多方安全计算研究综述

蒋凯元
（国家公共信用信息中心　北京　100045）
（jiangky@cegn.gov.cn）

Review of Multi-Party Secure Computing Research

Jiang Kaiyuan
（*National Center for Public Credit Information*, *Beijing* 100045）

Abstract　With the rapid development of the Internet, data resources have become an important competitiveness of all industries. However, as the owners and users of data cannot be unified, problems such as data security and personal privacy become increasingly serious, resulting in the phenomenon of "data islands". Secure multi-party computation (MPC) promises to solve these problems by ensuring both privacy of data input and correctness of data computation, as well as ensuring that data input from participating parties is not compromised through protocols without third parties. Based on the definition and characteristics of multi-party secure computing, this paper introduces the research status, component model, and application scenarios of multi-party secure computing.

Key words　multi-party secure computing; security model; personal privacy; cryptography; distributed computing

摘　要　随着互联网的快速发展，数据资源已经成为各行业的重要竞争力，但由于数据的拥有者和使用者无法统一，数据安全和个人隐私等问题日益加剧，因而产生了"数据孤岛"现象.多方安全计算技术能够同时确保数据输入的隐私性和数据计算的正确性，并且在无第三方的情况下通过协议保证参与计算的参与方输入的数据不泄露，有望解决以上类似问题.从多方安全计算的定义和特点出发，介绍多方安全计算的研究现状、组成模型和应用场景.

关键词　多方安全计算；安全模型；个人隐私；密码学；分布式计算

中图法分类号　TP309

在数据作为生产要素的大背景下，数据安全治理成为数字经济发展的重要前提和保障[1].为了提高数据安全能力，建设数据健康生态，各类专家学者探索和研究了不同的技术路线.多方安全计算由于其较高的理论价值和广阔的应用前景，成为近年来密码学界乃至科技行业的研究热点，国外研究团队将多方安全计算与云计算、大数据等技术结合进行研究探索，并尝试商业化落地.

传统的安全计算是指能够对数据进行计算同时还能保护数据隐私的计算方式，常用的算法有

收稿日期：2021-10-26

差分隐私、同态加密等.而多方安全计算由我国的姚期智教授在1982年首次提出,它主要研究的问题是如何在无可信第三方的情况下设计一个函数,可以让多方在不透露任何信息的前提下安全地得到输出.多方安全计算的组成利用了许多密码学知识,如零知识证明、数字签名等,也利用了分布式计算原理,如广播问题和Byzantine问题.但多方安全计算与传统的密码学、分布式计算之间又存在很大差异.

传统密码学是在非安全的环境下,通过加密的方式实现数据的保护作用.加密机制就是将原有的信息进行某种规则的变换,即使在传输过程中数据被篡改和泄露也能及时发现,及时补救不用担心数据被利用.密码学保护的是参与方在攻击方攻击下保护数据的隐私性.而多方安全计算研究的问题是在各数据参与方之间如何对各参与方的数据进行保护和确保正确.而不同于分布式计算的是,多方安全计算通过协议来控制计算进程,无第三方控制计算过程,各参与方的地位平等,不存在第三方的干预.

文献[2]研究了多方安全计算特定应用场景的匿名化认定,分析了目前匿名化与去标识化相关规定在适用上可能面临的问题,并从个人信息保护与数据流通的角度,对匿名化与去标识化及其相关规定提出建议.文献[3]构造了一个安全计算集合成员关系问题的多方协议.通过将判定集合成员关系问题转化为范德蒙行列式求值问题,该协议解决了已有研究成果中集合阶数的泄露问题,提高了安全性.

为了设计安全的多方计算(multi-party computation, MPC)协议,人们进行了许多研究,如不经意传输、乱码电路、同态加密和线性秘密共享方案[4].安全的MPC提供了增强的隐私性、正确性和输入的独立性,并保证了输出的交付.区块链非常适合安全的MPC协议,因为它们都处理分布式环境中的安全和信任问题[5-6].利用基于区块链的安全MPC受益的实际场景很多,如健康数据统计分析、匿名电子投票、首次公开发行(IPO)和边缘计算等[7].许多研究人员一直尝试将安全MPC与区块链结合起来处理隐私和信任问题.Zhou等人[8]使用同态加密、秘密共享和零知识证明构造了一个公开可验证的安全MPC协议,该协议由2部分组成,主要包括链上计算阶段和链外预处理阶段,并将该协议作为链码的一部分集成,以此保护交易数据的隐私.

多方安全计算早期用于电子选举、门限签名、电子竞拍等场景.随着研究的深入,现已拓展至面向分布式场景的协同计算,包括隐私信息查询、计算预测、联邦机器学习等,并且在政务、医疗、金融等领域具有广阔的应用前景.例如,2019年8月,谷歌(Google)开源了多方安全计算工具——Private Join and Compute,以帮助组织和机构协同处理机密数据集;2019年10月,脸书(Facebook)将安全机器学习(secure machine learning)框架CrypTen进行开源.我国相关机构和组织积极推动多方安全计算核心技术研发、标准规范制定以及商业应用落地.例如,蚂蚁金服推出了蚂蚁链摩斯多方安全计算平台;华控清交基于多方安全计算技术,实现了高性能通用的安全计算框架PrivPy平台;矩阵元推出了隐私机器学习开源框架Rosetta.

1 多方安全计算的概念

多方安全计算数学定义:假设存在n个参与方$P_1,P_2,\cdots,P_n$,每个参与方都有一个私有输入数据x_i,所有参与方共同计算某个函数$f(x_1,x_2,\cdots,x_n)$,且要求在计算结束时,每个参与方P_i只能得到私有输入数据x_i的输出,而不能获取其他参与方的输入信息及输出结果信息[9].多方安全计算技术架构如图1所示.

当MPC计算任务开启时,通过路由寻址的方式,根据所需类型选择数据进行协同计算,根据MPC节点的计算,从本地数据库中查询数据,进行计算.整个计算任务过程中,数据始终存在于本地数据库中,因此不存在数据泄露问题,且能根据数据参与方的需求进行数据的共享,确保各个参与方都能得到真实数据,保证计算的正确性.

多方安全计算理论主要解决的是各个数据参与者的信息保护和计算正确性问题,能够实现在无第三方的条件下,在保护数据不泄露的前提下保证计算的正确性,所以多方安全计算的特点有:

1)输入隐私性.多方安全计算在进行计算任务时根据MPC节点的计算,从本地查询数据,再根据计算任务进行数据计算,整个过程中,数据都

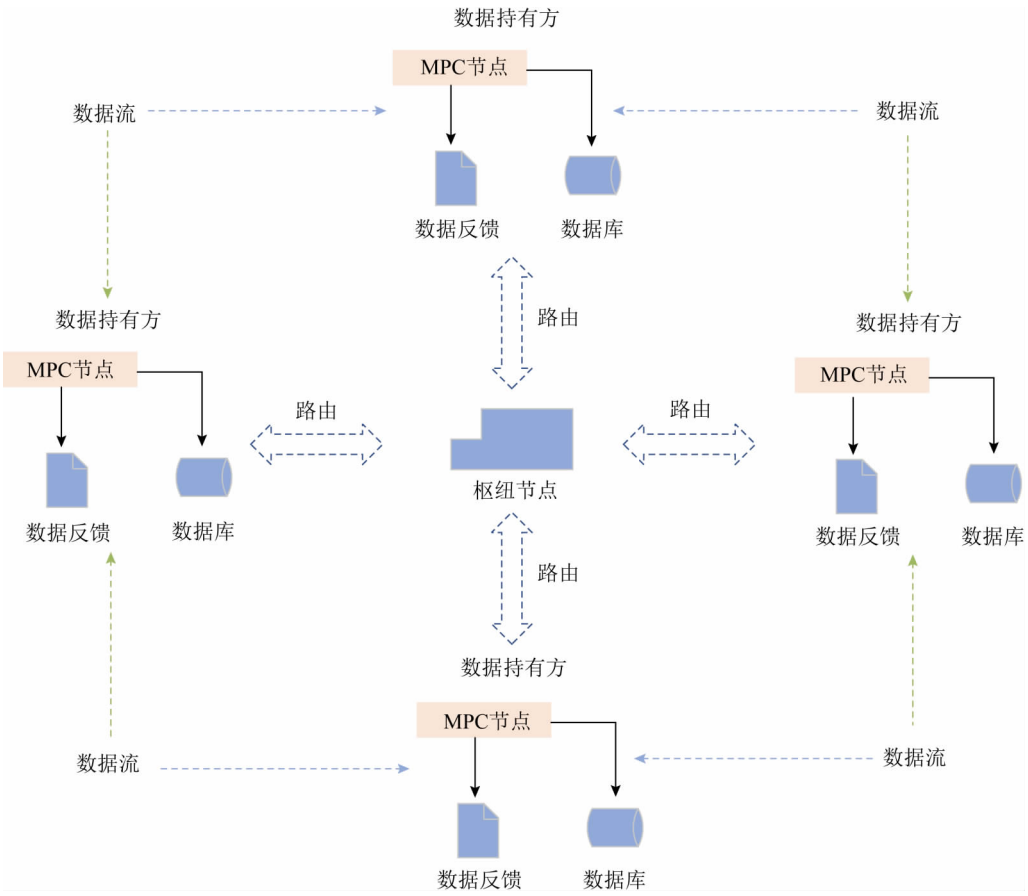

图 1　MPC 技术架构图

在本地数据库中保存，不存在数据泄露问题，因此保证了输入数据的隐私性.

2）计算正确性.多方安全计算数据参与方根据约定进行任务计算，通过多方安全计算协议进行计算数据的查询、协同计算，因此可以保证计算的正确性.

3）去中心化.多方安全计算不存在有特权的参与方或可信第三方，而是采用协议的方式代替第三方，通过协议保证各数据参与方地位权力平等，任何数据拥有者都可开启计算任务.

2　多方安全计算组成模型

多方安全计算的组成模型主要由 4 个部分组成：协议参与者、协议攻击者、网络条件和通信信道.

2.1　协议参与者

按照协议参与者在协议执行过程中的行为，可以把协议参与者分为[10-11]：

1）诚实的协议参与者.这类参与者是最理想的协议参与者，根据计算任务时约定好的过程，对照协议执行每个步骤.

2）半诚实的协议参与者.这类参与者不会像诚实参与者一样按照协议进行每个步骤，它根据现实情况私下收集所需数据，推导其他参与者的输入数据和输入值，但不会主动攻击或联合其他数据参与者破坏协议.这类参与者由于不主动攻击和联合其他参与者，一般很难被检测到，这对协议的安全性影响很大，一旦半诚实参与者被买通或攻破，其收集到的数据就会被泄露.

3）恶意参与者.这类参与者容易被攻击者买通，或者就是攻击者伪装形成的参与者，以此非法获取有用数据.

在现实情况下，主要存在的是半诚实协议参与者和恶意参与者，因此设计了半诚实模型和恶意敌手模型[12-13].

1）半诚实模型(the semi-honest model).在协

议执行时参与者按照协议规定的流程执行，但是可能会被恶意攻击者监听获取到在协议执行过程中参与者的输入输出以及在协议运行过程中获得的信息.

2）恶意敌手模型(the malicious model).在协议执行时，攻击者可以通过在其控制下的参与方进行不合法的输入，或者恶意篡改输入等方法来分析诚实参与者的隐私信息.还可以通过提前终止和拒绝参与等方式导致协议的终止.

2.2 协议攻击者

协议攻击者和协议恶意参与者参与协同计算的目的相同，都是通过非法途径来获取数据.与恶意参与者行为不同的是，它可以控制参与者，篡改协议参与者的步骤，使参与者按照其意愿继续执行协议来获取信息.

攻击者对不诚实参与者的控制可以分为以下2种情况：

1）被动攻击/窃听者.仅仅窃听信道或者获取不诚实参与者在协同计算过程中所得到的信息，不诚实参与者仍然按照协议约定执行协议步骤.

2）主动攻击者.攻击者会改变不诚实参与者的行为，不仅仅窃听或者获取不诚实参与者在协议进行中所得到的信息，还使其按照自己的意愿参与协议来达到窃取信息的目的.

2.3 网络条件

多方安全计算的各数据所有者在进行协同计算任务时都需要通过网络媒介进行连接，在同步网络媒介中，所有的数据参与者将共同拥有同一个、全局性的时钟.所有的信息会在同一个时间段内送达，并且每个数据参与者都能在下一时间段内收到属于自己的数据信息.但在非同步网络媒介中，所有数据参与者无法拥有同一个、全局性的时钟.信息从某个数据参与者的本地数据库中发出去，需要经过若干个时间段，数据参与者的接收方才能收到属于自己的数据信息，并且收到的数据信息由于来自不同参与者，接收到的数据信息顺序可能不是真实的发送顺序.

2.4 通信信道

多方安全计算的参与者之间的网络媒介需要信道相连，来完成与其他参与者的数据交换.由于协议攻击者会对不诚实协议参与者进行一定程度的控制，所以将通信信道分为3个级别：安全信道、非安全信道、未认证信道[14].攻击者对于安全信道没有控制能力；对非安全信道可以窃听参与者的通信信息，但不能篡改内容；对未认证信道可以完全控制，甚至可以伪装成诚实的参与者参与协议.

3 多方安全计算应用领域

3.1 政务应用

政务数据的公开共享与数据交易可促进政府部门和商业机构的共同发展，但在数据的公开交换过程中会导致政务数据和商业机构查询信息的泄露，利用多方安全计算技术，可以使数据一直存在于政务系统的本地数据库中，在不泄露数据的前提下实现政务数据共享.

多方安全计算为商业机构提供了统一的数据标准，可实现在协同计算中的信息检索、查询、数据跟踪等功能，保证了数据的安全性、隐私性，解决了“数据孤岛”问题.

3.2 金融应用

金融领域的营销、风控、反诈骗等，需要机构形成用户的完整画像来评判用户信用，因此需要跨机构联合共同刻画用户画像.跨机构获取数据需要从多个数据结构中获取数据，然后进行数据挖掘与分析.但在数据的获取过程中可能导致数据信息的泄露，数据被复制粘贴，数据所有者的所有权被复制.多方安全计算技术可以在数据无需归集与共享的情况下实现计算，保护目标数据所有方的隐私及数据资产安全[15-16].

3.3 医疗应用

对病患来说，医疗数据具有一定的敏感性、隐私性，也导致医疗数据得不到最大限度的使用，使得医院、药企、设备供应商之间难以实现数据的交换和共享.利用多方安全技术，在相对封闭的医疗数据参与方之间建立一个统一的、安全可信的数据交换网络，可在不泄露医疗数据的前提下，各参与方都可获取其需要的数据，由此实现医疗数据的有效使用[17].

3.4 创新应用

多方安全计算技术在与边缘计算、区块链、联邦学习、5G等新技术的融合中，能够创造更多的新应用.如基于多方安全计算和区块链技术的联合

征信[18]，可实现在无可信中心节点保护各参与方商业秘密及隐私数据的前提下，开展征信查询业务，对于解决多头借贷和过度授信问题具有重要意义；如“联邦学习＋区块链”多方安全计算引擎系统研究[19]，可实现多方隐私数据共享，构建数据生态，打破数据孤岛，挖掘数据联合价值，从而实现多方安全计算.

4 结　　语

多方安全计算技术的研究是为了解决数据拥有者和数据使用者的矛盾，数据拥有者不想泄露数据资源，而数据所有者需要正确的计算结果.一般情况下数据的拥有者也是其他数据的使用者，这就需要数据参与方协同计算，各取所需.多方安全计算在无第三方的条件下，采用协议标准的解决办法，让多个参与方共同合作，协同完成计算，计算所需的数据在整个计算过程中始终保存在本地数据库，这就保证了输入数据的隐私性，各个参与方协同计算，任务完成后返回各自的计算结果，保证了计算的正确性.多方计算安全适用于政务、金融、医疗等多个领域，能够实现数据共享交换、数据查询、数据联合分析、数据安全存储等多个功能.

参　考　文　献

[1] 吴振豪，高健博，李青山，等. 数据安全治理中的安全技术研究[J]. 信息安全研究，2021，7(10)：907-914

[2] 庄媛媛，靳晨，何昊青. 多方计算特定应用场景的匿名化认定与建议[J]. 信息安全研究，2021，7(10)：896-906

[3] 张茜，苏烨，秦静. 集合成员关系判定的安全多方计算协议[J]. 山东大学学报：理学版，2020，55(4)：118-126

[4] 张磊，马春光，杨松涛，等. 基于属性基加密的用户协作连续查询隐私保护策略[J]. 通信学报，2017，38(9)：76-85

[5] 刘峰，杨杰，李志斌，等. 一种基于区块链的泛用型数据隐私保护的安全多方计算协议[J]. 计算机研究与发展，2021，58(2)：281-290

[6] 王斌，张磊，张国印. 基于多方安全计算的属性泛化 mix-zone[J]. 通信学报，2019，40(4)：83-94

[7] 周俊，沈华杰，林中允，等. 边缘计算隐私保护研究进展[J]. 计算机研究与发展，2020，57(10)：2027-2051

[8] Zhou Jiapeng，Feng Yuxiang，Wang Zhenyu，et al. Using secure multi-party computation to protect privacy on a permissioned blockchain[J/OL]. Sensors，2021 [2021-10-26]. https://doi.org/10.3390/s2104154

[9] 李冬梅. 若干外包云计算中隐私保护的研究[D]. 上海：上海交通大学，2018

[10] 张硕. 安全多方计算协议及其应用研究[D]. 北京：北京邮电大学，2021

[11] 慈尚. 云环境下面向隐私保护的密度峰聚类方法研究[D]. 芜湖：安徽师范大学，2020

[12] 苏冠通，徐茂桐. 安全多方计算技术与应用综述[J]. 信息通信技术与政策，2019 (5)：19-22

[13] 李安康. 基于函数加密的多方统计计算研究[D]. 武汉：武汉理工大学，2017

[14] 杨萱. 安全多方计算中的若干应用问题研究[D]. 郑州：解放军信息工程大学，2009

[15] 王云河，李艺. MPC 与金融应用场景[J]. 金融电子化，2021 (2)：20-22

[16] 谭培强，谢谨. 多方安全计算金融行业应用初探[J]. 金融电子化，2020 (12)：11-12

[17] 张舒黎，邓春华，胡松，等. 安全多方计算体系架构及应用思考[J]. 通信技术，2021，54(9)：2182-2189

[18] 许健，关杏元，刘曦子，等. 基于区块链和多方安全计算技术的联合征信应用[J]. 银行家，2021 (7)：116-118

[19] 赵东明，刘静，徐晨兴，等.“联邦学习＋区块链”多方安全计算引擎系统研究[J].电子技术与软件工程，2020 (21)：184-186

蒋凯元

硕士，高级工程师.主要研究方向为政务信息化工程建设管理.

jiangky@cegn.gov.cn