北京郵電大學
Beijing University of Posts and Telecommunications

# 大数据安全与隐私保护

石瑞生
网络空间安全学院

# 安全多方计算的概念与应用

# 什么是安全多方计算（Secure Multi-Party Computation）？

- 首先，我们来考虑一种场景：1）两或更多方参与基于他们各自私密输入的计算。2）而且他们都不想其他方知道自己的输入信息。
- 那么，在保护输入数据私密性的前提下，如何实现这种计算?这就是“安全多方计算(Secure Multi-party Computation)”要研究的问题。

安全多方计算（Secure Multi-party Computing，SMC）是指多个参与方，每一个参与方拥有一个秘密信息，他们希望利用这些秘密信息作为输入，共同计算一个函数。例如，一个协会希望知道协会成员的平均收入，每个人又不希望泄漏自己的收入信息。

解决上述问题的策略之一是假设存在可信任的第三方。但是在目前多变和充满恶意的环境中，这是极具风险的。
因此，可以支持联合计算并保护参与者私密的协议变得日益重要。

# 什么是安全多方计算？

- 安全多方计算是无可信第三方的保护隐私计算协议。
  - 通常讲，一个安全多方计算问题在一个分布网络上计算基于任何输入的任何概率函数，每个输入方在这个分布网络上都拥有一个输入，而这个分布网络要确保输入的独立性，计算的正确性，而且除了各自的输入外，不透露其他任何可用于推导其他输入和输出的信息。
- 由此可见，通过安全多方计算技术，即实现了数据的共享，又保护了参与方的隐私信息，是大数据服务中实现安全与隐私保护的有力工具。

允许互不信任的参与方根据自己的输入计算任意一个函数，计算过程不泄露除输出结果以外的任何信息。

# 起源

- 安全多方计算问题最早是由著名的计算机科学家、2000年图灵奖获得者姚期智教授提出的，即百万富翁问题。
  - 这个问题是说，在没有第三方参与的情况下，两个百万富翁能够在互相不暴露自己的财产数额的情况下，比较谁更富有。
  - 1982年，姚期智首先介绍了安全多方计算的概念，并提出了著名的百万富翁问题。百万富翁问题就是两个富翁能够在相互不暴露自己的财产数额的情况下，比较谁更富有。
- 随后，Goldreich，Micali（2012图灵奖得主），Wigderson对该问题进行了推广，提出了具有密码学安全的安全多方计算协议，可以用来计算任意函数。1988年，Goldwasser，Chaum等人从理论上证明了安全多方计算的可解性。

安全多方计算的理论研究主要以以色列学者Goldreich等人的工作为主，研究工作已经得到了一般安全多方计算问题都是可解的结论。但是，这些协议几乎都需要使用电路计算、陷门置换等概念，不能直接用于具体的安全多方计算。

# 安全多方计算模型

- 安全多方计算模型有两种：半诚实模型和恶意模型。
  - 半诚实模型：如果所有参与者都是诚实的或者半诚实的，称此模型为半诚实模型，其中的攻击者是被动的。
  - 恶意模型：存在恶意参与者的模型称为恶意模型，其中的攻击者是主动的。

# 安全多方计算模型

- 参与者模型
  - 诚实参与者：在协议执行过程中，诚实参与者完全按照协议的要求完成协议的各个步骤，同时保密自己的所有输入、输出及中间结果。
  - 半诚实参与者：在协议执行过程中，半诚实参与者完全按照协议的要求完成协议的各个步骤，但同时可能将自己的输入、输出及中间结果泄漏给攻击者，也可以根据自己的输入、输出及中间结果推导其他参与者的信息。
  - 恶意参与者：在协议执行过程中，恶意参与者完全按照攻击者的意志执行协议的各个步骤，不但将自己的所有输入、输出及中间结果泄露给攻击者，还可以根据攻击者的意图改变输入信息、中间结果信息，甚至终止协议。

# 安全多方计算模型

- 攻击者模型
  - 攻击者是指企图破坏协议安全性和正确性的人。对攻击者进行分类时，可以有不同的分类准则，这些分类准则主要有攻击者的计算能力、网络同步状态、对恶意参与者的控制程度和动态性。
- 按照计算能力分类：按照攻击者的计算能力可以将攻击者分为拥有无限计算能力的攻击者和拥有有限计算能力的攻击者。
  - 对于拥有无限计算能力的攻击者而言，不存在诸如大素数分解困难等数学难题。在无限计算能力的攻击者模型下的安全的多方计算协议为信息论安全的多方计算协议。
  - 对于拥有有限计算能力的攻击者而言，破解目前公认的数学难题是不可能的。在有限计算能力的攻击者模型下的安全的多方计算协议为密码学安全的多方计算协议。

# 应用与挑战

- 几个实际生活中的例子
  - Alice认为她得了某种遗传疾病，想验证自己的想法。正好她知道Bob有一个关于疾病的DNA模型的数据库。如果她把自己的DNA样品寄给Bob，那么Bob可以给出她的DNA的诊断结果。但是Alice又不想别人知道，这是她的隐私。所以，她请求Bob帮忙诊断自己DNA的方式是不可行的。因为这样Bob就知道了她的DNA及相关私人信息。
  - 经过一次花费昂贵的市场调查后，A公司决定扩展在某些地区的市场份额来获取丰厚的回报。同时，A公司也注意到B公司也在扩展一些地区的市场份额。在策略上，两个公司都不想在相同地区互相竞争，所以他们都想在不泄露市场地区位置信息的情况下知道他们的市场地区是否有重叠。(信息的泄露可能会导致公司很大的损失。比如另一家对手公司知道A和B公司的扩展地区，提前行动占领市场；又比如房地产公司知道A和B公司的扩展计划，提前提高当地的房租等等)所以他们需要一种方法在保证私密的前提下解决这个问题。
  - 两个金融组织计划为了共同的利益决定互相合作一个项目。每个组织都想自己的需求获得满足。然而，他们的需求都是他们自己专有的数据，没人愿意透露给其他方，甚至是“信任”的第三方。那么他们如何在保护数据私密性的前提下合作项目呢?

# 应用与挑战

- 1)不可验证。
  - 这些方法忽略了对参与方输入、输出的验证，其正确性都依赖于计算参与方完全诚实遵循协议进行计算。计算的可验证性是一个非常重要，却被长期忽略的问题。
- 2)开销大。
  - 这些协议通常计算开销都非常大，难以应用于实际系统。

# 阅读材料 -


1) Hastings, Marcella, Brett Hemenway, Daniel Noble, and Steve Zdancewic. "Sok: General purpose compilers for secure multi-party computation." In *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 1220-1237. IEEE, 2019.
2) https://github.com/MPC-SoK/frameworks
3) David Evans, Vladimir Kolesnikov and Mike Rosulek, *A Pragmatic Introduction to Secure MultiParty Computation*. NOW Publishers, 2018. (This version: April 15, 2020)

北京郵電大學
Beijing University of Posts and Telecommunications

# 感谢聆听！