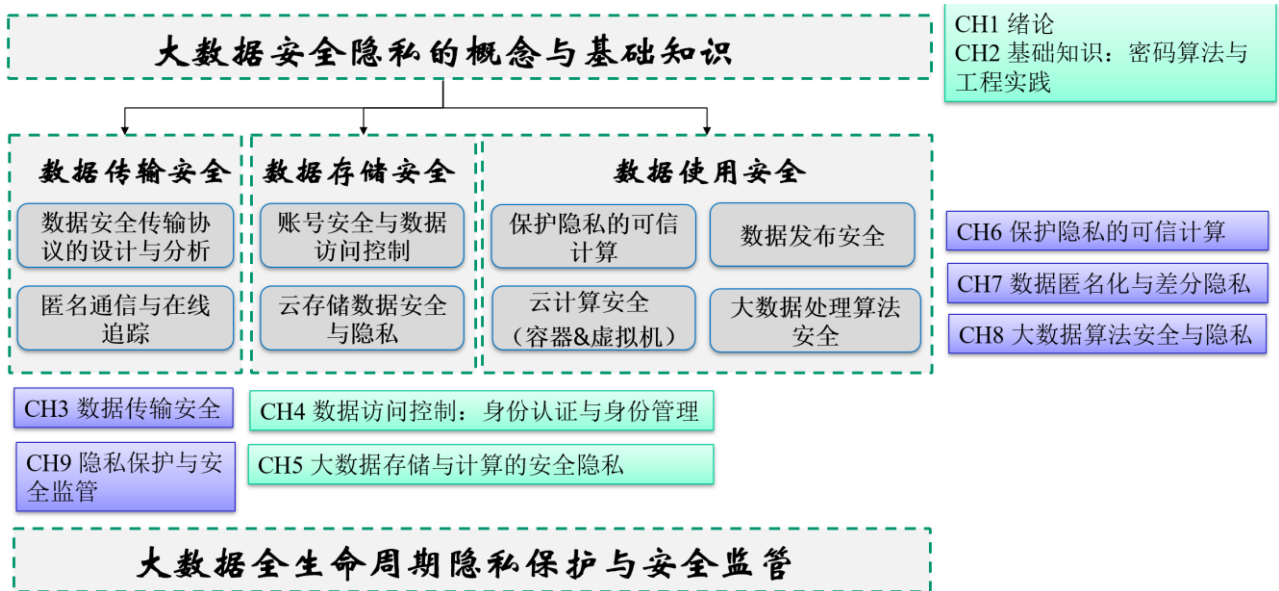


大数据安全（北京邮电大学 石瑞生）

学习的过程：从薄到厚，再从厚到薄



CH1 绪论

- 1、什么是大数据？
- 2、什么是大数据安全？
- 3、隐私权概念的发展经历了哪些主要变化？
- 4、技术进步与隐私保护的关系
- 5、为什么说斯诺登事件是全球网络安全的转折点？有哪几类可能侵犯大众隐私的敌手？
- 6、GDPR 的要点
- 7、在数据安全与隐私保护领域，我国相关的法律法规有哪些？
- 8、我国的数据资源跨境流通政策
- 9、隐私和安全之间的关系
- 10、个人数据安全与个人隐私之间的关系，举例说明。
- 11、个人隐私与国家安全之间的关系？为什么要保护个人隐私？

CH2 基础知识：密码算法与工程实践

- 1、加密强度与密钥长度的关系：给定安全等级和加密算法，如何选择密钥长度？【安全强度的选择通常是在 128 比特安全性和 256 比特安全性之间进行的，因为大多数标准加密

算法和实现对于实现这两个安全强度都是可行的。】

- 2、混合加密技术：KEM，DEM
- 3、几个核心概念：**攻击模型 (Attack Model)**、**安全目标 (Security Goal)**、**安全概念 (security notion，记为 GOAL-MODEL)**
- 4、3 类攻击模型：黑盒模型，灰盒模型，白盒模型。
- 5、5 种**黑盒攻击模型** (COA, KPA, CPA, CCA, CCA2)；
- 6、2 类安全目标 (不可区分性，不可塑性)；
- 7、解释语义安全的概念。实现语义安全的基本思路【随机加密或者概率加密】。
- 8、4 种常见的安全概念 (NM-CPA, NM-CCA, IND-CPA 和 IND-CCA)
- 9、第一个语义安全的密码算法是什么算法？
- 10、ElGamal 密码算法是语义安全的吗？ECC 加密算法是语义安全的吗？
- 11、RSA 算法是语义安全的吗？如何使用编码函数提高 RSA 算法的安全性？
- 12、AES 算法是语义安全的吗？如何构造基于 AES 算法的语义安全加密方案？
- 13、理解认证加密方案对安全性的提升的意义
- 14、典型案例分析：在工程实践中使用密码算法设计安全技术方案时需要注意的问题

CH3 数据传输安全

- 1、TLS 协议的基本原理：协议架构，报文结构，协议流程
- 2、TLS 握手协议的安全性分析与改进方案：前向安全性，中间人攻击，重放攻击
- 3、TLS 协议在实际部署中的安全性问题分析：HSTS，证书安全，CDN
- 4、HSTS 的 preload list 方案是为了解决什么问题？
- 5、TLS1.3 如何对抗降级攻击？如何对抗流量分析攻击？
- 6、TLS1.3 对性能做了哪些提高和改进？
- 7、QUIC 协议的特点
- 8、浏览器如何鉴别所访问网站证书的真伪？攻击者常用的欺骗手段有哪些？
- 9、ACME 协议的工作原理
- 10、证书透明化
- 11、公钥钉扎
- 12、数据传输安全协议设计的重点与常用技巧

CH4 数据访问控制：身份认证与身份管理

- 1、从用户视角，对身份认证方法的分类
- 2、从技术视角，对身份认证方法的分类
- 3、了解四种基本的身份认证方法及其面临的安全威胁
- 4、如何评估口令的安全性？掌握量化分析方法。
- 5、口令如何安全存储？了解主流的口令加密算法。
- 6、解释 FIDO 协议的工作原理。
- 7、Kerberos 协议的工作原理。
- 8、OAuth 协议的工作原理。

- 9、OpenID Connect 协议的工作原理
- 10、Cookie 的工作原理与安全属性
- 11、身份认证 Cookie
- 12、Cookie 劫持攻击的工作原理

CH5 大数据存储与计算的安全隐私

- 1、虚拟化技术：虚拟机，容器的三个核心概念与核心技术
- 2、三类容器安全问题：容器逃逸，镜像安全，集群入侵。
- 3、造成容器逃逸的两类原因（软件漏洞与错误配置）与缓解措施（与主机系统隔离，构建更强的信任边界）。
- 4、跨虚拟机的攻击与云计算环境下的安全挑战
- 5、大数据存储的数据完整性机制：POR
- 6、大数据存储的数据隐私保护机制（加密数据去重技术）
- 7、大数据存储的数据安全防护机制：PoW

CH6 保护隐私的可信计算

- 1、同态加密的概念；为什么需要同态加密技术？
- 2、乘法同态加密算法
- 3、加法同态加密算法：Paillier 算法
- 4、全同态加密算法的基本概念
- 5、CryptDB 的洋葱加密模式
- 6、安全多方计算的概念
- 7、安全多方计算的模型
- 8、百万富翁问题
- 9、混淆电路的工作原理
- 10、OT 的工作原理
- 11、秘密分享的工作原理
- 12、应用中的问题：性能，安全性（恶意参与者）

CH7 数据匿名化与差分隐私

- 1、隐私的定义
- 2、隐私的度量
- 3、什么是泛化？什么是抑制？
- 4、常见的隐私保护模型：K 匿名；L 多样化；T 相近
- 5、K 匿名与记录链接攻击
- 6、L 多样化与属性链接攻击

- 7、数据匿名化方法的局限性
- 8、差分隐私的基本思想：中间件（查询结果） + 随机噪声 = 带噪中间件。
- 9、差分隐私的定义
- 10、**基本概念：距离，相邻数据集，敏感度。**
- 11、常见的**查询函数及其敏感度**
- 12、拉普拉斯机制
- 13、指数机制
- 14、本地化差分隐私方法
- 15、差分隐私在机器学习中的应用

CH8 大数据算法安全与隐私（不考）

- 1、科学的发展与人类知识体系的构建：科学的发展是一个知识积累的过程。
- 2、什么是知识？怎么创造知识？
- 3、几乎所有的科学领域都在用模型拟合数据。数学模型在人类认知过程中扮演着重要的角色。
- 4、什么是数学模型？什么是好的数学模型？
- 5、数学模型与经典的大数据算法
- 6、四个要素：1) 假设：方向；2) 模型：方法；3) 实践：检验效果；4) 理论体系：推广。
- 7、机器学习：数据驱动的模式自动化构建方法
- 8、众包算法：人机结合克服 AI 的不足
- 9、对传统大数据算法的攻击：1) 通过伪造共同访问对推荐系统攻击；2) 搜索引擎优化（SEO）。
- 10、对机器学习算法的攻击（对抗样本攻击）：1) 诱导分类器产生错误分类；2) 诱骗视觉分类算法。
- 11、对抗样本的检测技术作为一种防御手段的补充应运而生。
- 12、除了对抗样本攻击，了解投毒攻击、后门攻击、模型萃取攻击、成员推理攻击、模型逆向攻击等多种针对机器学习模型的攻击方法。

CH9 隐私保护与安全监管（不考）

- 1、匿名通信（不考）
- 2、在线追踪（不考）
- 3、匿名支付（不考）