浙江大学 ZHEJIANG UNIVERSITY

蚂蚁集团 ANT GROUP

SECRETFLOW 隐语

# 隐私计算与数据合规

## 第一课：数据安全导论

授课老师：张秉晟

bingsheng@zju.edu.cn

# 教学安排

- 授课时间：
  - 周二： 第6、7节课 13:25 – 15:00
  - 周四（双周）： 第9、10节课 16:15 – 17:50
- 授课方式：
  - 现场教学（紫金港西2-410） + 学在浙大
- 考察方式：
  - 期末考试（30%）：随堂考试
  - 课程项目（40%）+ 课程实验（20%）
  - 出勤考察（10%）：助教记录

# 个人简介 BINGSHENG ZHANG

## 张秉晟

国家级青年人才项目获得者

- 教学经历：
  - 2019.9 - 至今： 浙江大学 “百人计划（A类）”研究员
  - 2015.9 - 2019.9：英国兰卡斯特大学 终身副教授
    - 网络空间安全中心负责人 （Head of security group）
    - 英国GCHQ认证的第一批 8 所ACE-CSR
    - 网络空间安全硕士学科主任
    - 英国GCHQ首批完全认证的英国4个网络空间安全硕士点之一
- 研究领域：
  - 密码学、安全多方计算、区块链、零知识证明、数据安全
- 学术成果：
  - 科技部国家重点研发计划项目“首席科学家”
  - 中国密码学会密码数学理论专业委员会委员、SIGSAC China委员、CCF数据治理发展委员会执行委
  - ISO/IEC 27565 《基于零知识证明的隐私保护指南》国际标准主编、ISO/IEC 27574《脑机接口隐私》国际标准联合编辑

# 现代社会数字信息化

## 现代社会高度数字化和信息化

**数字金融**

- 区块链
- 数字货币
- 银行业务
- 线上支付
- 智能IC卡

**电子政务**

- 智慧城市
- 数字政府
- 智能电网
- 电子公民信息
- 在线社区


**互联网领域**

- 互联网
- 云计算/存储
- 大数据
- 5G/6G
- 社交网路

**物联网领域**

- 物联网
- 车联网
- 工控系统
- 无人系统
- 嵌入式系统

# 网络空间安全涵义

- 网络空间安全（Cyberspace Security）是研究网络空间中的信息在产生、传输、存储、处理等环节中所面临的威胁和防御措施，以及网络和系统本身的威胁和防护机制。
- 不仅包括传统信息安全所研究的信息的保密性、完整性和可用性，还包括构成网络空间基础设施的基础设施的安全和可信。

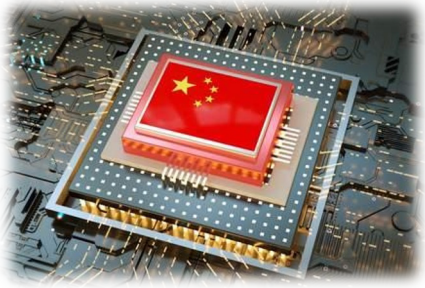
# 网络空间安全战略地位

## 网络空间安全是综合型、应用型学科

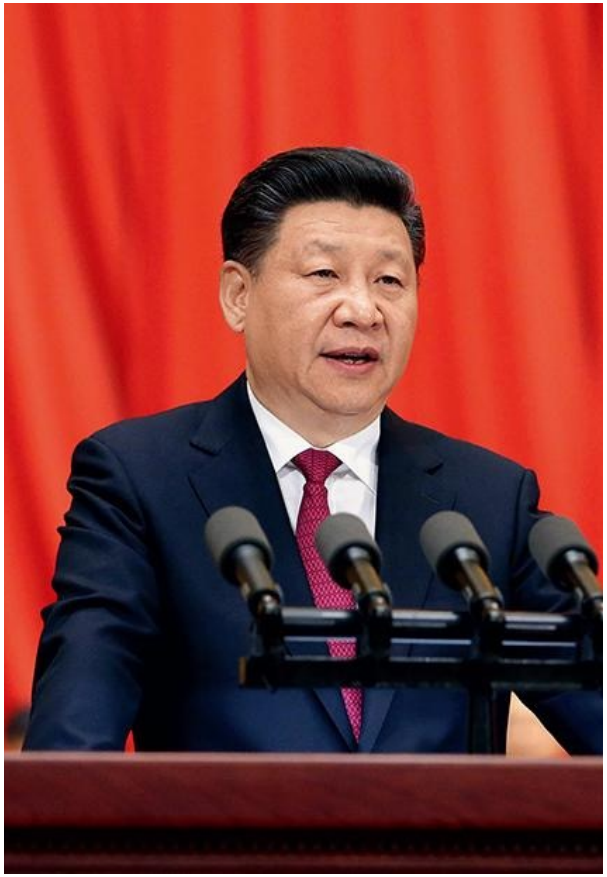
- 网络空间安全是计算机、电子、通信、管理、法律等融合而成的综合型、应用型学科，并非纯理论研究。
- 网络空间安全相关技术覆盖面广，迭代速度快，技术壁垒明确。目前，部分关键性技术仍需攻关和验证。
- 网络空间安全的理论体系尚未形成，相关的基础理论亟需构建，以适应快速发展的数字生态。

## 国际网络空间安全治理面临的困境

- 网络空间安全问题是当今时代背景下世界各国面临的最为严峻、最为复杂的问题之一。
- 网络空间安全的对象复杂且日趋泛化，所需保护的范围急剧扩大，治理监管能力亟待提升。
- 网络空间安全的核心技术研发和标准建设亟待加强，以实现技术自强自立，增强国际话语权。



# 我国高度重视网络空间安全战略重要性



**2014　中央网络安全和信息化领导小组第一次会议**

**没有网络安全就没有国家安全**。要有高素质的**网络安全和信息化人才队伍**。

**2016　网络安全和信息化工作座谈会**

培养网信人才，要**下大功夫**、**下大本钱**，请优秀的老师，编优秀的教材，招优秀的学生，建一流的网络空间安全学院。

**2017　《网络安全法》实施**

我国**网络领域的基础性法律**，明确了网络空间主权的原则，加强对个人信息保护。

**2018　全国网络安全和信息化工作会议**

我们要掌握我国互联网发展主动权，**保障互联网安全**、**国家安全**，就必须突破核心技术这个难题。**网络安全为人民，网络安全靠人民**。

**2021　《数据安全法》、《密码法》、《个人信息保护法》实施**

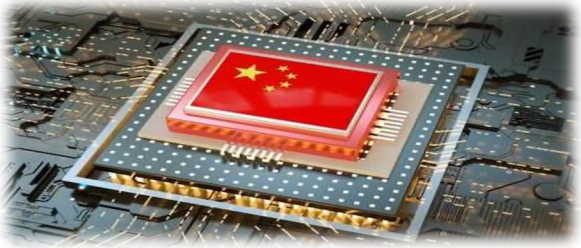
标志着数据安全工作首次升至**国家安全最高监管层级**。

# 数据要素

**数据被列为继土地、劳动力、资本、技术之后的第五大生产要素**

## 国家高度重视

- 我国将数据要素视为数字经济发展的核心引擎，十八大以来总书记多次提出“推动大数据和数字经济相关战略部署”

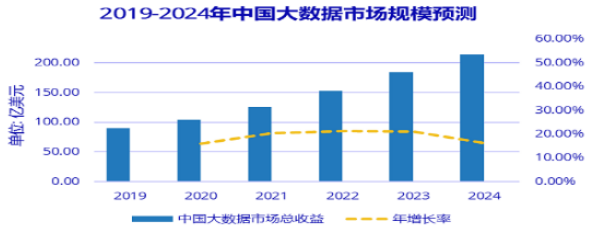


## 数据要素赋能新产业

- 数据要素与传统制造业相结合，形成数据采集、分析、处理等一系列以数据为核心的新业态，不断催生数据关联企业，实现数据产业化

## 数据市场规模日益增加

- 2023年全球大数据行业市场规模将达800亿美元，同比增长25.5%
- 2024年，我国大数据行业市场规模将超220亿美元




**激发数据要素在数字经济发展的重要作用，确保数据安全至关重要！**

# 数据安全重大意义

## 世界各国在数据安全领域积极进行布局

### 美国

2022年6月，《**美国数据隐私和保护法案**》：建立消费者数据隐私保护权利的美国联邦标准
2020年5月《**加州消费者隐私法案**》：全美最严隐私法

### 英国

2020年3月，《**国家数据战略**》：明确要从国家层面建立一套安全可信的数据机制，保障英国的数字安全

### 欧盟

2018年5月，《**通用数据保护条例**》：适用于所有涉及处理28个欧盟成员国公民信息行为的一致性隐私保护法

### 中国

2023年3月，第十四届全国人大《2023年政府工作报告》：

- 加强**网络、数据安全和个人信息保护**

2022年6月，中央深改委第二十六次会议：

- 审议通过**《关于构建数据基础制度更好发挥数据要素作用的意见》**

2021年3月，国家“十四五”规划和2035年远景目标纲要：

- 建立健全国家公共数据资源体系，确保**公共数据安全**
- 统筹数据开发利用、**隐私保护**和**公共安全**
- **加强网络安全保护**

2020年3月，中共中央 国务院《关于构建更加完善的要素市场化配置体制机制的意见》

- 加快培育**数据要素**市场，加强**数据资源整合**和**安全保护**

# 我国的数据安全相关立法

**2021 年《数据安全法》与《个人信息保护法》的出台，与2017年《网络安全法》一同形成了数据合规领域的“三架马车”，这标志着数据合规的基本法律架构已初步搭建完成**

**数据安全与个人信息保护达到立法高度**

- **《中国人民共和国网络安全法》** → 网络运营者收集、使用个人信息，应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意
- **《中国人民共和国数据安全法》** → 国家支持数据开发利用和数据安全技术研究，鼓励数据开发利用和数据安全等领域的技术推广和商业创新，培育、发展数据开发利用和数据安全产品、产业体系
- **《中国人民共和国个人信息保护法》** → 个人信息处理者应当定期对其处理个人信息遵守法律、行政法规的情况进行合规审计

网络安全法中 “个人信息”被提及21次之多；数据安全法中鼓励数据开发利用和数据安全等领域的技术推广和商业创新；个人信息保护法中首次对个人信息保护合规审计提出了明确要求，这为下一步各行业数据合规合法处理与流动，充分挖掘数据价值提供了法律保障。

# 数据安全重大事件

## 国家安全

### ■ 西工大机密数据遭窃取

2022年，美国国家安全局下属的特定入侵行动办公室TAO被曝出对西北工业大学进行**长期网络攻击**。TAO窃取西工大远程业务管理账号口令、操作记录等**关键敏感数据**，进而以“合法”身份进入中国某**基础设施运营商**服务网络，**控制相关服务质量监控系统**，窃取隐私数据

### ■ 乌克兰战争军事机密泄漏

2023年，五角大楼的一批**乌克兰战争作战数据简报** 在社交网络上泄露，包括战争伤亡人数、乌克兰防空地图、韩国为乌克兰提供33万发弹药的详细计划、埃及秘密为俄罗斯提供4万枚火箭炮等信息。据报道，泄露的机密文件不只俄乌战争，还有中东、印太等有关美国国安的议题。

*Ukraine War Plans Leak Prompts Pentagon Investigation*

Classified documents detailing secret American and NATO plans have appeared on Twitter and Telegram.

# 数据安全重大事件

## 国家安全

### ■ Facebook信息泄漏引发政治风波

2018年Facebook被曝出将**8000万用户数据**与剑桥分析公司分享，后者基于这些数据定制了2016年美国大选政治广告，助力特朗普胜选。2019年美国联邦贸易委员会对Facebook开出**50亿美元罚单**。近日，Facebook母公司Meta宣布将支付**7.25亿美元以和解用户集体诉讼。**



### ■ 滴滴数据跨境泄漏

2021年滴滴赴美上市后立刻遭到网信办审查，其应用程序因存在严重违法违规收集使用个人信息问题而被下架。此外，传闻称滴滴可能在上市过程中向境外泄漏中国道路信息和国人出行数据。2022年**滴滴被罚80.26亿元，并在美退市。**


# 数据安全重大事件

## 商业利益

- **企业技术资料遭窃取**

**微软——**与Bing 和 Cortana 在内的数百个项目相关的**37GB源代码遭窃取**。

**三星——**数据勒索组织从三星服务器窃取了大约**200GB的压缩材料**，其中包括敏感文档、代码等。

**索尼——**多年来从用户资料、商业数据、到PS设计图**累计几百TB**的数据被放上暗网交易。

**英伟达——**英伟达服务器遭黑客访问，包括驱动程序、原理图以及固件在内的**1TB数据**被泄露。

**中国工商银行——**2023年11月8日，中国工商银行的美国子公司(ICBCFS)遭受了LockBit勒索软件攻击，导致部分系统中断，约**90亿美元**的业务受到影响。

# 数据安全重大事件

## 商业利益

### OpenAI遭集体诉讼

加州律师事务所起诉OpenAI从互联网上抓取了数百万人未经同意的个人信息，要求**30亿美元**赔偿，如BERT、RoBERTa等其他LLM 大模型也纷纷中招。


### LLM使用导致三星机密泄露

三星半导体部门的工程师使用ChatGPT 参与修复源代码问题。过程中输入了包括新程序的**源代码**本体、**内部会议记录**等机密数据，导致数据外泄。


### 学而思抓取数据训练大模型

笔神作文6月起诉学而思，称其近期上线的“作文AI助手”用爬虫盗取数据，笔神作文超过**500万篇作文素材数据**资源短时间被爬取了超过两百万次。

# 数据安全重大事件

## 个人隐私

- **上海国家警察数据库超过十亿中国大陆居民资料泄漏**

2022年， 上海国家警察数据库遭到入侵，数据被标价为10比特币出售。这批数据总计超过**23TB**，涉及超过**十亿中国大陆居民**包括姓名、地址、出生地、身份证号码、照片、手机号码以及刑事案件信息。


- **45亿条国内个人地址信息遭到泄漏**

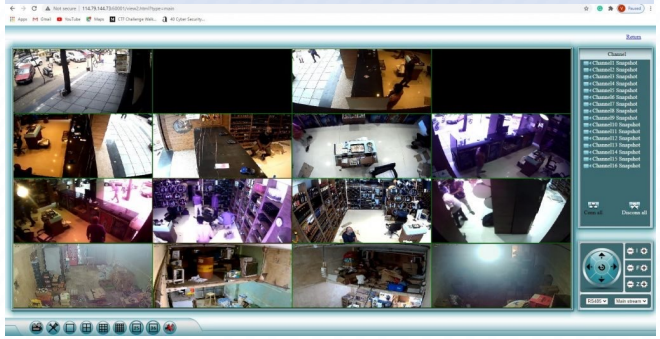
2023年，Telegram上爆出超**45亿条**国内个人地址信息遭到泄露，数据包括真实姓名、电话、地址等信息，并公开了免费查询渠道，据调查相关数据被大量用于电信诈骗、钓鱼邮件等犯罪活动。

# 数据安全重大事件

## 个人隐私

### ■ 智能安防用户数据遭泄露

智能安防企业SenseNets被曝泄漏了超过**256万**用户的身份证号、地址、工作单位以及基于摄像头所记录的过去24小时内经过的地点等信息。服务对象包括上海市公安局、兰州银行、上海第十人民医院等



### ■ 智能车用户数据泄露

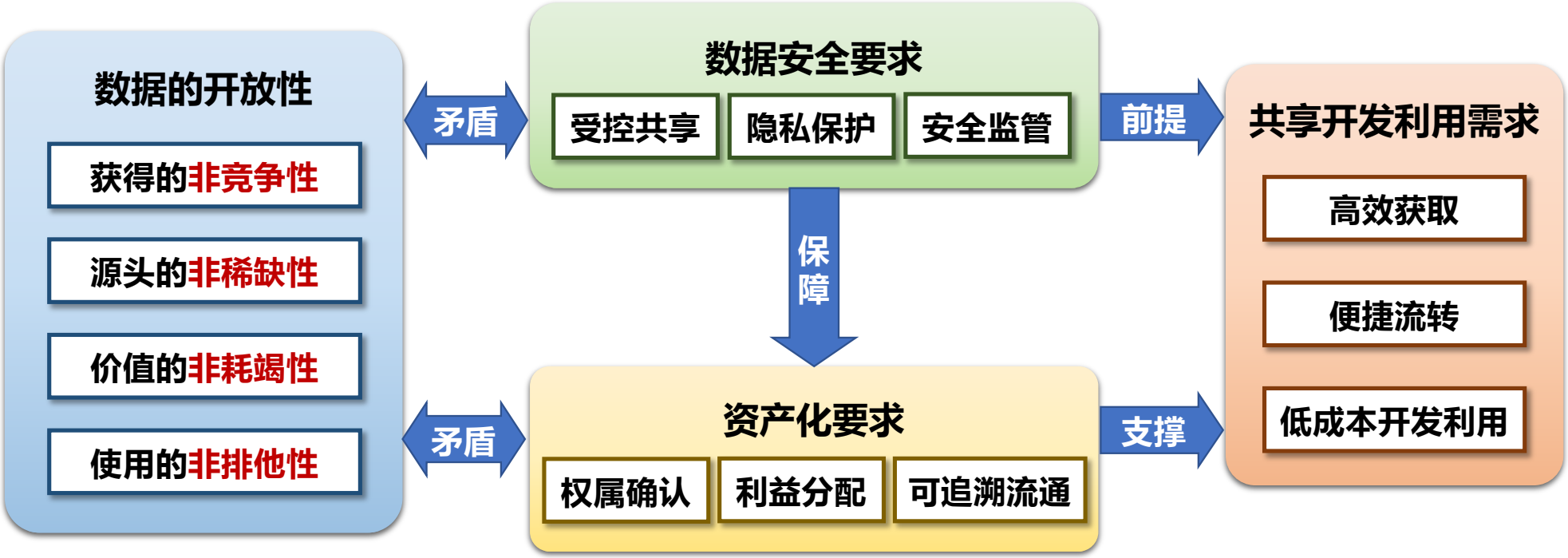
2022年，蔚来汽车因服务器配置错误导致**百万条蔚来的经营及客户隐私数据泄露，**包括蔚来员工信息、车主身份证、用户地址，甚至车主贷款数据等；并遭受**225万美元**等额比特币的勒索。

### ■ 智能医疗疫情管控数据泄漏

2022年， 上海疫情防控工具随申码数据库以**4000万美元**被 暗 网 拍卖，其中包含**4850万**用户的上海随申码数据，包含疫情期间居住或到访过上海的所有人的身份证、姓名及手机号。

# 数据要素安全前沿技术

研究**数据要素安全**相关技术，在**保障数据安全**的前提下实现数据**高效受控共享和有序开发利用**



# 数据要素安全前沿技术

## 数据在全生命周期过程中面临严峻安全挑战

### 数据泄露
- 数据被偷取、窃听、窃取、流量分析、公开数据推理等而造成的数据泄露

### 数据破坏
- 数据被篡改、删除、假冒，系统设备感染病毒，电磁干扰等造成的数据破坏

### 隐私泄露
- 通过数据分析、处理或推理等造成的用户身份、社交关系等个人隐私的泄露

### 数据失控
- 新型数据处理或应用模式（例如云计算）导致的数据失控

### 数据滥用
- 数据被非授权或越权使用
- 数据不可溯源、不可追踪

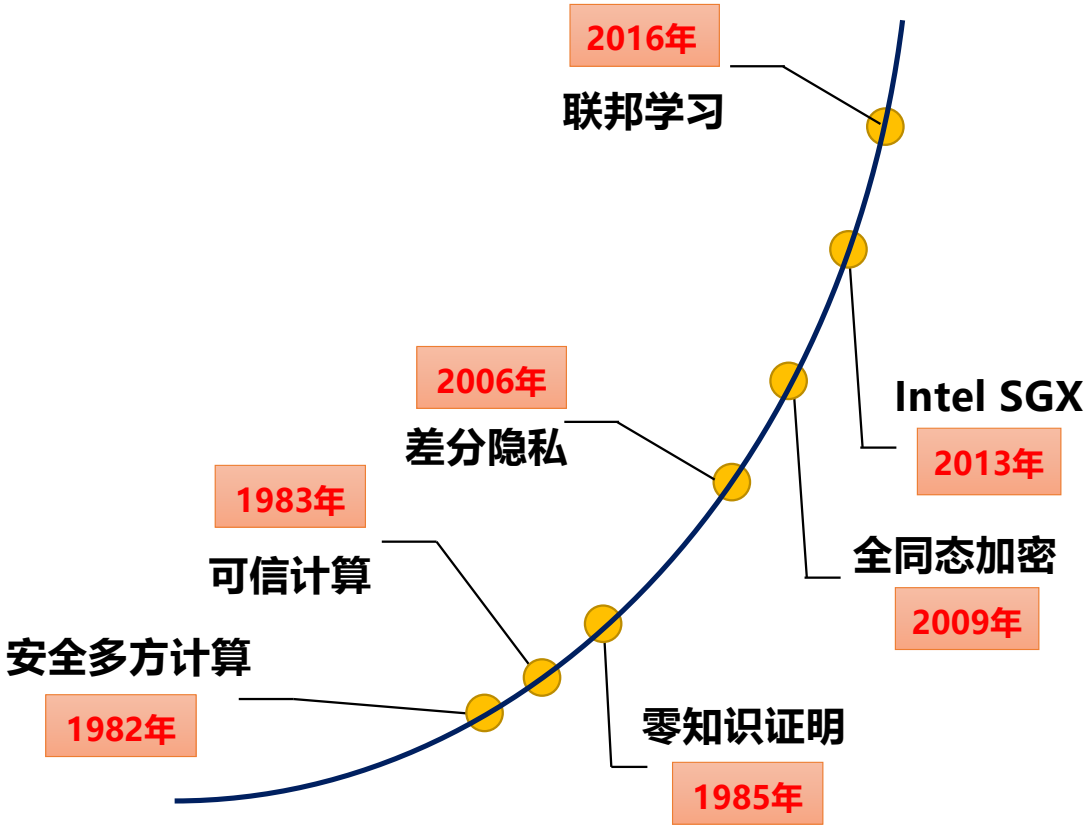
### 数据损坏或丢失
- 设备损坏、人为操作失误、自然灾害、电源供给故障等造成的数据损坏或丢失

## 支撑数据安全全生命周期的矩阵式技术体系

| 数据生产 | 数据采集 | 数据存储 | 数据传输 | 数据交换 | 数据共享 | 数据分析 | 数据使用 | 数据销毁 |
|---|---|---|---|---|---|---|---|---|
| 数据溯源 | 访问控制 | 密钥管理 | 身份认证 | 访问控制 | 数据脱敏 | 数据去重 | 数据脱敏 | 安全审计 |
| 安全审计 | 身份认证 | 访问控制 | 密钥交换 | 数字水印 | 数据格式 | 隐私保护 | 认证授权 | 永久删除 |
| 隐私保护 | 质量控制 | 存储加密 | 传输加密 | 认证授权 | 受控共享 | 数据挖掘 | 隐私保护 | 内容销毁 |
| ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |

| 安全多方计算 | 加密数据库 | 零知识证明 | 形式化验证 | 数据要素安全机制 | 抗量子密码 | 访问控制 | 全同态加密 | 安全计算模型 |
|---|---|---|---|---|---|---|---|---|

# 隐私计算技术分类




**隐私计算**

两个或多个参与方在不泄露各自数据的前提下通过协作对他们的数据进行联合计算处理。

V.S.

安全性　　高效性

# 隐私计算技术需求

隐私数据特点

- 数据海量
- 结构复杂
- 场景多样
- 管理困难

隐私计算需求

数据安全

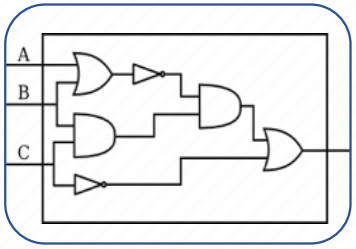
- 隐私性

数据效用

- 可用性
- 高效性


- 隐私性：有效保护主体的数据安全与隐私
- 可用性：不影响原有业务结果的正确性
- 高效性：计算开销在业务执行的允许范围内

# 什么是安全多方计算?


## 通俗定义

安全多方计算（secure multiparty computation）

- 密码学研究的一个重要分支
- 为解决一组互不信任的参与方之间在保护隐私信息以及没有可信第三方的前提下协同计算问题而提出的密码协议与理论框架。

## 狭义定义

狭义的安全多方计算主要包括以下两种实现方式：

- 针对布尔电路以姚氏混淆电路方式实现的两方协议
- 针对布尔电路或者代数电路以秘密分享方式实现的两方或者多方协议

## 广义定义

广义的安全多方计算包括通过以下技术在内实现的隐私保护多方计算协议：

- 全同态加密
- 可信硬件
- 联邦学习
- 第三方辅助服务器

# 安全多方计算的分类

## 通用安全多方计算

目标：
- 支持大多数（P/Poly）计算任务

方法：
- 实现常用基本计算算子协议，例如加，乘，比较，矩阵运算
- 将具体计算任务分解到基本算子

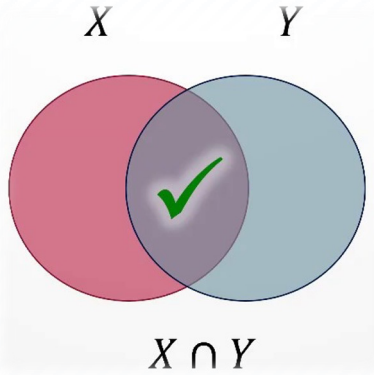
计算任务的表示形式：
- 布尔电路
- 代数电路
- RAM模型

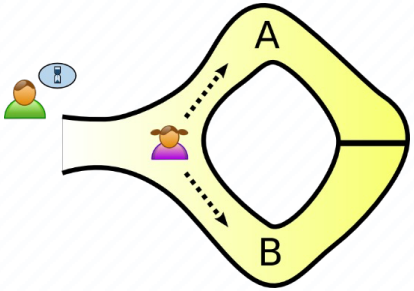
## 专用安全多方计算

目标：
- 高效实现专用实用计算任务

方法：
- 针对专用计算任务和应用场景定制多方安全计算协议

常见的专用协议：
- 隐私保护求交集（PSI）
- 隐匿查询（PIR）
- 零知识证明（ZK、SNARK）
- 联邦学习（FL）
- 电子投票（e-voting）



# 安全多方计算的“三维”

## 性能

- 通信量限制
- 通信轮次限制
- 参与节点算力是否对称
- 参与方数据是否平衡
- 两方/三方/多方

## 安全假设

- 网络假设：同步、异步、半同步
- 敌手模型：半诚实、恶意、隐匿作恶
- 敌手门限：<1/4、 <1/3、 <1/2
- 入侵假设：静态、自适应、动态
- 设置假设：PKI、TEE、RO、CRS
- 算力假设：DDH、RLWE

## 安全保障

- 隐私性（Privacy）
- 正确性（Correctness）
- 公开可验证性（Public Verifiability）
- 健壮性（Robustness）
- 公平性（fairness）


浙江大学
ZHEJIANG UNIVERSITY