大数据安全与隐私保护

浙江财经大学数据科学学院 夏雨晴

xiayq0121@zufe.edu.cn

课程要求

- 考试时间：11月12日 9:00-11:00，考试地点：C201
- 题型：选择、填空、判断、名词解释、简答
- 成绩：期末考核60%+作业30%+签到10%

第一章 绪论

- 大数据概述
  - ✓ 大数据的特点：体量大、速度快、种类多、价值高
  - ✓ 种类：结构化数据、半结构化数据、非结构数据
- 课程安排
  - ✓ 大数据安全与隐私保护
  - ✓ 区别与联系
- 基本密码学工具
  - ✓ 加密技术（对称/非对称/混合加密）、数字签名技术、Hash和MAC、密钥交换技术
  - ✓ 具体方案：基于大整数因子分解实现公钥加密和数字签名（RSA算法）；基于离散对数问题实现密钥交换技术

第二章 安全存储与访问控制技术

- 背景
  - ✓ Unix系统的权限管理
- 早期访问控制技术
  - ✓ 基本概念：引用监控机、主体、客体、操作、访问权限
  - ✓ 访问控制模型
    - ✓ 自主访问控制模型：访问矩阵、能力表、访问控制表
    - ✓ 强制访问控制模型：BLP（机密性、下读上写）和Biba（完整性、上读下写）
    - ✓ 基于角色的访问控制模型：RBAC0～3四个模型及其相互关系
    - ✓ 基于属性的访问控制模型：各组成部分的功能及流程
  - ✓ 局限性
- 基于数据分析的访问控制技术
- 基于密码学的访问控制技术

第二章 安全存储与访问控制技术

- 背景
- 早期访问控制技术
- 基于数据分析的访问控制技术
  - ✓ 角色挖掘技术
    - ✓ 基于层次聚类的角色挖掘技术：凝聚式的角色挖掘、分裂式的角色挖掘
    - ✓ 层次聚类方法存在的问题
    - ✓ 生成式的角色挖掘技术：LDA和ATM
  - ✓ 风险自适应的访问控制技术
    - ✓ 常见的风险要素
- 基于密码学的访问控制技术

第二章 安全存储与访问控制技术

- 背景
- 早期访问控制技术
- 基于数据分析的访问控制技术
- 基于密码学的访问控制技术：依赖于密钥的安全性，无须可信引用监控机
  - ✓ 基于单发送者广播加密的访问控制
    - ✓ 基本概念：参与方、用户密钥树
  - ✓ 基于单发送者广播加密的访问控制
    - ✓ 基本概念：参与方、公钥服务器
    - ✓ 基于公钥加密的访问控制过程：加密存储、授权、数据文件访问
  - ✓ 基于属性加密的访问控制
    - ✓ 基本概念：单/多属性权威方案的参与方、访问结构

第三章 安全检索技术

- 基本概念
  - ✓ 密文检索的概念及参与方
  - ✓ 密文检索系统的流程
  - ✓ 密文检索分类：应用场景和数据类型
- 早期安全检索技术
  - ✓ PIR问题（定义3-1）、PIRK问题、SPIR问题、ORAM技术
- 对称密文检索
- 非对称密文检索
- 区间检索

第三章 安全检索技术

- 基本概念
- 早期安全检索技术
- 对称密文检索
  - ✓ 适用场景和算法组成：Setup算法、BuildIndex算法、GenTrapdoor算法、Search算法
  - ✓ 基于全文扫描/文档-关键词索引（布隆过滤器）/关键词-文档索引的方案
  - ✓ 三种方案的效率对比
  - ✓ 安全性：非适应性/适应性语义安全
- 非对称密文检索
  - ✓ 适用场景和算法组成：Setup算法、BuildIndex算法、GenTrapdoor算法、Search算法
  - ✓ 三种非对称密文检索方案在通信量、服务器端存储量、检索效率和加密效率上的对比
- 区间检索

第三章 安全检索技术

- 基本概念
- 早期安全检索技术
- 对称密文检索
- 非对称密文检索
- 区间检索
  - ✓ 早期工作：基于桶式索引的方案、基于B+树的加密方案
  - ✓ 基于矩阵加密/谓词加密/等值检索/保序加密的方案
  - ✓ 四种方案优缺点比较

第四章 安全处理技术

- 同态加密技术
  - ✓ 同态加密：算法组成、正确性、全同态加密、语义安全性、紧凑性
  - ✓ 自举加密
  - ✓ 具体加密方案（对称/非对称类同态加密方案、全同态加密方案）：基于近似最大公因子问题和稀疏子集和问题
- 可验证计算技术
- 安全多方计算技术
- 函数加密技术
- 外包计算技术

第四章 安全处理技术

- 同态加密技术
- 可验证计算技术
  - ✓ 基本概念：比特承诺、交互证明、完全性、合理性、零知识证明/论证系统、完美/统计/计算不可分的概念及性质
  - ✓ 基于承诺的可验证计算：朴素地证明x=y的协议（协议4-1）
  - ✓ 基于同态加密的可验证计算：线性MIP和线性PCP，具体方案（协议4-8+协议4-9）
- 安全多方计算技术
  - ✓ 百万富翁问题
  - ✓ 参与者分类：诚实、半诚实、恶意
  - ✓ 模型分类：半诚实模型、恶意模型
- 函数加密技术
- 外包计算技术

第四章 安全处理技术

- 同态加密技术
- 可验证计算技术
- 安全多方计算技术
- 函数加密技术
  - ✓ 函数加密技术定义
  - ✓ 公钥加密和谓词加密都是函数加密的特例
- 外包计算技术
  - ✓ 多个服务器的外包计算方案（协议4-14）
  - ✓ 两个服务器的外包计算方案：弱秘密隐藏假设
  - ✓ 单个服务器的外包计算方案：抢秘密隐藏假设

第五章 隐私保护技术

- 基本知识
  - ✓ 隐私保护方案的参与方：个人用户、数据采集/发布者、数据使用者、攻击者
  - ✓ 分类：根据隐私保护需求、根据数据类型
  - ✓ 典型的隐私保护技术手段包括：抑制、泛化、置换、扰动，裁剪等
- 关系型数据隐私保护（k-匿名）
  - ✓ 身份匿名：链接攻击、k-匿名
  - ✓ 属性匿名：同质攻击、熵L-多样化、递归(c,l)-多样化、t-贴近
- 差分隐私
  - ✓ 基本差分隐私：拉普拉斯差分隐私机制及证明、序列组合性、并行组合性
  - ✓ 本地差分隐私：Rappor协议和SH协议
  - ✓ 两者的区别