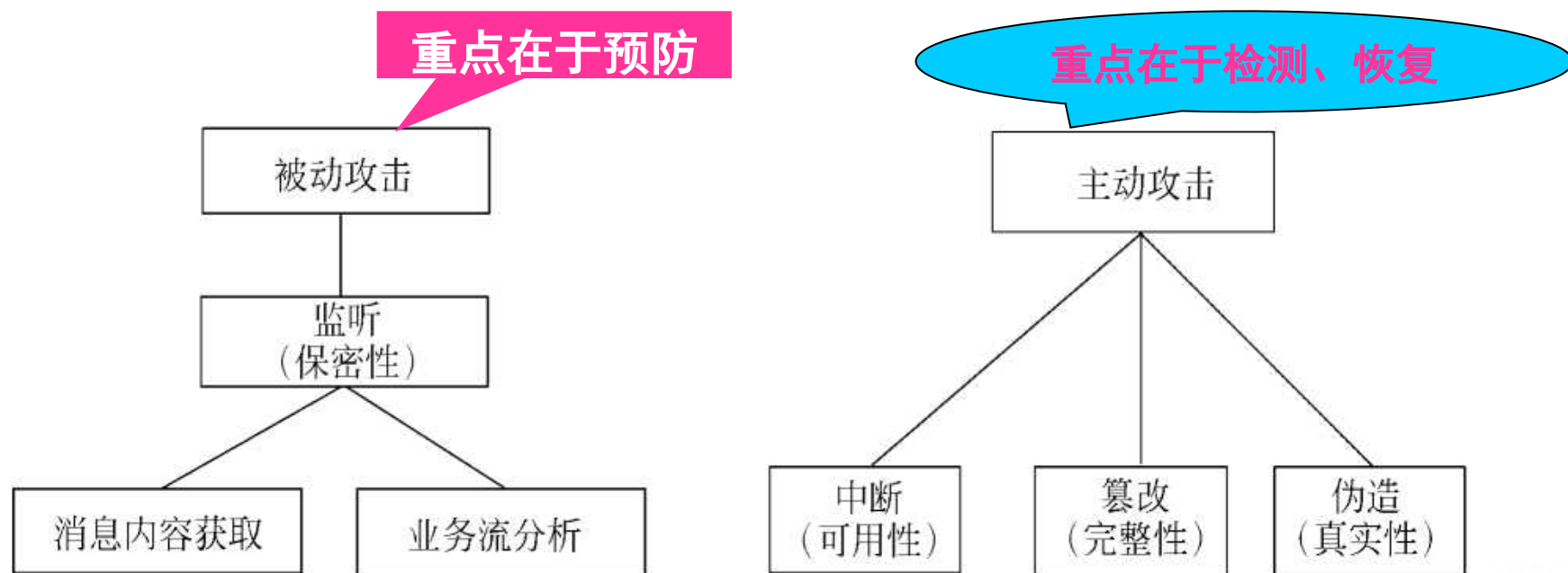
5.1 大数据平安

- ◆ 信息平安〔Information Security〕：是一个综合、交叉的学科领域，研究内容涉及平安体系结构、平安协议、密码理论、信息分析、平安监控、应急处理等。分为3个层次：
- ◆ 系统平安：操作系统、数据库系统。
- ◆ 数据平安：数据的平安存储与传输。
- ◆ 内容平安：病毒防护、不良内容过滤
- ◆ ***密码学技术是保障信息平安的关键技术

◆ 信息平安的模型



- ◆ 信息平安威胁：宏观地分为人为威胁和自然威胁
- ◆ 人为威胁：对信息的人为攻击，主要来自用户（恶意的或无恶意的）和恶意软件的非法侵入。
- ◆ 自然威胁：可能来自于自然灾害、恶劣的场地环境、电磁辐射和电磁干扰、网络设备自然老化等。



人为攻击

- ◆ 信息的平安性〔平安属性〕
- ◆ 机密性：是指信息不泄露给非授权实体并供其利用的特性。
- ◆ 完整性：是指信息不能被未经授权的实体改变的特性。
- ◆ 不可否认性：是指通信双方信息真实同一的平安属性，收发双方均不可否认。
- ◆ 可控性：是指授权实体可以对信息及信息系统实施平安监控，控制信息系统和信息使用的特性。
- ◆ 可用性：指信息能够被授权实体访问并按要求使用，信息系统能以人们所接受的质量水平持续运行，为人们提供有效的信息效劳的特性。

- ◆ 平安效劳
- ◆ 鉴别效劳：提供对通信中对等实体和数据来源的鉴别。
- ◆ 访问控制效劳：对资源提供保护，以对抗其非授权使用和操作。
- ◆ 机密性效劳：保护信息不被泄露或暴露给非授权的实体。
- ◆ 完整性效劳：对数据提供保护，以对抗未授权的改变、删除或替代
- ◆ 抗抵赖效劳：防止参与某次通信交换的任何一方事后否认本次通信或通信内容。

◆ 平安机制

- ◆ 加密机制：为数据提供机密性，也为通信业务流信息提供机密性。
- ◆ 数字签名机制：证实消息的真实来源和抗抵赖。
- ◆ 访问控制机制：对资源访问或操作加以限制的策略。
- ◆ 完整性机制：保护数据以防止未经授权的数据乱序、丧失、重放、插入和篡改。
- ◆ 鉴别交换机制：通过密码、密码技术、实体特征或占有物提供对等实体的鉴别。
- ◆ 通信业务填充机制：提供通信业务流机密性，对抗通信业务分析。
- ◆ 路由选择控制机制：使路由能动态地或预定地选取，以便通信只在具有适当保护级别的路由上传输。
- ◆ 公正机制：公正人为通信双方所信任，并掌握必要信息以一种可证实方式提供所需的保证。

- ◆ 数据平安：是一种主动的包含措施，包括数据本身的平安和数据防护的平安。
- ◆ 数据本身的平安，主要是指采用现代密码算法对数据进行主动保护，如数据保密、数据完整性、双向强身份认证等；
- ◆ 数据防护的平安，主要是采用现代信息存储手段对数据进行主动防护，如通过磁盘阵列、数据备份、异地容灾等手段保证数据的平安。

- ◆ 大数据平安内涵：应该包括两个层面的含义：
- ◆ 保障大数据平安，是指保障大数据计算过程、数据形态、应用价值的处理技术；
- ◆ 大数据用于平安，利用大数据技术提升信息系统平安效能和能力的方法，涉及如何解决信息系统平安问题。

行业	安全需求
互联网	有效的安全存储和智能挖掘分析，严格执行大数据安全监管和审批管理，从海量数据中合理发现和发掘商业机会与商业价值
电信	核心数据和资源的保密性、完整性和可用性、保障用户利益、体验和隐私的基础上发挥数据的价值
金融	对数据访问控制、处理算法、网络安全、数据管理和应用等方面提出安全要求，期望利用大数据安全技术加强金融机构的内部控制，提供金融监管和服务水平，防范和化解金融风险
医疗	数据隐私性高于安全性和机密性，同时要安全可靠的数据存储、完善的数据备份和管理，以帮助医生与病人进行疾病诊断，药物开发，管理决策、完善医院服务，提高病人满意度，降低病人流失率
政府	隐私保护的安全监管、网路环境的安全感知、大数据安全标准的制定、安全管理机制规范等内容

- ◆ 大数据平安威胁
- ◆ 大数据根底设施具有虚拟化和分布式特点，容易受到非授权访问、信息泄露或丧失、破坏数据完整性、拒绝效劳攻击、网络病毒传播；
- ◆ 大数据应用的生命周期中，数据存储是一个关键环节，尤其是非关系数据库带来如下平安挑战：模式成熟度不够、系统成熟度不够、效劳器软件没有内置的足够的平安机制、数据冗余和分散性问题。
- ◆ 网络平安是大数据平安防护的重要内容，平安问题随着网络节点数量的增加呈指数级上升；平安数据规模巨大，平安事件难以发现；平安的整体状况无法描述；平安态势难以感知，等等。

- ④ 网络化社会使大数据易成为攻击目标：在网络化社会中，信息的价值要超过根底设施的价值；大数据蕴涵着人与人之间的关系和联系，使得黑客成功攻击一次就能获得更多数据。
- ⑤ 大数据滥用的风险：一方面，大数据本身的平安防护存在漏洞，平安控制力度不够。另一方面攻击者利用大数据技术进行攻击，最大限度地收集更多用户敏感信息。
- ⑥ 大数据误用风险：大数据的准确性和数据质量不高，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。

- ◆ 传统数据平安的缺乏
- ◆ 传统数据平安往往是围绕数据生命周期来部署，即数据的产生、存储、使用和销毁。由于大数据的规模没有上限，且许多数据的生命周期极为短暂，因此，常规平安产品想要继续发挥作用，那么需要解决如何根据数据存储和处理的动态化、并行化特征，动态跟踪数据边界，管理对数据的操作行为等。
- ◆ 随着大数据应用越来越多，数据的拥有者和管理者相别离，原来的数据生命周期逐渐转变成数据的产生、传输、存储和使用。数据以外包形式存储在云端。

- ◆ 保障大数据平安
- ◆ 围绕大数据全生命周期，即数据的产生、采集、传输、存储、处理、分析、发布、展示和应用、产生新数据等阶段进行平安防护。
- ◆ 大数据保障技术可以从物理平安、系统平安、网络平安、存储平安、访问平安、审计平安、运营平安等角度进行考虑。
- ◆ 保障大数据平安的目标：最大程度地保护具有流动性和开放性特征的大数据自身平安，防止数据泄露、越权访问、数据篡改、数据丧失、密钥泄露、侵犯用户隐私等问题的发生。
- ◆ 对大数据全生命周期阶段合并与精简，大数据应用过程可划分为4个环节：采集、存储、挖掘和发布。

1、采集环节平安技术

平安威胁：数据损坏、丧失、泄露、窃取。

平安机制：

身份认证

数据加密

完整性保护

抗重放攻击

端到端的数据平安传输：采用SSL VPN〔平安套接层，虚拟专用网〕，有Web浏览器模式、客户端模式、Lan到Lan模式。

2、存储环节平安技术

平安威胁：数据被窃取、丧失、破坏、机密数据泄露、隐私数据泄露。

平安机制：

隐私保护：数据变换、数据加密、匿名化。

数据加密：静态数据是指文档、报表、资料等不参与计算的数据—先加密再存储；动态数据是指需要检索或参与计算的数据—同态加密。

备份与恢复：异地备份、RAID〔独立磁盘冗余阵〕、数据镜像、快照〔数据副本、复制品〕、

3、挖掘环节平安技术

数据挖掘是大数据应用的核心局部，是开掘大数据价值的过程，即从海量的数据中自动抽取隐藏在数据中 useful 信息的过程，有用信息可能包括规那么、概念、规律和模式等。

平安威胁：第三方在进行数据挖掘的过程中植入恶意程序、窃取系统数据。

平安机制：

身份认证：基于秘密信息、基于信物、基于生物特征

访问控制：自主访问控制DAC，可将拥有的权限自主地授予其他主体，并在随后的任何时刻将这些权限回收，控制是自主的；强制访问控制MAC—系统根据使用系统的机构事先确定的平安策略，对用户的访问权限进行强制性的控制，用户不能改变他们的平安级别或对象的平安属性；基于角色的访问控制RBAC—在用户和访问权限之间引入角色的概念，将用户和角色联系起来，通过对角色的授权来控制用户对系统资源的访问。

4、发布环节平安技术

发布是指大数据在经过挖掘分析后，向数据应用实体输出挖掘结果数据的环节，即数据“出门”环节，必须对即将输出的数据进行全面的审查，确保输出的数据符合“不泄密、无隐私、不超限、合规约”等要求。

平安威胁：泄密、泄露隐私、不合标准。

平安机制：

平安审计：基于日志的、基于网络监听的、基于网关的、基于代理的

数据溯源：就是对大数据应用周期的各个环节的操作进行标记和定位，在发生数据平安问题时，可以及时准确地定位到出现问题的环节和责任者，以便于对数据平安问题的解决。

溯源方法：标注法、反向查询法、数字水印。

- ◆ 大数据用于平安
- ◆ 平安监测与大数据的融合技术，利用大数据技术实时监控海量、多样、快速和复杂的数据，有效提高平安监测的效果和能力。
- ◆ 大数据技术将极大扩展平安分析的深度和广度，把传统的数据平安分析的被动的事后分析变成主动地事前防御，能够更好地感知网络平安态势。
- ◆ 在信息平安领域借助大数据处理技术，可以针对APT (Advanced Persistent Threat, 高级持续性威胁，恶意商业间谍威胁) 攻击隐蔽能力强、长期潜伏、攻击路径和渠道不确定等特征，设计具备实时检测能力与事后回溯能力的全流量审计方案。

5.2 隐私及其保护

- ◆ 隐私是与个人相关的、具有不被他人搜集、保存和处分的权利的信息资料集合，并且它能够按照所有者的意愿在特定时间、以特定方式、在特定程度上被公开。
- ◆ 隐私的根本属性：保密性、个人相关、能够被所有者处分。
- ◆ 隐私分类
- ◆ 个人隐私：任何可以确定特定个人或与可确定的个人相关，但个人不愿意暴露的信息，譬如：就诊记录。
- ◆ 共同隐私：不仅包含个人隐私，还包含所有个人共同表现出但不愿意被暴露的信息，譬如：平均薪资。

◆ 隐私权

- ◆ 1890年，哈佛大学法学院教授Samuel D. Warren和Louis D. Brandeis发表题为“隐私权”的论文，标志隐私权理论的诞生。
- ◆ “公民权利和政治权利国际公约”第17条：任何人的私生活、家庭、住宅和通信不得任意或非法干预，其荣誉和名誉不得加以攻击。人人有权享受法律保护，以免受非法干预或攻击。我国将侵犯隐私权的行为视为侵害名誉权。

- 信息时代的隐私权保护要比传统的隐私权保护重要得多。信息隐私权保护的客体可分为以下4个方面：
- 个人属性的隐私权：譬如姓名、身份、肖像、声音等，直接涉及个人领域的第一层次，是保护的首要对象。
- 个人资料的隐私权：譬如消费习惯、病历、犯罪前科等记录，假设其涉及的客体为一个人，那么这种资料含有高度的个人特性而常能辨识该个人的本体，即“间接”的个人属性，也应予以隐私权加以保护。
- 通信内容的隐私权：通信内容包含个人的思想与情感，原本存于内心，别人不可能知道，因而应以隐私权保护，以保护个人人格的完整开展。
- 匿名的隐私权：匿名权利的适度许可，可以鼓励个人的参与感，并保护其自由创造力空间，真知直谏推动社会的整体进步。

◆ 隐私保护

◆ 隐私保护〔Privacy Preserving〕：是对个人隐私采取一系列的平安手段防止其泄露和被滥用的行为。

◆ 隐私保护面临的威胁

◆ 数据搜集：Internet效劳提供商在搜集、下载、集中、整理和利用用户个人隐私资料极为方便。

◆ 信息效劳：个性化需求的信息效劳需要用户提供更多的个人信息，才能提供更好的用户体验。

◆ 搜索引擎：无法对自己搜索到的网页数据库信息进行监督，不会对搜索到的内容信息负责。

◆ 数据挖掘：从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。

- 隐私保护面临大数据的威胁
- 大数据时代带来信息存储和管理的集中化，一个大规模生产、分享和应用数据的时代，一切皆可量化，通过社交网络将社会关系和活动数据化，实现了过去不可想象的情绪数据化。
- 大数据通常包含了大量的用户身份信息、属性信息、行为信息，在大数据应用的各阶段内，如果不能保护好大数据，极易造成用户隐私泄露。
- 大数据的多源性，使得来自各个渠道的数据可以用来进行交叉检验，有可能发现匿名化数据后面的真实用户，因而导致隐私泄露。大数据时代，人们无法防止失去隐私。

- ④ 数据存储过程：大数据环境下，用户无法知道数据确切的存放位置，用户对其个人数据的采集、存储、使用、分享无法有效控制。
- ⑤ 数据传输过程：大数据环境下，数据传输将更为开放和多元化，传统物理区域隔离的方法无法有效保证远距离传输的平安性，电磁泄漏和窃听将成为更加突出的平安威胁。
- ⑥ 数据处理过程：大数据环境下，部署大量的虚拟技术，资源动态共享增加了访问控制和身份认证的管理难度，根底设施的脆弱性和加密措施的失效可能产生新的平安风险。

- ◆ 隐私保护技术
- ◆ 隐私保护技术是用于保护用户隐私的各种平安策略的功能集合。
- ◆ 访问控制：隐私数据的内容和用途相关。
- ◆ 推理控制：防止依据低密级的数据和模式的完整性约束推出高密级的数据。
- ◆ 数据变换技术：将用户的真实隐私数据进行伪装或经轻微改变，而不影响原始数据的使用。
- ◆ 密码和密码协议：平安多方计算、盲签名
- ◆ 匿名化技术：通过对需要保密的数据进行泛化和隐匿处理，防止攻击者通过准标识符将某一个体与其敏感属性值关联起来。
- ◆ 保护隐私数据挖掘技术：基于启发式的〔集中式数据对象，方法：值替代和分组〕，基于密码学的〔分布式数据对象，水平或垂直分割〕，基于重构的〔集中式对象，使用方法构造原始数据的分布〕。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/767026155066006110>