浅谈数据库的隐私计算

童咏昕[1] 李书缘[1] 毛 睿[2]
[1] 北京航空航天大学
[2] 深圳大学

关键词：数据库 数据全生命周期 隐私计算

近年来，我国数字经济蓬勃发展。数据资源正是其中的核心引擎。2020 年，中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，首次将数据列为与土地、劳动力、资本、技术同等地位的**生产要素**。2022 年 1 月，国务院办公厅印发《要素市场化配置综合改革试点总体方案》，提出探索**“原始数据不出域、数据可用不可见”**的新型交易范式，实现数据使用的“可控可计量”，推动完善分级分类的数据安全保护制度。

根据发展数字经济的政策指导，北京、深圳和上海等地先后成立大数据交易所，探索数据流通的基础设施建设。2021 年 3 月 31 日，北京国际大数据交易所（以下简称“北数所”）正式成立，成为国内首家基于新型交易范式建立的数据流通枢纽。北数所提供数据、算法和算力三类产品，它就像一所大型超市，既有“生鲜区”提供数据原材料，又有“食品百货区”提供各类数据成品，还有“加工区”提供计算服务。例如北数所的“企业普惠金融数字画像”应用，结合某银行北京分行自有数据，提供精准的小微企业数字画像服务，该过程仅交易计算结果而不交换数据。

隐私计算为实现上述交易中数据“可用不可见”的目标提供了有力支撑。数据安全流转是数据交易的底线。我国《数据安全法》与《个人信息保护法》分别于 2021 年 9 月与 2021 年 11 月开始施行，为数据的合规使用提供法律准绳。例如，为保护数据隐私安全，数据交易所通常先审查数据提供商的数据源是否合法？是否对预发布的数据进行适当的脱敏处理？这些问题在数据库领域早有研究，并涌现了匿名化等多种隐私计算方法，这些隐私计算方法成为大数据交易与数据要素流通服务的基础技术。

数据库的隐私计算技术不仅能支撑我国对数据要素发展的长远规划，也是国际领域近年来的热门研究方向。加拿大统计局指出，借助隐私计算技术，可以在遵循隐私政策的前提下充分挖掘数据价值。国际工业界与学术界对隐私计算技术展开了多方面研究，从应用程序（APP）采集用户信息到 DNA 数据共享，从可信硬件到密码学工具库，从专用算法到应用系统，均在数据库的隐私计算领域做出了颇具意义的探索尝试。

本文将从现今生活中的几个隐私安全问题切入，浅析数据库隐私计算技术的发展历史，探讨在新一波数据要素发展的时代浪潮下，隐私计算将何去何从。

何谓数据库的隐私计算？

中国信息通信研究院在 2021 年发布的《隐私保护计算与合规应用研究报告》与《隐私计算白皮书》中提出，**隐私计算是面向数据采集、传输、存储、处理、共享、销毁等全生命周期中隐私保护的计算理论和方法**，可以在不泄露数据提供方原始数据的前提下，完成对数据的分析计算。隐私计算是一套在数据所有权、管理权和使用权分离时，计算隐私

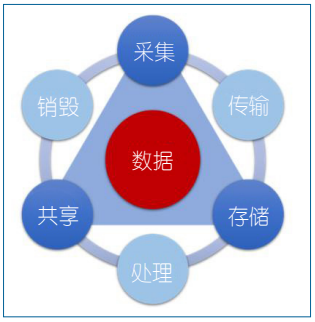

图1 数据全生命周期示意图

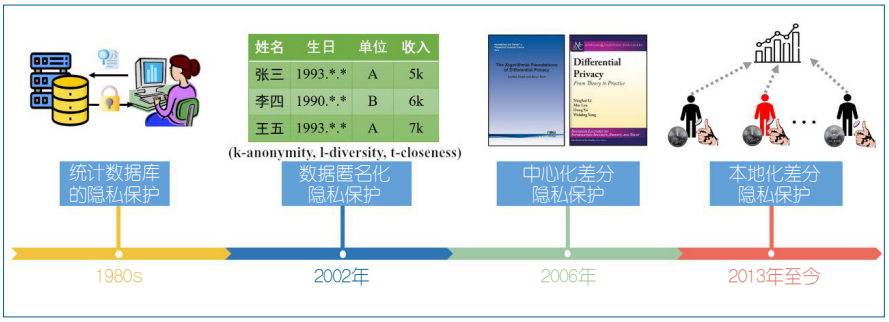

图2 数据采集隐私保护技术的发展历程

度量、隐私泄露代价、隐私保护与隐私分析复杂性的可计算模型与公理化系统。

隐私计算契合我国面向全生命周期的数据治理需求。“十四五”规划指出要做好数据采集、传输、存储、处理、共享、销毁等全**生命周期管理**（如图1所示），并明确指出我国大数据产业发展的四个制约因素之一为**安全机制不完善**，敏感数据泄露、违法跨境数据流动等隐患依然存在。而隐私计算包含同态加密、差分隐私和安全多方计算等多种技术方案，可以提供不同的隐私安全保护。其保护的对象可以是原始数据、计算结果，也可以是训练模型等。根据数据生命周期不同阶段产生的不同的隐私保护需求，可以选择不同的隐私技术方案。

数据库在数据全生命周期中提供数据采集、存储、共享以及对应的处理操作。因此本文将从数据生命周期中的采集、存储与共享三个阶段着手，追溯隐私计算在这三个阶段的研究发展脉络。此外，一些经典数据库安全问题，例如访问控制、数据审计等，在全生命周期的安全保障上也发挥重要作用，并在近年来结合区块链等技术焕发新的生机，本文对此不进行讨论。

数据采集：APP会泄露隐私吗？

如今手机上的各种APP为人们提供各类服务，覆盖了衣食住行等生活的方方面面。然而这些APP在提供服务时往往需要获取用户的信息数据，其中不乏一些敏感信息，如交通出行类应用需要获取用户的地理位置信息；购物类应用需要收集用户的住址信息；支付类应用则绑定了用户的银行卡、身份证，甚至指纹与人脸信息，等等。这些APP收集的各类信息，使用户仿佛成为“透明人”，令人不禁担忧这些APP是否会泄露隐私数据。

这一担忧关注的是全生命周期中数据采集阶段的隐私安全问题，在数据库领域中对应数据发布这一研究方向，早在20世纪就已有相关研究。根据内容的不同可以将数据发布分为发布原始数据、发布查询结果两类，下文将分别论述。图2展示了数据采集隐私保护技术的发展历程。

数据采集的隐私安全基础

最早关注发布结果中隐私安全问题的工作来源于20世纪80年代的统计数据库领域[1]。统计数据库是用于统计分析的数据库。为保证数据安全，统计数据库仅面向聚合数据而非单独某条数据做查询，因此一般仅支持求和、计数和求均值等聚合查询。然而结合多个聚合结果，仍然有可能推断出其中涉及的个体数据信息。加利福利亚大学霍默（Homer）教授的研究发现，通过分析DNA数据，可判断出某人是否参与了Genome-Wide-Association基因数据公开项目。随后美国国立卫生研究院（NIH）就移除了公开的dbGap数据集的所有聚合结果。因此该时期的一些研究工作通过向聚合结果添加噪声以实现隐私保护。

21世纪进入大数据时代，越来越多的服务平台如医疗机构、银行、电商和社交媒体等均通过收集大量的用户数据进行分析建模，从而提升平台的服务质量。随着数据挖掘的火热发展，各行各业对于数据集的需求也水涨船高。然而各个平台收集的数

据往往涉及敏感的个人隐私，因此多个国家和地区均对数据保护相关法律法规进行了完善更新。1998 年与 1999 年，英国与美国分别实施了《数据保护法》与《互联网个人隐私保护政策》。我国也在 2000 年 9 月实施的《互联网信息服务管理办法》中规定电信网络和个人信息的安全受法律保护。受制于这一时期的法律法规，各类机构显然不能直接发布原始数据。因此如何避免从数据集中推断出个人信息成为数据发布中需要解决的隐私安全问题。

为满足数据挖掘对原始数据的需求以及法律法规与隐私安全的约束，数据需要在预发布前进行脱敏处理。最简单直接的脱敏方式是将数据去标识符（identifier）后再发布，即去掉如个人的身份证号码、姓名等可以唯一识别个体身份信息的标识符。然而这种方法仍然存在还原标识（re-identified）的风险。2002 年，卡耐基梅隆大学的斯威尼（Sweeney）教授就还原出了一份公开医疗数据集中 87% 的记录对应的个体。此外，美国互联网公司 AOL 曾公开一批用户的搜索记录数据集，但去除了用户的标识信息。然而《纽约时报》通过通讯录信息仍还原出了其中部分用户的标识信息。

从 2003 年开始，陆续出现了一系列基于数据泛化、抑制和扰动等手段的隐私保护方法，例如 k 匿名（k-anonymity）、l- 多样性（l-diversity）、t- 保密（t-closeness）等，用以实现数据脱敏。较前述简单方式，匿名化方法能有效加大隐私保护力度，此类技术的核心思想是增强数据的不可区分性。以 k 匿名为例，其是指数据集中任一条数据都难以和其他 k-1 条数据区分开 [2]。可以通过数据抑制，即将一些属性或部分属性值变为星号，或是通过数据泛化，即将精确值变为一段数据范围等手段，实现 k 匿名，如表 1 所示。保护隐私的数据发布方法多种多样，根据发布次数不同，有针对单次发布和多次发布的方法；根据发布数据类型不同，有针对图数据、空间数据等类型的方法。

然而这些数据发布技术并非万无一失，总是存在一些难以抵御的攻击。因此，2006 年，哈佛大学德沃克（Dwork）教授延续了统计数据库中隐私保护的思路，提出了差分隐私（Differential Privacy，DP）技术 [3]。该技术有以下三个优点：（1）支持更通用的计算函数。20 世纪 80 年代的统计数据库仅支持简单的聚合查询，而差分隐私将其拓展至通用函数。（2）抵御更强的攻击。其安全模型假设攻击者拥有更多信息。（3）更严谨的数学模型。该技术可以从统计学意义上给出隐私安全的保证。此外，该方法通过添加噪声的方式保证添加或删除任意一条数据不会对查询结果造成很大的影响。添加噪声的大小与给定的参数 ε 以及数据本身的灵敏度有关。一个满足 ε - 差分隐私的随机算法 M 是指，对于两个相邻数据集 D 与 D'，M 任意可能的输出 o，算法 M 满足：

$$\frac{\Pr[M(D)=o]}{\Pr[M(D')=o]} \leqslant e^{\varepsilon}$$

其中 ε 称为差分隐私的隐私预算（privacy budget）。经过随机化算法处理后，两个相邻数据集输出相同结果的概率非常接近，从而使差分攻击无效。随机性是差分隐私的核心，可以通过对输出结果添加随机噪声实现，常见的噪声有拉普拉斯噪声与高斯噪声等。差分隐私技术应用示例如图 3 所示。

差分隐私技术近年来发展得如火如荼，并衍生出本地化差分隐私（Local Differential Privacy，LDP）、Rényi DP 等变种，以及置乱模型（shuffle model）等优化技术。本地化差分隐私产生自移动互联网时代，作用于移动端设备进行数据采集的阶段 [4]。前述介绍的 ε - 差分

表1　匿名化方法应用示例

姓名	省份	年龄	疾病
*	河*	15-25	心脏病
*	河*	15-25	颈椎病
*	河*	15-25	颈椎病
*	河*	15-25	心脏病

图3　差分隐私技术应用示例

隐私中均假设存在可信第三方为数据添加噪声。该可信第三方可以查看所有数据，并保证公布的结果不泄露单一数据。然而在现实应用中，往往不存在符合这些条件的第三方。例如一些生活服务 APP 会收集用户的地理位置信息，然而这些 APP 是不完全受用户信任的。因此本地化差分隐私作为一种无须可信第三方的隐私保护机制逐渐兴起。不同于差分隐私作用于整个数据集，本地化差分隐私以每一条数据为单位独立添加噪声，例如用户可以对自己的地理位置添加噪声后再发送给各类 APP。

数据采集的应用案例

保护数据采集过程的隐私计算技术在实际生活中已有许多落地应用。其中匿名化技术最为常见，例如，人们在各类应用程序上绑定银行卡后，一般仅显示银行卡号的头与尾 ；身份证号被绑定后，也会将中间的出生年月变成星号。

而差分隐私技术其实已经在各类 APP 中有所应用。以某手机操作系统中提供的服务为例，当 APP 向手机系统申请获取 GPS 位置时，系统会将该位置信息添加噪声，并将扰动后的信息发送给 APP，从而通过差分隐私技术保护用户位置信息的隐私安全。

数据存储：云服务安全吗?

如今人们已经习惯使用各类便捷的云服务，将文档、视频等存放在云盘，将手机等便携式移动设备上的数据传至云服务。云服务一方面可以节省设备本地存储空间，另一方面便于多个设备间同步数据。然而云服务却频频发生隐私泄露问题，例如苹果 iCloud 服务泄露用户照片的事件曾引起轩然大波。我们不禁要问 ：交给云服务的数据能得到安全保护吗？

这一问题关注数据在存储阶段的隐私安全保护，它源自数据库领域的数据即服务（Data as a Service，DaaS），也称为外包数据库（outsource database），旨在让第三方服务提供商向用户提供数据管理服务，从而减少用户自己部署数据库所需的硬

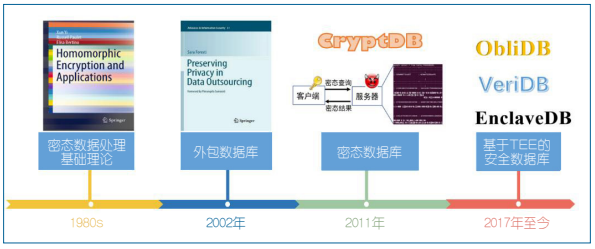

图 4　云服务的隐私安全技术的发展历程

件、软件以及人力等开销。21 世纪初，云计算兴起使用户可以将数据外包给云服务商。然而云服务导致数据不再完全受控于数据拥有者，存储在云服务器的数据可能会遭到恶意攻击，甚至服务提供商本身也未必可信，这会带来数据隐私泄露的风险。因此数据库领域也出现了一系列关于外包数据库隐私保护的研究工作。下面将简单介绍此类研究工作涉及的安全基础，回顾外包数据库领域的相关系统工作。图 4 回顾了云服务隐私安全技术的发展历程。

数据存储的隐私安全基础

在云服务场景下，为了保护数据隐私安全，用户需要将数据加密后存储在云端。而云端为了向用户提供一定的数据管理服务，则需要基于密态数据进行一定的计算，这称为密态数据处理。密态数据处理关注数据的加密以及如何在加密后的数据上进行各种类型的操作。密态数据处理对数据进行加密存储，在查询时用户将初始查询转化为加密后的查询发送给云服务器，云服务器将加密后的查询结果返回给用户进行解密以及进一步处理，从而得到真实结果。

密态数据处理需要使用保序加密（Order-Preserving Encryption，OPE）、同态加密（Homomorphic Encryption，HE）、可搜索加密（Searchable Encryption，SE）等技术。保序加密是指一种密文保持明文顺序的加密方式，可以基于密文进行比较排序。同态加密允许对密文进行特定代数运算，并且解密后的结果与明文进行相同运算的结果一致，因此可基于密文进行查询计算而无须解密数据。可搜索加密是指对数据文件加密后仍能对其进行检索的一类技术，可实现将加密文档存储在第三方的同时，提供

关键词检索、近似检索等功能。为提高查询与检索的执行效率，有研究工作尝试在密态数据上构建不同类型的索引 [5]。

以上技术主要针对数据的加密存储及对应处理过程中的隐私保护，然而云端对数据的访问过程还存在安全隐患。云服务器可以通过分析程序执行过程中对数据的访问模式（access pattern）推理出数据的隐私信息。此时可以借助不经意随机访问机（Oblivious Random Access Machine，ORAM）技术保护数据在不可信系统下的访问模式 [6]。该技术的主要思想是通过添加冗余的读写操作来保护原始的读写模式。基于该技术，Oblix 构建一种密态数据上的搜索索引，保护数据的访问模式并支持动态更新 [7]。

由此可见，针对云服务存储场景，密态数据处理提供了加密、处理、读写等多个步骤的隐私保护方案，具有较高的应用探索价值。

数据存储的系统构建

21 世纪初，外包数据库研究聚焦于关系型密态数据的查询处理技术，例如结合保序加密技术构建选择、投影与连接等关系运算符 [8]，实现对 SQL 查询的支持。后续研究也扩展到更丰富的数据类型上，如在空间数据上基于同态加密技术设计安全 k 近邻查询等方法 [9]，在半结构化数据和图数据上也均有相关研究。此外还有工作关注验证外包数据库的完整性、完备性等系统特性 [5,10]。

2011 年麻省理工学院研发了支持不同加密等级的密态数据处理系统 CryptDB[11]。该系统聚焦于关系型数据处理，可以对接 MySQL 系统。CryptDB 结合保序加密、同态加密等多种加密技术，通过感知 SQL 查询动态调整加密层级，从而提供更加灵活的多粒度隐私安全保护。

近年来以可信执行环境（Trusted Execution Environment，TEE）为代表的新型硬件技术也为密态数据处理系统注入新的活力。TEE 可以从硬件和操作系统层面为数据和程序提供一个隔离的运行环境，避免不可信的操作系统或应用程序窃取或篡改数据。代表性的 TEE 技术主要有 ARM 的 TrustZone 与英特尔的 SGX。

基于 TEE 技术，2018 年面世的 EncalveDB 系统将内存数据库置于可信内存中运行，从而保证该数据库的安全性 [12]。2019 年，ObliDB 结合 SGX 与不经意随机访问机技术构建了可保护数据访问模式的数据查询系统 [13]。2021 年推出的 VeriDB 系统则利用 SGX 技术构建低吞吐的可验证数据库系统 [14]。此外，TEE 还可与差分隐私、安全多方计算等多种技术结合。软硬件结合正在成为隐私计算的重要发展趋势之一。

数据共享：为什么是联邦计算?

近年来，随着人们对数据隐私的日益关注，各国均出台了相关法律来规范数据的管理和使用。2018 年 5 月 28 日，欧盟正式出台《通用数据保护条例》（*General Data Protection Regulation*，*GDPR*），该条例对企业收集、使用和共享用户数据做出了详尽的要求与说明，因此也被冠以“史上最严”的称号。2021 年亚马逊就因违反该条例被处以单笔高达 7.46 亿欧元的罚单，这也成为 *GDPR* 生效以来单笔金额最高的罚款纪录。在美国，加利福尼亚州的《消费者隐私法》（*California Consumer Priacy Act*，*CCPA*）于 2020 年 1 月正式生效；我国也在 2021 年相继施行了《数据安全法》与《个人信息保护法》两部数据隐私安全的相关法律。

各国保护数据隐私的措施日益严格，加剧了长久以来存在的“数据孤岛”问题。为破除数据孤岛、加强数据共享，联邦计算这一概念应运而生。联邦计算是指多个数据拥有方在原始数据不出本地的前提下，联合进行查询分析的一种计算范式。根据计算任务不同，联邦计算可以分为面向数据库查询任务的数据联邦与面向机器学习任务的联邦学习。联邦计算需要结合秘密共享、混淆电路等技术达到隐私安全保护的目的。下文将回顾联邦计算所需的隐私安全基础。不涉及隐私计算的传统数据共享技术不在本文的讨论范围内 [15]。图 5 展示了联邦计算的

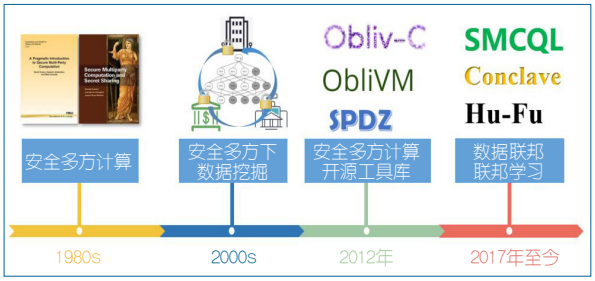

图 5　联邦计算的隐私安全技术的发展历程

隐私安全技术的发展历程。

数据共享的隐私安全基础

20 世纪 80 年代，姚期智最早提出安全多方计算的概念 [16]，这种计算方法旨在不泄露多方原始数据的条件下，共同完成给定的计算任务。安全多方计算可以通过经典的“百万富翁问题”具体地阐述：有两个百万富翁互相好奇究竟谁更富有，但同时又不想透露自己具体拥有多少财富，那么他们如何完成财富的比较呢？萨莫尔（Shamir）教授提出的秘密共享、姚期智院士提出的姚氏混淆电路等协议均能实现上述计算任务 [16,17]。

21 世纪初期，如何在数据挖掘过程中保护隐私安全也引起了来自数据挖掘和安全多方计算两个领域的学者的关注，基于安全多方计算的频繁模式挖掘、决策树分类和 K-means 聚类等方法 [18,19] 被提出。这一时期的研究主要受限于当时安全多方计算技术的发展状态，以算法设计与理论分析为主，缺乏实践验证。因此，安全多方计算的后续研究转向了以落地应用为目标的通用工具库的构建与高效专用协议的设计。

通用工具库是指面向通用安全多方计算构建的系统。在 2012 年以后涌现出一批安全多方计算开源的工具库，例如 ABY 工具库实现了布尔电路与算术电路，支持丰富的运算。后续出现了结合多种安全协议的工具库，例如开源库 SPDZ 内包含了多种混淆电路与秘密共享协议的实现。2017 年开始出现面向开发的工具库。马里兰大学开发的 ObliVM 为用户提供高层封装的类面向对象开发语言，可以将用户输入的程序编译为电路并执行。随后，2018 年面世的基于 C 语言开发的 OblivC 系统在性能表现上更胜一筹。此类面向开发的工具库大大降低了非专业人员使用混淆电路的开发门槛，例如对于矩阵分解问题，原本需要 5 名专业博士生耗时近 4 个月才能完成，而使用开源工具库仅需 1 名硕士生耗时 1 天即可解决。

安全多方计算在专用协议的性能上也取得了突破性进展。专用协议是指针对某一类求解目标设计的安全协议。例如，对于隐私保护集合求交（Private Set Intersection，PSI）这一问题可以设计出比通用协议性能高出多个数量级的安全协议。其中，面向电路的隐私保护集合求交协议（circuit-based PSI protocol）允许在该协议之后接入任意的安全多方计算函数，同时要求各参与方不能泄露交集之外的任何额外信息，如输入集合的大小等。因此，该类协议在数据库连接 – 聚合操作、数据对齐等问题上能够发挥重要作用 [20]。

联邦计算的系统构建

2012 年以来，ABY、SPDZ 等安全多方计算通用工具库的出现以及高效隐私保护集合求交专用协议的突破，推动了新一轮联邦计算的研究。

如上所述，得益于安全多方计算的技术突破，学者们能够相对容易地将安全多方计算技术引入数据库系统中，从而支持多方自治场景下保护隐私安全的数据查询，这一研究方向也被称为数据联邦（data federation）。近年来，学术界已经涌现了一系列在数据联邦上的探索性系统工作。例如，美国西北大学数据库团队于 2017 年研发的 SMCQL 系统是早期结合安全多方计算技术构建的数据联邦系统 [21]。该系统使用 ObliVM 将 SQL 执行计划编译为混淆电路执行，能够联合两个参与方的关系型数据库执行安全的 SQL 查询，并且不泄露除查询结果之外的任何其他数据。于 2018 年提出的 Conclave 系统则将数据联邦扩展到大数据处理引擎 Spark 上 [22]，并借助了较为成熟的商业安全多方计算库 Sharemind，能够联合三个参与方进行联合查询。以上系统均基于半诚实的安全模型，即参与者均遵守协议执行。而

之后的 Senate 系统则考虑如何在恶意模型下构建数据联邦系统 [23]。恶意模型下参与者可能违背协议，Senate 系统能够抵挡恶意攻击者且支持多个参与者。以上工作主要面向关系型数据。2022 年发布的虎符（Hu-Fu）系统进一步丰富了数据联邦系统支持的数据类型 [24]，该系统聚焦于时空类型数据，其底层能够对接多个时空大数据计算系统，如 PostGIS、SpatialLite 与 GeoMesa 等。

结语

本文面向数据全生命周期的隐私保护需求，从数据库视角出发，选择数据采集、存储和共享三个阶段，浅述了隐私计算技术的发展动向。然而就数据库隐私计算的宏观发展而言，未来还有以下方向值得深入思考与探讨。

隐私与性能的平衡 近年来尽管一些新型隐私安全技术在性能上已经有了突飞猛进的提升，但仍不足以适用于工业界的大规模落地应用。因此根据不同的应用场景，需要确定不同等级的隐私安全保护级别，并设计对应的技术方案以在隐私安全与性能之间寻得平衡。

全周期保护的流程 各种隐私安全技术方案在设计时所针对的隐私保护目标具有一定局限性，有些技术专注于保护计算过程中的数据安全，而有些技术则致力于保护数据采集或发布阶段的数据隐私。因此需要结合多种技术方案，构建数据生命全周期的隐私安全保护流程。

技术与法规的桥梁 本文从技术层面浅谈了数据隐私安全发展的一些动态。在应用层面，保护数据隐私安全需要以国家出台的相关法律法规为准绳。然而技术与法规之间还存在一定距离，如何判定技术方案是否合规还需要在二者之间搭建起桥梁。

2022 年 4 月 10 日国务院发布《关于加快建设全国统一大市场的意见》，再次强调了数据要素的重要价值，明确提出加快培育统一的技术和数据市场。数据库的隐私计算技术作为数据要素流通中不可或缺的重要一环，将迎来新一轮的蓬勃发展。 ■

童咏昕

CCF 高级会员,CCF 会员与分部工委副主任，CCF 走进高校工作组组长。北京航空航天大学教授。主要研究方向为联邦学习、隐私计算、时空大数据分析、数据库技术与群体智能。yxtong@buaa.edu.cn

李书缘

CCF 学生会员。北京航空航天大学博士研究生。主要研究方向为大数据分析处理、隐私计算与联邦学习。
lishuyuan@buaa.edu.cn

毛　睿

CCF 杰出会员。深圳大学计算与软件学院副院长。主要研究方向为大数据管理分析处理、高性能计算与数据库技术。
mao@szu.edu.cn

参考文献

[1] Beck L L . A security mechanism for statistical database[J]. *ACM Transactions on Database Systems*, 1980, 5(3):316-3338.

[2] Sweeney L. k-anonymity: A model for protecting privacy[J]. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2002, 10(05): 557-570.

[3] Dwork C. Differential privacy[C]// *ICALP 2006*:1-12.

[4] Cormode G, Jha S, Kulkarni T, et al. Privacy at scale: Local differential privacy in practice[C]// *SIGMOD*, 2018: 1655-1658.

[5] Li F, Hadjieleftheriou M, Kollios G, et al. Dynamic authenticated index structures for outsourced databases[C]// *SIGMOD*, 2006: 121-132.

[6] Pinkas B, Reinman T. Oblivious RAM revisited[C]// *CRYPTO 2010*: 502-519.

[7] Mishra P, Poddar R, Chen J, et al. Oblix: An efficient oblivious search index[C]// *S&P 2018*: 279-296.

[8] Hacigümüş H, Iyer B, Li C, et al. Executing SQL over encrypted data in the database-service-provider model[C]// *SIGMOD*, 2002: 216-227.

[9] Wong W K, Cheung D W, Kao B, et al. Secure kNN computation on encrypted databases[C]//*SIGMOD*, 2009: 139-152.

[10]Xie M, Wang H, Yin J, et al. Integrity auditing of outsourced data[C]// VLDB 2007: 782-793.
[11]Popa RA, Redfield CMS, Zeldovich N, et al. CryptDB: Protecting confidentiality with encrypted query processing[C]// SOSP 2011: 85-100.
[12]Priebe C, Vaswani K, Costa M. EnclaveDB: A secure database using SGX[C]// S&P 2018: 264-278.
[13]Eskandarian S, Zaharia M. ObliDB: Oblivious Query Processing for Secure Databases[C]// Proceedings of the VLDB Endowment, 13(2): 169-183.
[14]Zhou W, Cai Y, Peng Y, et al. Veridb: an SGX-based verifiable database[C]// SIGMOD 2021: 2182-2194.
[15]Doan A H , Ives Z , Halevy A . Principles of Data Integration[J]. Morgan Kaufmann Publishers Inc. 2012.
[16]Yao A C. How to generate and exchange secrets (Extended Abstract)[C]// FOCS 1986: 162-167.
[17]Shamir A. How to share a secret[J]. Communications of the ACM 1979, 22(11): 612-613.
[18]Agrawal R , Srikant R . Privacy-preserving data mining[C]// Proceedings of the 2000 ACM SIGMOD international conference on Management of data. ACM, 2000.
[19]Vaidya J, Clifton C. Privacy preserving association rule mining in vertically partitioned data[C]// SIGKDD 2002: 639-644.
[20]Wang Y, Yi K. Secure Yannakakis: Join-aggregate queries over private data[C]// SIGMOD 2021: 1969-1981.
[21]Bater J, Elliott G, Eggen C, et al. SMCQL: Secure query processing for private data networks[C]// Proceedings of the VLDB Endowment, 2017, 10(6): 673-684.
[22]Volgushev N, Schwarzkopf M, Getchell B, et al. Conclave: Secure multi-party computation on big data[C]// EuroSys 2019: 1-18.
[23]Poddar R, Kalra S, Yanai A, et al. Senate: A maliciously-secure MPC platform for collaborative analytics[C]// USENIX 2021: 2129-2146.
[24]Tong Y, Pan X, Zeng Y, et al. Hu-Fu: Efficient and secure spatial queries over data federation[C]// Proceedings of the VLDB Endowment, 2022, 15(6): 1159 - 1172.