WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

白皮书

金融服务业新一代数据共享：利用隐私增强技术解锁全新价值

世界经济论坛与德勤联合呈现

2019年9月

世界经济论坛
91-93 route de la Capite CH-1223
Cologny/Geneva
Switzerland
电话：+41 (0)22 869 1212
传真：+41 (0)22 786 2744
电邮：contact@weforum.org
网址：www.weforum.org

©2019世界经济论坛 版权所有 保留一切权利。严禁以任何方式（包括复印和刻录）或通过任何信息存储和检索系统复制或传播本出版物的任何内容。

本白皮书由世界经济论坛出版，致力于推动某一项目、领域洞察或某种互动的发展。本报告所述调查结果、诠释和结论均在世界经济论坛的推动和支持下完成，但并不一定代表世界经济论坛或其成员、合作伙伴及其他利益相关者的观点。

目录

序言 4

前言 5

第一章：金融行业隐私现状 6

数据共享的收益 6

数据共享的潜在弊端 6

改变数据共享现状 7

第二章：隐私增强技术 8

第一项技术：差分隐私 9

第二项技术：联合分析 11

第三项技术：同态加密 13

第四项技术：零知识证明 15

第五项技术：安全多方计算 17

第三章：金融服务业应用 20

为金融机构解锁新价值 20

为客户解锁新价值 22

为监管部门解锁新价值 24

结语 25

附录 26

技术优势和限制 26

相关阅读资料 29

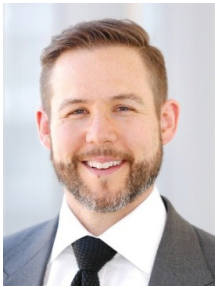
鸣谢 30

尾注 32

序言

Matthew Blake
世界经济论坛
未来金融和货币体系部门负责人

Jesse McWaters
世界经济论坛
金融业创新项目主管



Rob Galaski
德勤管理咨询
银行业及资本市场全球领导人

现今，数据之于第四次工业革命转型的重要意义不言而喻，数据被喻为新一代的石油、黄金，炙手可热。毫无疑问，在数据变得日益重要的同时，企业的工作重点也在转变。然而，媒体竞相报道各企业积累海量数据的竞赛，却甚少关注企业对发掘机构间数据共享潜能的兴趣。尤其在金融服务行业，企业对机构间协作的需求大大增加，涵盖了从改进欺诈检测手段到赋能新型个人理财咨询服务的各种应用场景。

当然，数据共享存在风险。企业在发掘数据潜在价值时，须妥善降低对客户隐私的影响、保障数据安全，并对竞争性敏感信息加以管控。从过往经验来看，金融服务行业在隐私保护与数据应用上的目标往往是矛盾的，需要在数据共享价值与潜在的隐私风险间进行权衡，这也直接导致许多原本似乎很有希望落地的数据共享项目被束之高阁。

新兴的“隐私增强技术”或将通过消除（或降低）过往的相关协作风险，从根本上推动数据共享领域的变革。随着隐私增强技术的成熟，企业会期望利用这些技术重新审核许多搁置的数据共享项目，借此探索此前难以实现的项目机会。

隐私增强技术可以为金融业带来巨大价值——前提是行业高管和监管部门能够了解这些需要应用到复杂数学和计算的技术以及具体应用。本文旨在概述当前最有前景的一些技术，帮助读者理解有关理论概念，并展示如何在金融体系中应用这些技术。我们希望通过这种方式助力打造高效协作的金融环境，期望金融机构、消费者和更广泛的金融体系都能从数据共享中受益。

前言

古有盲人摸象的故事，形容不窥得事物全貌无法得出正确的结论。如今，金融服务业面临着同样的问题。在“消费者（客户）是否值得信赖”、“交易商是否互相串通”或“某项交易是否是欺诈性交易”等重要问题上，每个机构都仅握有一块拼图（即数据），但因为各自手上的数据有限，金融机构就像故事中的盲人一样，都存在得出错误结论的风险。而信息分享是解锁事物全貌、全面了解事物的关键，可惜金融机构间数据共享并不容易。由于面临数据存储、管理和共享方面的诸多限制，因此，金融机构至今仍无法对客户和运营环境有全面的了解。

完整的数据才能发挥最大的价值，但获取最大价值的过程却十分复杂，其间还伴随目标冲突：例如，金融机构通过数据共享将能更好地识别隐藏的交易欺诈模式，减少金融犯罪检测误报。但是金融机构对于披露与自身客户有关的竞争信息相当谨慎，通常尽量避免违反隐私监管规定。值得一提的是，数据共享不仅能使金融机构受益，还能让客户得到更加个性化、更为具体和细致的建议，但客户可能担心自身信息被误用、滥用或在未经本人同意的情况下被共享。

这些都说明了数据共享的矛盾：数据共享可以创造价值，但对于信息被共享的个体而言，其隐私信息不再是秘密，对于开展数据共享的机构而言，其机密性也会受到影响。各方已投入大量精力希望以一种机构、客户、行业协会和监管机构都能够接受的方式平衡目标之间的冲突，保障金融体系的运营。“隐私增强技术”能够使机构、客户和监管部门在不损害“数据所有者”（客户）隐私和“数据管理者”（金融机构）机密性的情况下发掘共享金融数据的价值。这些技术并不新鲜，但近年来的巨大发展已使其从单纯的探索性研究转变为可服务于生产实践的技术，或将为数据共享带来根本性的改变。

本文就隐私增强技术的工作原理及其可能为金融机构带来的价值作简要概述，供各金融子行业（如保险、银行、投资管理）高管使用。我们将围绕下述议题进行分析和论述：

第一章：金融行业隐私问题现状概览

第二章：隐私增强技术工作原理介绍

第三章：隐私增强技术如何应用于数据共享

本文包含三个章节：

第一章：
金融行业隐私现状

第6页

第二章：
隐私增强技术

第8页

第三章：
金融服务中的应用

第20页

第一章：金融行业隐私现状

因在数据使用上相互竞争，金融机构往往难以就如何存储、管理和共享数据达成一致。金融机构、监管部门和消费者间的矛盾冲突也由来已久。

我们将在下文探讨这三个领域的目标冲突（数据共享的益处与弊端）。

数据共享的益处

金融机构可从以下三种数据共享方式中受益：

- 输入式数据共享（从第三方获取数据）
- 输出式数据共享（向第三方提供数据）
- 协作式数据共享（与第三方就形式相近的数据互通有无）

首先，输入式数据共享使机构可借助更多信息来丰富其决策系统，获得更高质量的输出结果，助力精准运营。例如，贸易公司可使用汤森路透的MarketPsych Indices[1]等第三方服务，基于社交媒体数据的分析支撑采购/销售相关决策，或更准确地了解市场行情；其次，输出式数据共享使机构可在自身缺乏相关能力的情况下，借他山之石攻玉（并最终使客户受益）。例如，智能投资顾问Wealthsimple可通过安全链接将客户投资组合信息导入Mint.com，[2]使客户可同时查看日常支出和投资余额，进而全面了解自身的财务状况；最后，协作式数据共享使机构能够获得单靠自身之力所无法得到的海量数据，从而收获更深、更广的洞察。例如，六家北欧银行近期宣布合作开发共享的“了解客户”（KYC）实用程序[3]，以强化其金融犯罪防御系统的能力。

对于监管部门而言，数据共享提供了将金融数据的控制权和所有权交还客户的机会，进而促进创新和竞争，这在监管法规中均有体现：如英国的《开放银行标准》（Open Banking Standard）、欧盟的《欧盟支付服务修订法案第二版》（PSD2）、澳大利亚的《消费者数据权利法案》（Consumer Data Right）以及新加坡、中国香港和日本所采取的其他形式的开放应用程序接口（API）监管条例。这些法规中均有相应条文规定机构应按客户要求将其拥有的客户数据（如交易数据）提供给经认可的第三方，使市场的新参与者得以访问这些数据并制定新的价值主张，监管部门认为这将最终改善公民的财务状况。[4]

对于客户而言，数据共享使其可获得更高质量的产品和更高效的服务。例如，Lenddo通过分析客户的社交媒体数据、通信数据和交易数据为其提供更高质（即准确性可能更高）的信用评分。[5]客户正逐步认识到个人信息的价值，愈发倾向于仅在交换中获得实际利益时才会共享其个人信息（直接向金融机构提供更多信息或授权其作为代表与第三方共享数据）。[6]

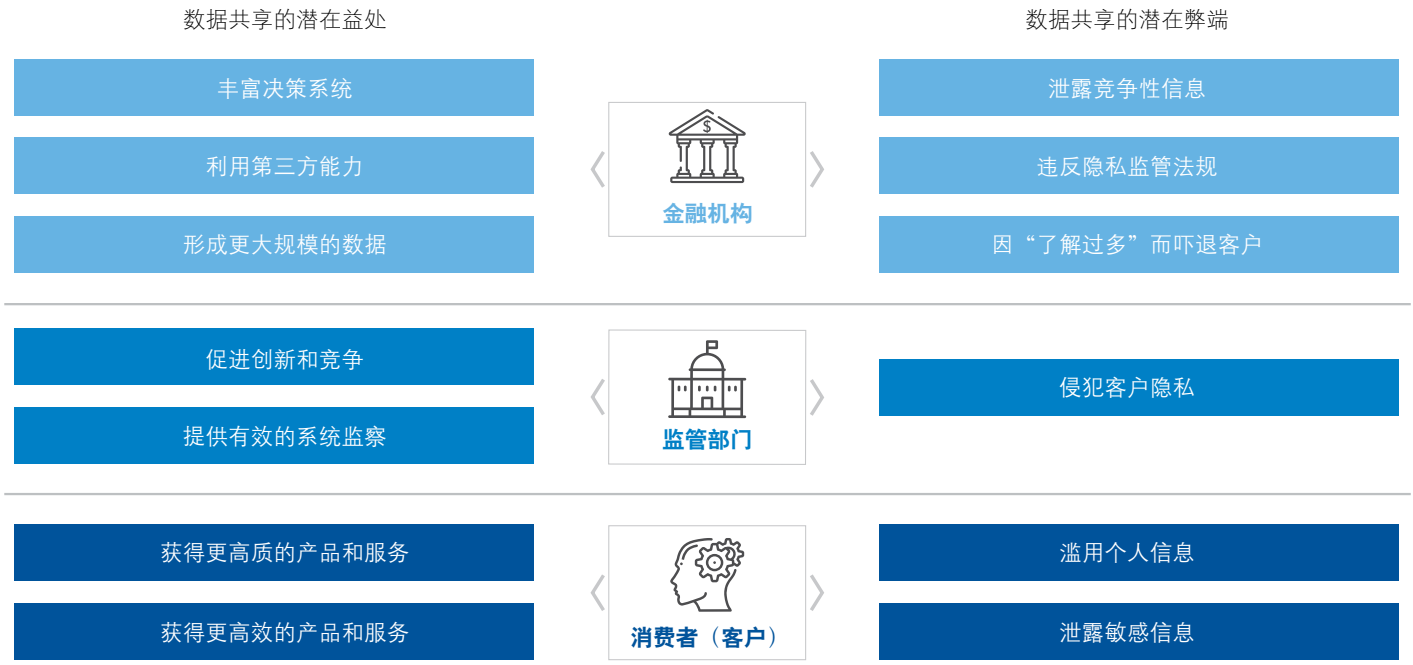
数据共享的潜在弊端

一些因素会阻碍金融服务中的数据共享。对于金融机构而言，任何输出式数据共享都使其面临可能被第三方滥用竞争性信息（如客户身份及特征）的风险；此外，数据共享还可能违反诸如《通用数据保护条例》（GDPR）等隐私监管法规，或因必需流程（如建立新机制以确保知情同意）过于复杂而导致投入超出数据共享所能带来的益处。随着人工智能和其他高级分析技术应用的不断增加，大型金融机构的高管已经开始担忧因掌握过多客户数据而使其感到不安进而对金融机构心生恐惧。

监管部门长期以来通过限制数据共享的方式达成其一项重要职责，即保持消费者金融和非金融信息的机密性。[7]例如，美国1999年《金融服务现代化法案》（Gramm-Leach-Bliley Act of 1999）要求金融机构了解其客户敏感信息是如何被共享的，并允许金融机构选择退出数据共享或采取特定措施保护共享内容。[8]近年来，全球监管部门还推出了新的更为严格的客户隐私保护要求：例如，欧盟的GDPR规定机构应让客户能够更便捷的查阅保存在机构处的个人信息；其他法规则禁止公司跨国共享个人身份信息（PII）以保护本国客户隐私，这或将阻止跨国机构分析其整个组织所拥有的内部数据。这些要求意味着无法共享某些类型的数据，或因共享变得过于昂贵、复杂且耗时而使机构不愿更多开展数据共享。

虽然客户寻求从自身数据共享中获得更多利益，但他们也愈发警惕其数据可能被滥用：Harris Poll的一项调查显示，只有20%的美国消费者“完全信任”与其打交道的公司会妥善保护其信息隐私。[9]

2018年发生的几起备受关注的安全和隐私泄露事件 (包括Cambridge Analytica[10]、Capital One[11]、Google +[12] 和 Aadhaar[13] 等) 无疑加剧了客户的担忧。客户担心其数据可能被用于会损害自身权益的情形 (如身份盗用)，更有可能被未经授权的第三方得知自己的隐私信息 (如敏感的购买记录)。[14]



改变数据共享现状

如图所示，金融服务业中的每个利益相关者都面临隐私相关冲突，而这些冲突一直阻碍着数据共享巨大价值的实现。新兴的隐私增强技术能够使金融机构、消费者和监管部门能够在竞争机会与保护义务之间取得平衡，实现既符合监管原则，又保护消费者隐私，并保持金融机构业务流程机密性的数据共享。这些技术或将扩大金融服务中数据共享范围，从而使金融机构掌握全局信息，进而为自身、消费者、监管部门乃至全社会创造全新价值。

第二章：隐私增强技术

数据作为第四次工业革命的动力，推动了人工智能和互联设备等新技术的发展。为了真正从这些新技术中受益，机构要充分利用内外部所能获取的数据。管理数据隐私的技术能够帮助机构发掘新价值。我们将在下文介绍五类关键技术[15]。

我们将探索各隐私增强技术的潜在益处，以假设的案例演示技术运用，通过既往隐私泄露实例展示相应隐私增强技术的用处，并评估其在金融服务中的可行性。随后，我们将探讨如何结合这些技术在金融行业打开数据共享协作的新局面。

第一项技术：差分隐私

概述：

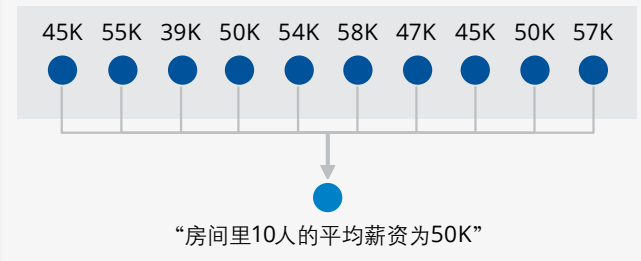
当机构想要与第三方共享数据时，删除或匿名化个人身份信息的方式并非总能充分保护数据库中个体的隐私。例如，将数据与其他数据集相结合就可重新识别数据库中的特定个体。对此，一种行之有效的解决方法是在流程中（输入、计算或输出）添加噪声，确保特定数据“行”的保密性，但仍可通过查询汇总数据获得有意义的洞察。例如，人口普查数据通常采用添加噪声的方式实现匿名化处理，以保护受访个体的隐私；美国也将在2020年的联邦人口普查中应用差分隐私技术。[16]

2006年，Cynthia Dwork等人[17]发表了有关“差分隐私”的标志性论文，提出了一种普遍适用的方法来计算为保护数据库中每一个体的隐私所需添加的噪声量，[18]后经大量深入研究提升其效率和可扩展性后，该方法目前已投入各种实际应用。目前，差分隐私已在苹果等公司的大规模生产中得以运用（如自动完成网上搜索[19]），并已嵌入各种广泛运用的分析和机器学习库中（如PyTorch[20]和TensorFlow[21]）。

注：差分隐私本身并非一种技术或机制，而是对添加噪声的各种技术和方法的一种度量，这些添加噪声的技术和方法可限制各不相关方试图从分析结果中推断出输入数据的能力。

原理揭示：

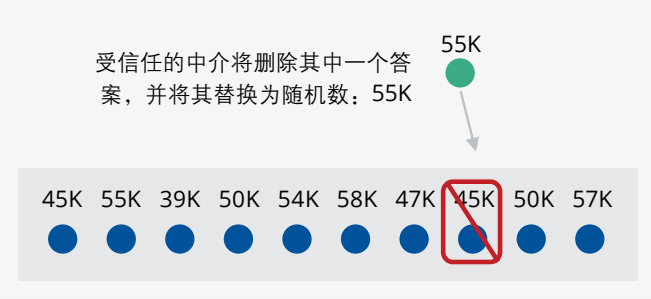
假设一组10名从事相同工作的个人想要共享薪资信息，以了解自己的薪资是偏高还是偏低，但又不想向其他任何人透露自己的实际薪资。为此，他们找到一个独立且受信任的第三方充当中介，中介会将所有人输入的信息匿名化，同时根据汇总数据得出有用洞察。中介对他们的数据取平均值，并告知10人的平均薪资为50K。这对于个人是很有用的信息，因为他们可以确定自己的薪水是偏高还是偏低。



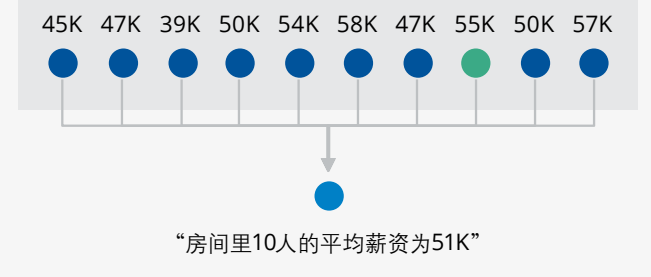
但是，如果某人已经知道房间中其他八人的薪资，只剩一人的信息未知。

一旦知道了房间里众人的平均薪资，此人就可推断出第十人的确切薪资为45K，并可公开或使用该私人信息。

为防止这种侵犯隐私的行为，中介可在他/她的平均值计算中添加噪声。例如，调查员可删除十人中某一人的答案，并用一个在收到的答案范围内（即39K和58K之间）的随机数进行代替。



随后照常计算平均薪资，中介提供了51K这一带有轻微噪声的答案，同时任何第三方都无法逆向分析得出输入数据。



知道其他八人薪资的那个人无法推断出房间里最后一人的确切薪资，因为添加噪声会带来两个不确定性：

- 八个已知薪资的任何一个都可能被一个未知数所代替，当得知平均数为51K时，只能推定未知薪水范围为36K-74K。而这个范围太大，没有任何价值。
- 如删除的刚好是未知薪资，甚至无法逆推出薪资范围。

想要窥探他人隐私的人不知道发生了以上两种情况中的哪一种，因此无法逆推出房间中最后一人的薪资信息。同时，其他人仍然可以定向确定自身薪资高低。

如果不信任中介能够对个人信息保密，那么他们也可在与中介共享之前即在个人输入数据中添加噪声。例如，每人都可在提供给中介的薪资数上增加或减少一定的数额（如2K），输出数据的大方向仍将是正确的，每个人既可确定自己的薪水高低，同时又能保护其输入的隐私信息。

隐私泄露实例探讨：

二十世纪九十年代中期，美国某州政府保险机构公开了经匿名处理的健康记录，以鼓励医疗保健领域的公共研究，其间使用了多种技术对数据进行匿名处理，例如删除地址、将姓名替换为随机字符串等。但是，研究人员仍然能够将该信息与可从公开渠道获取的选民登记数据进行比较和关联，进而重新识别数据库中的特定个体，[22] 甚至此前向公众保证患者隐私受保护的州官员亦被识别出。与其直接公开数据库，不如仅提供数据集查询功能，再应用差分隐私系统在反馈结果中添加噪声，从而防止患者个人信息泄露。例如，研究人员可以查询“邮政编码为ABCDE的人中有多少人患有糖尿病？”，差分隐私系统则会回复“邮政编码为ABCDE的人中有12,045人患有糖尿病”，这是围绕真实值的“模糊”响应。如果查询过于具体，例如“邮政编码为ABCDE的人中有多少人患有菲尔德病（一种极为罕见的疾病）？”，回复可能是只有一两个人患有这种疾病，这可能泄露私人信息。为了保护这部分人的隐私，差分隐私系统会添加噪声，返回诸如“邮政编码为ABCDE的人中有五人患Fields症”之类的回复，而这与现实情况大不相同。

金融服务业的应用：

该技术已足够成熟，可应用于金融机构；其益处显而易见，并且将其整合到现有数据系统中不会增加过多成本。添加噪声在数据精度和隐私保护之间作取舍，因此该技术最适合评估总体趋势，而不适用于异常检测（如欺诈分析）或精确的模式匹配（如光学字符识别）。目前Immuta等多家公司已开发出差分隐私解决方案，为金融机构提供服务。

第二项技术：联合分析

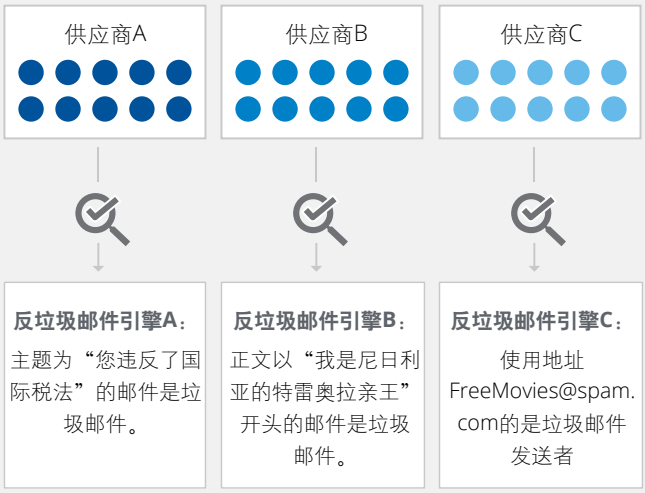
概述：

如某家机构想要分析跨多个数据库或设备保存的大量数据，可将所有数据整合入一个数据库中，再对整个信息集合进行分析，但这就出现了三个问题：1）在某些情况下，机构可能无权将本地存储的数据向外传输（例如，基于不同司法管辖区就隐私或其他本地化的限制）；2）数据可能为敏感信息（如就医记录、私人交易信息）且数据主体（即客户）可能不愿意共享此类信息；3）数据集中化风险：如果中央数据库被第三方恶意攻击，就会集中泄露大量敏感信息。所以，机构和数据主体可能都不愿意以这种方式共享数据。对此，解决方案是分析不同的数据集，然后共享分析所得洞察。[23]

近年来，联合分析作为此类问题的解决方案，已被谷歌等大型科技公司广泛用于学习个人计算设备（如手机和笔记本电脑）中的用户输入数据。[24]该领域的研究仍在继续，联合分析模型已与人工智能等其他新兴技术结合使用：2019年3月，TensorFlow（一个广泛使用的机器学习开源库）发布了名为TensorFlow Federated的开源框架，[25]该框架支持基于联合数据集的机器学习。

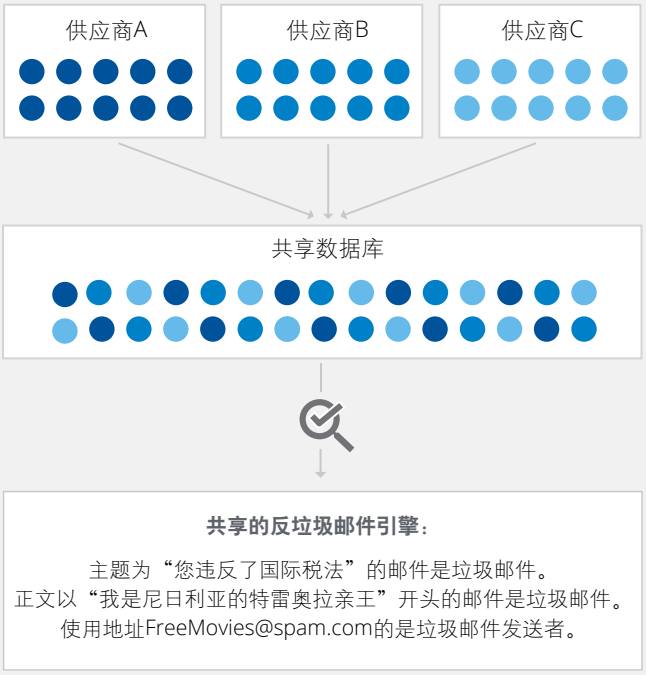
原理揭示：

假设三家电子邮件提供商想要帮助用户减少收到的垃圾邮件数量，一种方式是基于各自数据集分别对被报告为垃圾邮件的电子邮件进行分析，开发各自的垃圾邮件过滤器。



在这种情况下，三家机构可能会重复工作，因为垃圾邮件发送者的特征可能在三个用户群中均有出现。此外，分析或输入数据集的任何差异都可能导致各自的垃圾邮件检测引擎出现漏洞。

为解决漏洞问题，这些机构可将其报告的垃圾邮件数据合并到中央数据库中，而后创建一个共享的垃圾邮件检测引擎。

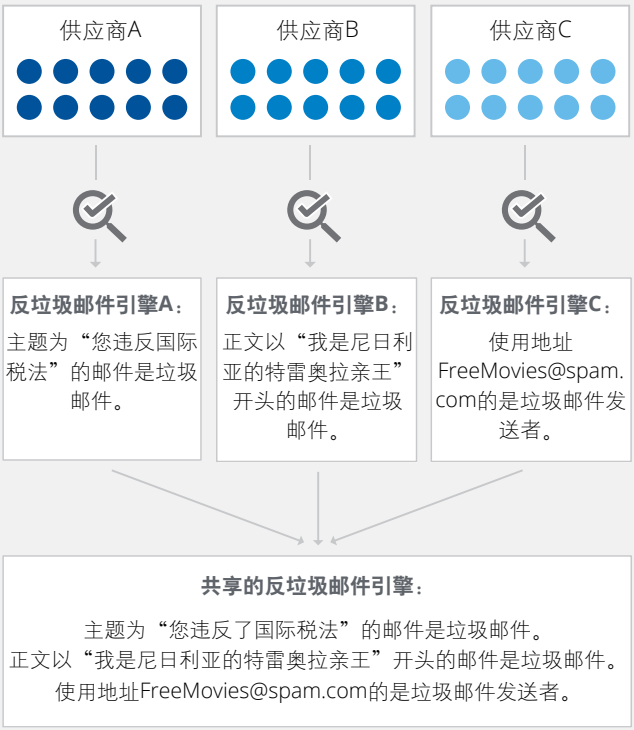


集合三家机构信息的数据规模使得这一引擎更具优势，所有用户都可从中受益。但是，该解决方案的问题在于：1）各电子邮件提供商的用户可能都不希望与第三方共享自己的邮件（即使已经声明共享目的是为用户自身利益而改进垃圾邮件过滤器）；2）各机构均面临竞争性信息（如用户相关信息）被公开的风险；3）该共享数据库还有可能成为恶意第三方集中攻击的目标——破坏一个数据库就可访问三家电子邮件提供商所有用户的敏感信息。尽管这种共享数据的方法达到了改进反垃圾邮件引擎的预期目标，但同时也带来了巨大的风险。

而使用联合分析就可在不产生新风险的情况下达到同样的目的。三家机构可共享其垃圾邮件检测模型并创建汇总模型，而不用共享底层原始数据。

通过这种方法仍然可以建立强大的反垃圾邮件引擎，同时降低共享底层数据所带来的风险。各机构能够从更大规模的数据中获益，同时，因未与其他电子邮件提供商共享用户数据，从而不会违反在共享用户数据方面需遵循的限制性规定。从安全性角度来看，也不存在恶意第三方会集中攻击的目标。

值得一提的是，该模式下得到的模型与通过将初始训练数据合并到中央数据库得到的模型并不完全等效；多数情况下，通过联合机器学习训练的模型不如在集中化数据集上训练的模型。用例一中就有一个此类示例。



隐私泄露实例探讨：

2017年，安全研究人员能够通过名为ai.type[26]的Android应用程序访问3,100万名用户的个人信息，该应用程序是第三方键盘，允许用户自定义手机/平板电脑键盘并提供个性化的输入联想建议。ai.type程序收集了各种数据（如能够提供数据输入建议的联系信息或改进自动联想功能的击键历史），存储于一个中央数据库中；然后，该数据库会先删除其中的私人信息（如密码字段），再分析数据从而为用户提供自动输入填充功能。但是，研究人员能够在清理私人信息之前访问数据库，并能公开所有3,100万名用户的电子邮箱地址、密码和其他敏感信息。ai.type原本可以使用联合分析在每个用户的手机上创建本地预测模型，而不是将所有数据集中到一个数据库，然后再汇总3,100万名用户的模型而非数据本身，从而保护每个用户的输入历史记录。[27]随后再将汇总模型通过更新推送回到每个手机中，并不断重复学习过程，这样就可以使键盘基于汇总模型对全体用户的学习分析提供高级建议。这也是谷歌和苹果在Android和iOS的默认键盘上采用的方法。[28]

金融服务业的应用：

尽管此项技术已被熟知并且已足够成熟，但目前在金融服务中的应用仍然有限。当存在大量单独的数据源（如手机、物联网（IoT）设备、笔记本电脑等）时，联合分析能够发挥最大的价值。在数十万个单独数据源中存储敏感信息，如此规模在金融服务实践中是罕见的。金融机构会集中存储交易信息和客户信息等数据，且多数地区均由排名前十的机构占据了绝大部分市场份额。但无论如何，联合分析作为一种技术上成熟的方法，是能够为金融服务行业带来益处的。第三章中探讨了一个相关用例。

第三项技术：同态加密

概述：

某些情况下需由第三方进行数据分析，原因包括：

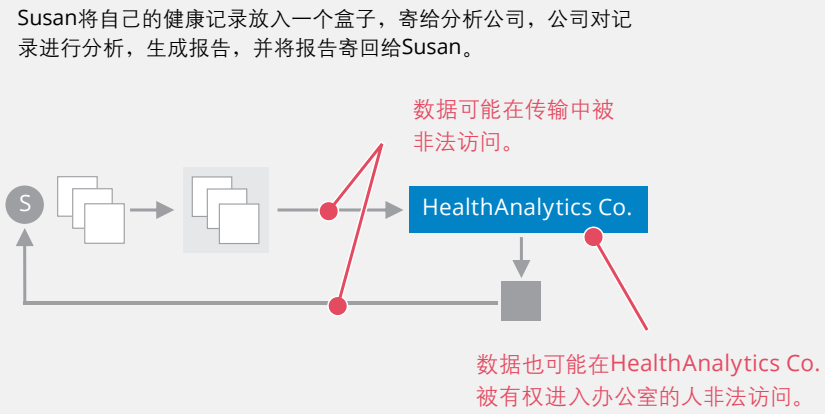
- 第三方具备数据管理方所没有的能力，并且愿意在不共享其所用底层函数的情况下提供分析服务；
- 第三方能访问数据管理方无法访问的其他辅助性数据，从而能提供更好的分析和洞察，这是数据管理方独自分析所不可及的。

但与联合分析一样：1）数据管理方可能无权转移数据；2）如果数据管理方并不信任第三方，或者担心第三方或其合作伙伴的内部人员滥用数据，则数据管理方不愿进行数据共享；3）如果第三方发生信息泄漏，原数据管理方可能因先前与第三方共享了数据而被客户追究责任。为解决这些问题，可使用同态加密（HE）技术对数据加密，以便在进行数据分析时维持信息不可读。并且，分析结果对除原定接收方（通常为输入数据的所有者）以外的任何人也都不可读。

“RSA”加密系统是首批广泛用于传输数据的加密方案之一，在1977年问世后，同态加密理论也在1978年被首次提出。[29] 在RSA加密系统下，（公开）密钥被用于加密数据并维持数据不可读；而后，该数据可被传输至原定接收者，接收者再使用另一种（私人）密钥对其进行解密；1978年，有人提出：能否使用加密数据执行不同类型的函数（如加法、乘法），而无须首先解密数据并因此暴露敏感信息；此后三十多年间，开发的各类解决方案已能使用加密数据执行特定函数，但仍未开发出可执行任何转换的全同态加密系统；2009年，Craig Gentry开发出了首个全同态加密（FHE）系统[30]，自2010年以来，全同态加密系统的效率和可行性得到了显著提升。

原理揭示：

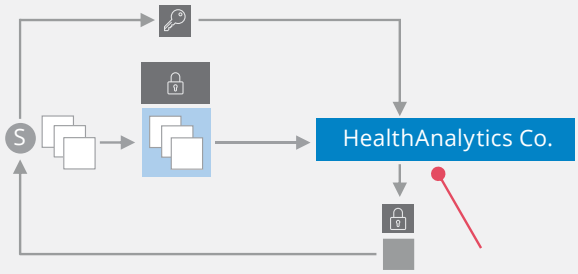
假定Susan希望深入分析自己的健康记录，以识别并预测潜在的健康风险，但她自身并无能力做这样的分析，需借助第三方——该领域的领军企业HealthAnalytics Co.进行分析。为了与HealthAnalytics Co.共享数据，Susan可将自己所有的健康记录放入一个盒子中，再将盒子寄给分析公司，但这会产生一些风险：盒子可能在传输中或在HealthAnalytics Co.的办公室被未经授权的第三方拦截；此外，HealthAnalytics Co.雇员也可能非法使用这些资料。



避免冒险，Susan可使用加密技术保护自己的信息。为此，Susan会将所有健康记录放入保险箱，并在不附密钥的情况下将其寄给HealthAnalytics Co.，再通过其他渠道将密钥单独发送给该公司。这消除了非原定接收方获取保险箱内文件的风险：即使有恶意第三方企图在传输间或在HealthAnalytics Co.办公室获取保险箱内文件，因没有密钥，也无法得逞。恶意方必须同时攻破HealthAnalytics Co.的数据库和Susan用于共享其数据访问密钥的传输渠道。

通过提升恶意方的执行难度，安全风险得以降低。但该公司得到密钥后会否将这些文件用于非预期目的或复制这些文件就不再是Susan能够掌控的了。因此，这种形式的“加密”也并非完全可靠。

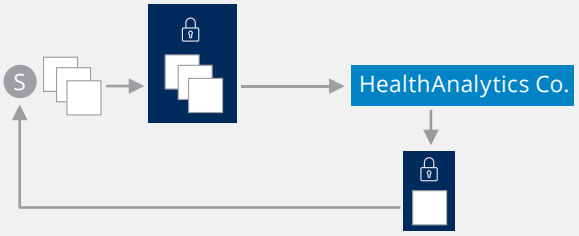
Susan将自己的健康记录放入上锁的保险箱，寄给分析公司，并向该公司单独发送密钥以便其分析保险箱内数据。该分析报告被放入另一个上锁的保险箱，并寄回给Susan。



数据可能在HealthAnalytics Co.被该公司内部能同时接触保险箱和密钥的员工非法访问，或在分析期间（数据被从保险柜中取出后）被公司外部的恶意方非法访问。

为全方位保护自己的数据，Susan可借助同态加密——一种特殊的保险箱。她将自己的健康记录锁入这个特殊的保险箱，在不附密钥的情况下将其寄给HealthAnalytics Co.。如果第三方试图在传输过程中或在HealthAnalytics Co.的办公室获取保险箱内文件，因没有密钥，将无法打开保险箱。不同于之前的情况，HealthAnalytics Co.无须打开这种特殊的保险箱，就可完成所需分析。对这个特殊保险箱进行的分析将其转换为另一个包含分析结果的特殊保险箱，同样也只能使用仍由Susan持有的密钥才能解锁。HealthAnalytics Co.随后将此保险箱寄回给Susan，Susan再用自己的密钥解锁保险箱，并阅读该公司对其健康记录的分析。该公司本身无法读取健康记录，甚至无法读取其数据分析结果，因为记录和分析结果都受特殊保险箱的保护。在整个传输/存储过程中，信息也受Susan所持同一密钥的保护。

Susan将自己的健康记录放入一个同态加密的保险箱，寄给分析公司。该公司像分析其内的健康记录一样分析此保险箱，并生成另一个也只能由Susan解锁的保险箱。此保险箱被寄回给Susan，Susan再用密钥将其转换为分析报告。



隐私泄露实例探讨：

2018年，Cambridge Analytica被卷入了数据泄露事件，此前该公司已收集了超过5,000万脸谱（Facebook）公司用户的数据。[31] 该公司向一个性格测试应用程序公司购买数据，此应用程序通过收集用户的姓名、电子邮箱地址、个人资料照片、社交网络、喜好和其他信息，再向用户返回高水平的个性特征图解。此应用程序存储了其“抓取”的数据，然后与第三方Cambridge Analytica共享，而Cambridge Analytica利用这些数据建立了详细的消费心态档案，以向目标受众投放数字广告。脸谱公司本可强制实施同态加密，或涉事的性格测试应用程序公司也可主动采用此法，以树立负责任的数据管理方形象。同态加密虽然可能不是防止数据滥用最高效或最直接的方法，却不失为一种可行的办法。使用同态加密，用户数据在被共享给第三方性格测试应用程序前将被加密。然后，该应用程序将分析此加密数据，并将个性特征图解返回给个人用户。应用程序本身无法读取用户的个性特征图解。用户可使用基于自己脸谱账户密码的私钥解密这些分析结果，而Cambridge Analytica或其他任何第三方均无法使用数据，甚至无法读取数据。

但一定要注意，以同态或其他方式加密数据并不意味着免除机构的隐私义务。经加密的数据本质上仍属于个人信息，需要有力的管理和监督以确保数据共享和使用方式合乎道德规范。

金融服务业的应用：

以当前的技术成熟度，同态加密未能实现大规模应用有两个关键原因：技术的局限性及公认标准缺失。

许多同态加密方案仅能执行一种类型的运算（例如，加法或乘法，而非两者皆可），且分析全同态加密（可能采取任何一种加密算法）数据要比分析未加密数据慢几个数量级。因此，该技术仅能用于采用特定函数的场景（同态加密案例），或者计算速度和计算成本并非优先考虑因素的场景（全同态加密案例）。但这些技术近期的快速发展和改进（几秒到几分钟）使应用同态加密来保护高度敏感信息成为可能。目前，该领域的开发活动仍十分活跃，诸如Ziroh Labs和Inpher等初创公司已经开发了在实际用例中具备计算可行性的同态加密或全同态加密方案。

传统加密系统存在公认标准，具备高度的互操作性，因而应用广泛。由于目前尚未建立起针对同态加密或全同态加密方案的公认标准，各类同态加密方案的可用性均大为受限。目前推进的一些举措（如同态加密标准化）正力图为此技术建立共同标准。

第四项技术：零知识证明

概述：

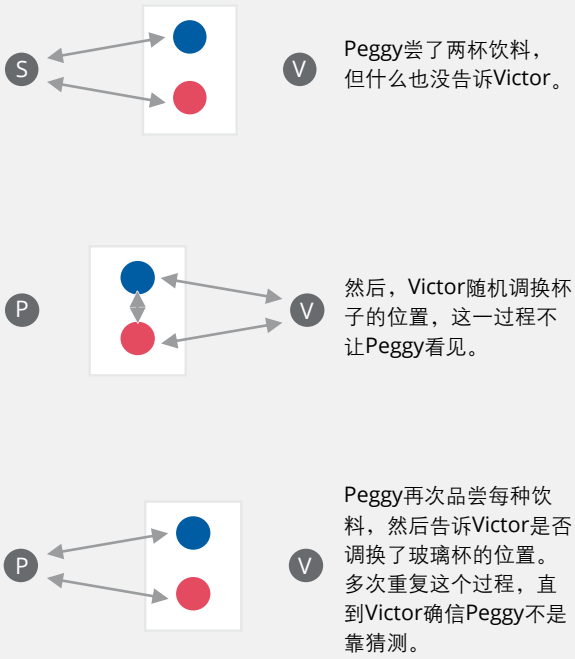
一些情况下，用户期望共享特定信息，同时又不泄露其他任何数据。当用户对另一方会否将此信息用于非预期用途存疑时，采用这一方式共享信息就非常重要。例如，在填写租赁申请表时，某人要证明自己的收入超过了房东的最低要求，但同时又不愿让对方得知自己确切的收入信息——如果他的收入远高于最低要求，则房东有可能很快寻机提高租金。在此情况下，接收收入证明的第三方可利用其收到的额外信息（确切的薪资）得出申请人希望保密的其他信息。零知识证明（ZKP）使一方能向另一方证明某些特定信息，而无须共享除预期信息以外的任何信息。

1985年，Shafi Goldwasser（麻省理工学院）、Silvio Micali（麻省理工学院）和Charles Rackoff（多伦多大学）在论文《交互式证明系统的知识复杂性》（The Knowledge Complexity of Interactive Proof-Systems）中首次提出了零知识证明。[32] 此后，零知识证明不断发展，涵盖了广泛的用例，包括证据不可区分证明、非交互式证明和可抵抗量子攻击的证明等。与联合分析一样，该技术也结合其他新兴技术使用——最著名的是结合分布式账本使用，从而在完全隐私的情况下跨P2P系统转移资产。

原理揭示：

假定Peggy想向Victor证明她能分辨出装在两个完全相同的玻璃杯中的两种苏打水的区别。Peggy还有另外两个要求：她不想让Victor知道区分两种苏打水的方法（例如，通过甜度不同来区分），也不想让Victor知道每杯苏打水的品牌。如果Peggy能满足上述要求，她就完成了“零知识证明”——既证明了自己能分辨两种饮料的区别，又未暴露与自己或玻璃杯内饮料相关的其他任何信息。

为此，Peggy应品尝每个玻璃杯里的饮料，然后背对桌子；接着，Victor应随机调换玻璃杯位置或者保持它们原本的位置（概率各约50%），然后让Peggy再次品尝每个玻璃杯里的饮料；Peggy应回答指出是否调换了玻璃杯的位置，但不透露每个玻璃杯内苏打水的品牌，亦不解释她是如何得知玻璃杯是否换过位置。第一次进行测试时，Peggy仅凭猜测就有50%的正确几率。但是，若她能分辨两种苏打水的区别，则应在重复测试过程时能够始终回答正确，且她靠猜测得到正确答案的几率应大大降低。



到第20次测试时，Peggy猜对的几率大约为百万分之一，因此Victor有理由确定Peggy的确知道这两种苏打水的区别。这就是零知识证明，因为Victor既不知道每杯苏打水的区别，也不知道Peggy是如何分辨出两杯苏打水之间的区别的。

隐私泄露实例探讨：

2019年1月，美国某大型零售商的一名员工被捕，因其涉嫌将客户的信用卡号透露给同伙，然后由同伙使用窃取的信用卡信息进行购物。[33] 该员工会在客户购物时记住他们的卡号并抄录下来，随后将号码通过短信发送给同伙。在2018年到2023年期间，零售商遭遇的无卡交易欺诈涉案总值预计将达1,300亿美元，其中与上述案件类似的信用卡盗刷不容忽视。[34] 而今我们可设想采用零知识证明支付系统来避免此类损失：零知识证明支付系统允许个人在零售商处验证其银行信息和余额，而无须向任何第三方（如收银员）透露账户信息和信用卡验证值（CVV）代码。

金融服务业的应用：

随着技术方法不断成熟，零知识证明近期终于投入了实际应用，例如，支付（如Zcash[35]）、互联网基础设施（如NuCypher[36]）和数字身份（如Nuggets[37]）等用例。荷兰国际集团（ING）等大型机构已投入资金在金融服务领域推广零知识证明技术的运用，[38] 并且该技术还有望在广泛推动分布式账本技术发展中发挥关键作用（因为它使个人和机构能够保护公共分布式账本上的私人信息）。

第五项技术：安全多方计算

概述：

与同态加密和零知识证明一样，此项技术可在共享信息给不可信的第三方时维护个人隐私。安全多方计算（SMC）可确保公司在集中分析多家机构所持私人信息时不泄露输入数据。在过去，这需要一个中介充当数据共享的中间人，但会产生一些问题：

- 该中介的内部人员可能滥用数据（例如，将数据出售给另一方用于非预期的用途）。协作（例如，银行共享交易数据以识别付款欺诈）涉及的第三方/中介甚至可能是公司的竞争对手，这增加了竞争机密泄露的风险。
- 如果中介受到外部恶意攻击，则公司的敏感信息将会泄露，且尽管安全漏洞不直接归咎于公司，公司却仍可能被消费者和监管机构追究责任。

使用安全多方计算技术后，上述中介将被一种可靠的算法所替代，在此情形下，即便遭遇攻击也不会泄露任何敏感信息。安全多方计算从根本上依靠“秘密共享”：[39]即每个参与者的敏感信息均以加密“份额”的形式分存于所有其他参与者处；这些份额即便被第三方恶意拦截或被某个参与者滥用，也毫无用处，因为被拦截的部分只有与其他各方分存的信息结合起来才能解密。

二十世纪七十年代后期，计算机技术在世界各地越来越多的用于办公领域，也逐渐进入万千家庭，也正是此时安全多方计算解决方案被首次提出，用于在不具备可信第三方的环境中建立可信的系统（例如，当对游戏网站是否会在幕后操纵系统存疑时，该如何玩在线扑克？）[40]。随着时间的推移，新的方案不断推出以适用于更广泛的用例。2008年，安全多方计算首次投入实际应用，用于在不泄露农民个人经济状况的条件下确定丹麦甜菜的市场价格。[41]自2010年以来，研究重点转向了提高安全多方计算协议的运行效率/可扩展性上。

原理揭示：

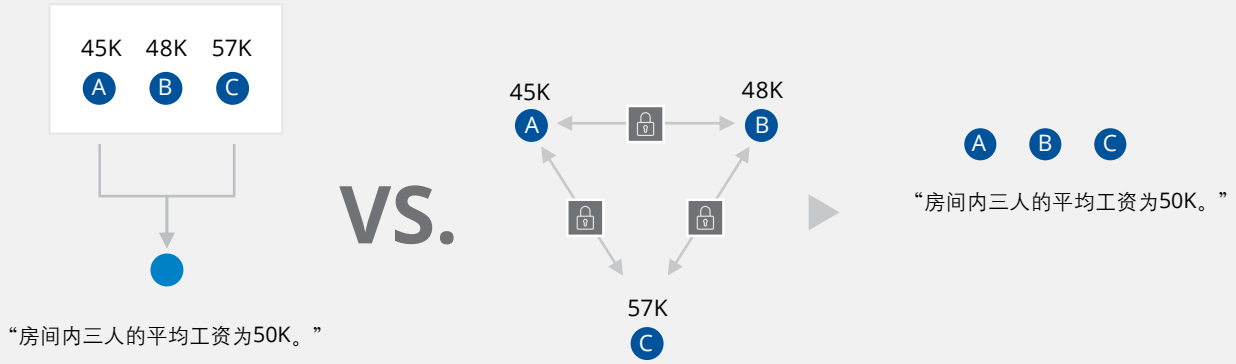
安全多方计算背后的特定逻辑极其复杂，可能比本文概述的其他技术都复杂。为确保读者在没有大量技术专业知识的情况下理解基本过程，我们提供了两种不同的描述：

1. 简要概述：是指对本技术及其优势的简短概述。
2. 详细说明：详细的案例分析，逐步分析第一项技术（差分隐私）中探讨的假设示例，且每个步骤均有实例计算。

另外，您也可点击此处观看波士顿大学讲解安全多方计算的实用视频。

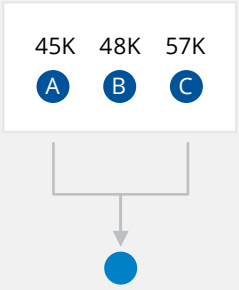
1. 简要概述：

安全多方计算的本质：加密信息在多方间共享，按照所需分析和计算进行配置，各方并不共享敏感信息，但仍能得到正确的最终结果。安全多方计算系统按照各方负责计算的部分进行配置，所以不需要可信的中介。



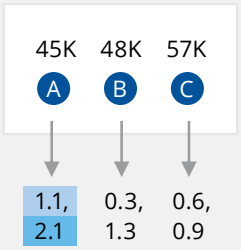
2. 详细说明：

让我们回到第一项技术中探讨的假设案例，并将房间内人数由10人简化为3人。原案例中存在可信的中介，了解房间内人们平均工资的过程相对简单。

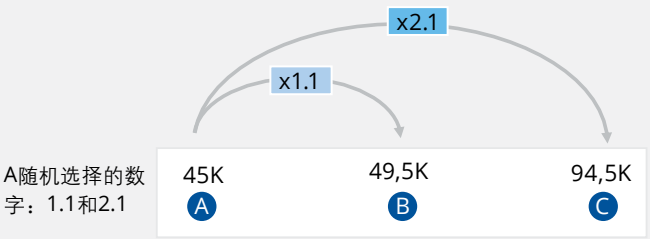


“房间内三人的平均工资为50K。”

该假设案例基于一个前提，即该中介是廉洁可信的，而中介并不一定始终满足此前提。中介可能与房间内的某个人（或第三方）串通，然后共享私人信息；或中介记录泄露，导致第三方未经许可访问私人信息。安全多方计算可用于降低这些风险——在不使用中介的情况下，采取算法来实现相同目的。首先，各方随机选择0到3之间的两个数字（上限3是数据共享协作参与者的数量）。



然后，每个参与者将其薪资数分别乘以上述随机选择的数字，并将相乘后失真的数字告诉另外两个参与者。让我们看一下参与者A的计算过程：



此时，B和C所知A的薪资数差异巨大，即便两者串通，也无法逆推出原始数字。参与者B和C也执行相同的操作，用自己选择的随机数字来改变薪资数，并共享给其他参与者。这产生了失真薪资信息矩阵。

	A	B	C
A告知...		49.5K	94.5K
B告知...	14.4K		62.4K
C告知...	34.2K	51.3K	

为得出三人的总薪资（除以三可算出三人的平均薪资），每个参与者用本人的实际薪资数，加上其他参与者告知的数字（在本人姓名所在列），再减去告知其他参与者的数字（在本人姓名所在行）。让我们梳理一下A的计算过程：

	A	B	C
A告知...		49.5K	94.5K
B告知...	14.4K		62.4K
C告知...	34.2K	51.3K	

参与者A用自己的实际薪资数加上蓝色高亮部分的数字再减去红色高亮部分的数字。然后，A将得出的结果共享给其他参与者。

因此，A的结果为= 45+14.4+34.2–49.5–94.5

= **-50.4K**

参与者B用他/她自己的薪资数执行相同的计算过程：

	A	B	C
A告知...		49.5K	94.5K
B告知...	14.4K		62.4K
C告知...	34.2K	51.3K	

B的结果为= 48+49.5+51.3–14.4–62.4

= **72K**

参与者C也如此执行：

	A	B	C
A告知...		49.5K	94.5K
B告知...	14.4K		62.4K
C告知...	34.2K	51.3K	

C的结果为= 57+94.5+62.4–34.2–51.3

= **128.4K**

将上述三个结果相加就得到房间内三人的总薪资，只须除以三便得平均薪资：

-50.4+72+128.4=150K
150K/3=50K

重要的是，整个过程的任何环节均未泄露任何参与者的实际薪资数：任何中间步骤均未出现45K、48K和57K。想要根据参与者在任一中间环节提供的输入数据逆推得出这些数字也不可能，因为输入数据经随机修饰因子（0到3之间的任意数）处理均已失真。

但由于各方都从分析中获取了真实而准确的输出数据，因此一方仍可能将输出数据与其他信息相对照，推断出一些敏感信息（如第一项技术的案例所示，其中一人可根据平均薪资和所涉及的其他参与者的已知薪资，逆推某一人的薪资信息）。可将差分隐私技术应用于安全多方计算的输出数据，这样在分析数据和共享分析结果时都能保护隐私。在下节的用例中，我们将探讨如何将不同技术结合起来并实际应用到金融服务。

隐私泄露实例探讨：

2009年2月10日，美国的铱星33通信卫星与俄罗斯的宇宙2251卫星相撞，两者当即损毁。[42] 如果美俄共享各自卫星的位置信息，本可以探测到即将发生的撞击并采取措施防止其发生，但为了保护国家安全及双方公民与军方的隐私，卫星的轨道数据受到严密保护。安全多方计算协议可用于仅共享关键信息（即“美俄各自的卫星是否即将相撞？”），而不共享底层的位置数据。

金融服务业的应用：

安全多方计算属相对较新的技术，因此在金融服务业（及更广泛的领域）中的应用有限。部分原因在于安全多方计算需要针对每个用例进行完全定制化的设置，带来极高的设置成本（与之不同的是，差分隐私可在各用例中使用通用的算法）。但目前正在开发可将底层协议分离出来的“编译程序”以实现通用计算，从而使安全多方计算能更广泛地运用于数据科学和机器学习应用程序。

目前的安全多方计算系统通信成本高，持续运行的费用高昂。但该技术还在不断发展，Inpher（由摩根大通投资）等金融科技公司已开发了针对金融服务业的安全多方计算产品和服务。假以时日，这项技术在金融服务行业的应用或将持续增加。

第三章：金融服务业应用

上述技术各有侧重，在整个金融行业中具有众多的潜在应用场景。将各项技术互相结合，用一项技术的优势弥补另一项技术的局限性，可更好地实现隐私性、安全性和实用性。本章中，我们将探讨如何采用隐私增强技术解决隐私实践中的矛盾问题，以及如何通过新型的数据共享方式创造新价值。

为金融机构解锁新价值

用例一：侦测车险欺诈
使用联合分析、差分隐私和零知识证明

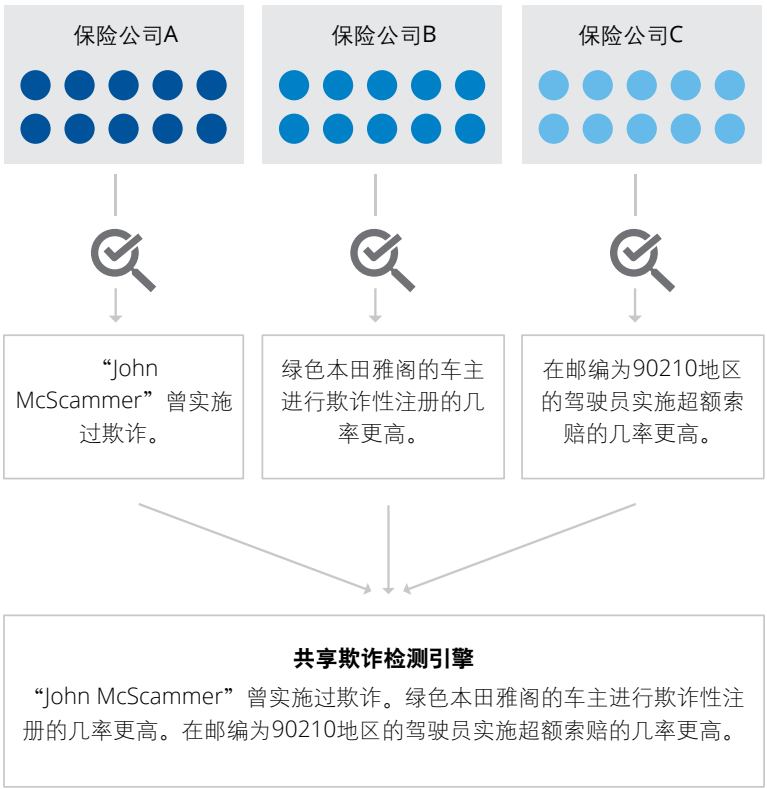
背景：
在美国，每年保险业中非医疗保险诈骗造成的总损失估计超过400亿美元，平均每个家庭每年为此多支出400至700美元保费。[43] 特别是在车险领域，消费者和金融机构都承担了诈骗损失：消费者支付了高于实际风险所需的保费；金融机构则赔付了欺诈性索赔，导致赔付率升高，从而最终损害了其盈利能力。

数据共享：
保险公司有望通过数据共享来减少诈骗，数据范围包括注册、索赔、车联网、投保车辆和客户数据，以及其他非结构化数据，如医疗报告等。这对保险公司有两大益处：

- 共享数据规模扩大，有助于改善预测和分析。例如，索赔数据、车联网数据和其他非结构化数据规模的扩大将有助于保险公司更好地识别欺诈性索赔的模式。
- 识别针对同一车辆或事故向多家保险公司提出的重复索赔。

因牵涉众多敏感信息，存在重大隐私问题。消费者不愿与第三方共享自己的注册数据、索赔数据、个人信息等私人信息，以及医疗报告等其他数据。保险公司本身也对与竞争对手共享此类信息持谨慎态度，因为这些信息可能被滥用于推断自身的承保和定价策略，以及其他敏感的竞争情报。

隐私增强技术应用设想：
联合分析可用于创建涵盖注册和索赔的诈骗侦测/预防主模型，而无须各保险公司共享底层客户数据。这使保险公司受益于扩大的数据规模，同时又保障了客户的隐私安全和业务运营的机密性。



需要注意的是，联合分析可能无法跨保险公司识别所有重复性信息。例如，假设John McScammer投保了保险公司A的车辆责任保险，以及保险公司B的健康保险。后来，McScammer先生出了事故，并向保险公司A和保险公司B提出了索赔，且均获得了赔偿。联合分析会遗漏此类“领双份赔偿金”的情形，因为其在每家保险公司单独数据集的记录和分析中是合理正当的。

为解决联合分析的这一弱点，可结合其他隐私增强技术，具体技术类型视数据共享协作的确切架构而定。保险公司1）可将各自的数据集并入一个同态加密的中央数据库进行分析，就能够识别类似上述McScammer先生案例的重复索赔；2）也可直接通过该中央数据库进行查询/分析，结果与采用联合分析模型所得结论相同，同时采用差分隐私技术确保不会在分析时泄露客户隐私。在索赔过程中，零知识证明也可用于查询每家保险公司单独的数据集。例如，当McScammer先生向保险公司A提出索赔时，保险公司A将查询保险公司B和C的数据集，以确认McScammer先生是否已提出过索赔，或近期是否针对同一投保车辆提出过索赔。通过使用零知识证明技术，这些查询将能够被检测匹配，且不会使保险公司B或C知道被查询的客户或车辆信息，从而防止了敏感信息的泄露（例如，保险公司A的客户名字被保险公司B或C获知）。

用例二：成为可信的数据守卫者
使用零知识证明

背景：
过去，只要具备现代化系统、高级分析法和专有数据集，科技公司就能妥善保护客户敏感信息（如电子邮箱地址）和身份信息（如社交平台登录信息）。但在大量数据泄露丑闻被曝光，众多公司因违反《通用数据保护条例》（GDPR）等法规被罚款后，过往做法的局限性逐渐显露，换言之许多科技公司核心业务的盈利能力就是建立在其所能提供的数据之上。

数据共享：
不同于科技公司和其他行业的许多公司，金融机构历来都不靠数据获取收入。金融机构在所持数据的安全性方面也受到严格的监管，并在另一种重要资产——资金保管方面，历经数十年甚至数百年建立起了值得信赖的品牌。因此，金融机构被广泛认定为新一代数据管理的驱动者，将建立起基于信任和监管义务的数字服务新模式。这为金融机构带来了机遇，通过更频繁的互动增强客户粘性；同时在相关产品和服务方面，为金融机构扩展业务创造了机会。

对隐私增强技术应用的设想：
为了证明金融机构是值得信赖的数据保护者，我们再次回顾零知识证明概述中简单介绍的用例：某人想要在不透露自己确切收入的情况下，向房东证明自己的收入达到了最低要求（或在某些情况下，向雇主提供证明）。作为个人薪资直接存款的接收方，零售银行已掌握该信息，且被认为可信（即通常值得房东信赖）。

要满足房东要求不必使用隐私增强技术，一份经公证的银行证明足矣，这也是目前住房租赁市场主要采用的方式。然而，使用零知识证明有以下两个优势：

- 客户能够自行准备必要文件，而不必依赖理财顾问或银行客服代表协助取得公证文件，这有助于以更低的成本实现更高效的服务。
- 证明更可信。通过零知识证明（ZKP）系统，房东能够更直接地核验，而公证文件也有可能是伪造或经篡改的。

对于任何金融机构而言，仅为收入验证目的而作重大技术投资创建ZKP系统是不合算的。但同一个系统还可轻松用于更多客户特征的验证，包括交易数据等金融数据，以及年龄、地址等非金融数据。由于银行按规定需保持客户最新的信息，其确实极有可能握有大量最新的个人身份验证信息（如护照或驾照信息），甚至比传统的身份验证机构掌握的此类信息还多，因为政府身份验证服务部门往往仅在数年一次的护照/驾照更新时才会收到相关更新信息。

作为值得信赖的客户数据管理者，金融机构通过协作将能够解锁更多价值：各机构都持有几类数据（如借记卡和信用卡交易）而非所有数据（如贷款余额和投资余额信息就可能没有）。对此，多个数据管理方可建立协作网络，根据个案与客户具体情况将第三方请求分配至适当的金融机构。

这或将为客户带来更多价值，包括但不限于：

- 在不告知具体年龄的情况下进行年龄验证（例如，租车不缴纳低龄风险驾驶金）。
- 在获取政府服务（如纳税评估报告）以及金融服务（如信用评分免费查询）时进行简单快捷的验证。
- 证明自身信用评分在贷款机构评审系统规定的范围内，而无须共享确切的分值。

用例三：效仿开放式银行且无需监管
使用安全多方计算

背景：
全球多个司法管辖区包括英国、欧盟、日本、澳大利亚、中国香港、以及加拿大均推行开放式银行监管。然而，世界其他地区尚未建立相关法规，并且特定的环境导致难以实施自上而下的有效监管。例如，在美国，银行同时受到州以及联邦的监管，并且美国有超过5,000家在联邦存款保险公司投保的机构，因而无论从监管角度还是机构角度来看，美国的银行格局都高度碎片化。[44]

开放银行规定了机构须共享的数据及数据共享方式，这在碎片化的格局下是难以实现的。许多机构认为不进行数据共享为妙，因为他们认为，与第三方共享数据会加剧对自身不利的竞争，且数据共享会耗费大量技术成本。

数据共享：
即使不共享数据，第三方仍可通过“屏幕抓取”服务获取上述数据。这种服务通过请求客户提供网上银行用户名和密码，并利用自动登陆工具定期进入账户抓取客户交易信息。此类服务将带来巨大的安全风险（客户被要求共享其安全证书）以及占用大量带宽（自动登陆系统需持续加载银行网页，加载的数据量远超过所需抓取的数字和文本）。

一些机构已经意识到数据共享能够成为一种提升竞争力的有力武器：为客户提供增值产品和服务。由机构自身而非法规所搭建起的数据共享生态系统有助于在同第三方签署数据共享协议时扩大共享范围和提升协议条款的灵活度，且便于加强对数据共享系统的管控。例如，西班牙对外银行 (BBVA) 建立了“银行即服务”平台，第三方可通过该平台验证客户身份、转移资金、甚至通过代码找到账户来源。[45] 这使BBVA从竞争者变成了金融技术的提供商，并最终拿到机构客户的存款资金。

对隐私增强技术应用的设想：
金融机构 (假设前述金融技术能够得到应用) 可使用安全多方计算系统以确保共享给第三方的数据 (如客户交易信息) 仅用于预期目的。例如，某机构与一家金融科技公司合作，为小型企业提供现金流预测；金融科技公司通过与客户的开票软件 (如Xero) 以及银行实时连接，自动预测现金流详情并预测何时可能需使用过渡性贷款。

不仅客户能从银行与金融科技公司之间的合作中获益，银行的贷款业务亦可受益于此，银行须确保客户的数据不被滥用。基于一般的数据共享协议，或借助于屏幕抓取，金融科技企业能够获得用户交易数据的所有访问权限，并可能在银行和客户不知情的情况下，滥用该数据。

通过对安全多方计算系统进行限定，系统只能进行指定分析工作（例如，交易求和、识别经常性的资金流入和流出），这样金融科技公司就无法访问交易数据，进而降低了第三方滥用数据的风险并减少了因第三方遭遇安全事件而导致敏感信息泄露的可能性。让作为合作方的金融机构和使用金融科技公司服务的潜在客户能够对数据共享更为放心。[46]

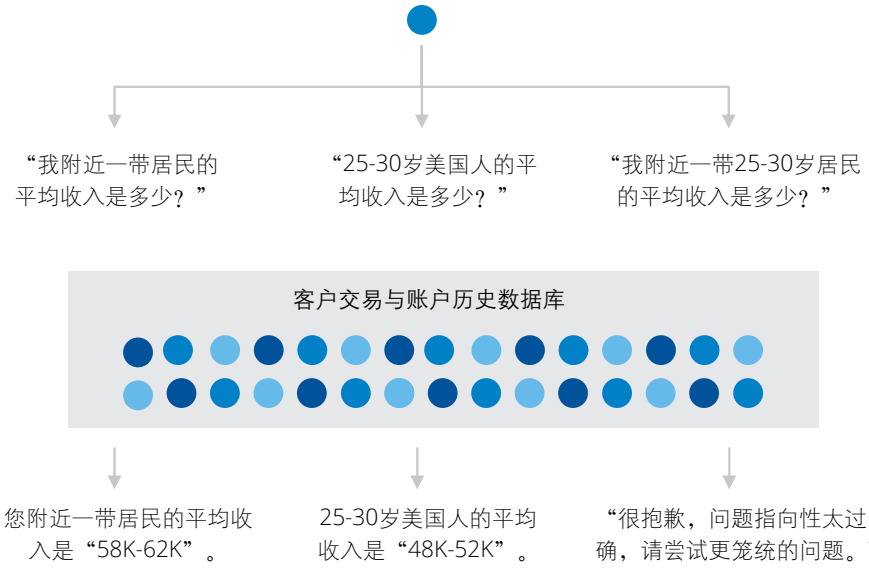
为客户解锁新价值

用例四：智能化、自动化的个人财务管家（PFM）
使用差分隐私

背景：
隐私增强技术既能用于增强机构自身的竞争力，亦能用于增进机构间的协作。随着数据获取渠道的增加 (通过开放式银行) 以及自动分析的复杂化，开发面向大众的个人财务管家 (PFM) 产品正受到各方追捧。鉴于仅30%的美国人有包括储蓄和投资在内的长期理财计划[47]，而近半数美国人对自身当前财务状况感到担忧和焦虑，所以该产品的潜在优势十分明显。[48]

数据共享：
通过对零售银行的整个数据库进行自动分析，可基于相似群体的特征，提供高级建议。第三方将基于开放式银行所提供的信息（在征得客户同意后）对客户群进行整体分析后，提供定制化的咨询建议。例如，某机构可解答客户“我在酒吧的花费比同年龄段的平均水平高或低多少呢？”的疑问，尽管能够帮助客户解决疑惑，然而就算以匿名形式共享，某些客户也会因自身消费习惯被他人所知而感到不适。若特定群体足够小，你就能通过“相似群体”比较功能了解使用同一PFM产品的特定群体中其他个体的消费习惯。

对隐私增强技术应用的设想：
对于开放式银行数据的接收方（如PFM自动化顾问），差分隐私是解锁跨机构数据集价值的关键工具。差分隐私能在洞察形成过程中添加噪声并确保数据集内的个人隐私不被侵犯，打破以隐私换取建议这一怪象，使个人在隐私受到保护前提下获得定制化的财务建议。

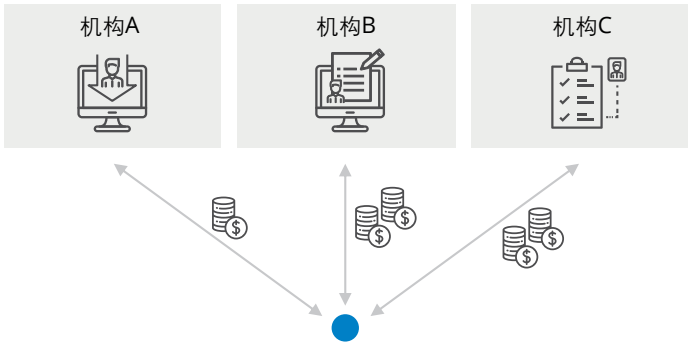


对于大型金融机构而言，跨区域的分析和比较使其能够利用某市场中可用的数据为另一市场提供高质量的服务和产品，这将为其带来巨大的价值。设想一下，一家大型美国银行，开发了利用差分隐私的PFM顾问产品，该银行欲在加拿大发展业务，鉴于加拿大客户与美国客户有许多共同的行为和偏好，该银行希望从一开始就能为加拿大客户提供同等质量的顾问服务。

根据隐私监管法规，该银行或不被允许跨境共享交易数据、账户余额、客户的人口统计信息，因此，该银行需先积累大量加拿大客户才能进行此前为美国客户进行的分析工作。然而，通过采用差分隐私技术，加拿大机构得以在不访问底层数据的情况下对美国客户进行分析并利用所得到的洞察。通过向其美国机构汲取经验，让其加拿大客户受益，并最终为银行解决冷启动相关问题：为尚未直接积累其大量信息的客户群提供高质量的咨询建议。

用例五：零售银行客户注册程序规范化
使用零知识证明

背景：
目前，零售银行独立管理“了解您的客户”（KYC）与“反洗钱”（AML）注册流程。虽然该流程包含多个步骤，但本小节旨在介绍其中的客户识别步骤（CIP）。CIP是美国金融机构按规定必须完成的步骤，即，当个人想要通过某金融机构的基础设施进行交易时，该金融机构需对其进行身份验证，需验证的信息至少包括注册时所填写的个人姓名、出生日期、有效身份证号码。对客户而言，在每家金融机构申请产品进行注册时都需反复提交相同的资料。而对于整个金融服务业，这也带来了大量的重复性工作。虽然CIP是美国的金融机构按规定所必须履行的义务，但与之相似的合规义务在世界其他国家的金融机构运营过程中也并不少见。



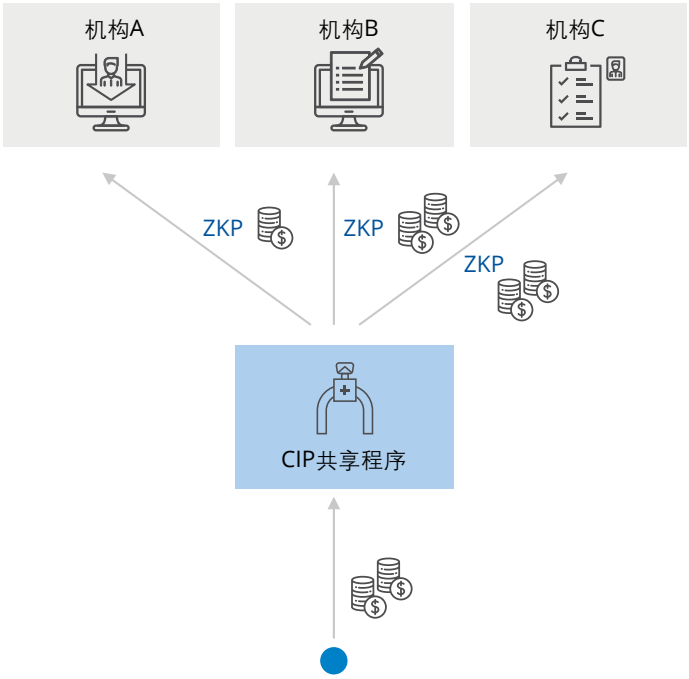
同一客户在不同机构注册时，需提供相同的资料，这造成了金融机构间不必要的摩擦与重复劳动。

上述步骤虽繁冗，却仍不够有效：在谷歌中搜索就能轻易查到著名诈骗犯Daniel Fernandes Rojo Filho的金融犯罪前科，但该诈骗犯仍于2014年年中在17家大型金融机构用这个名字成功开户[49]。

数据共享：
共享CIP注册流程能够使客户在不同金融机构注册时无须重复提交相同的信息，此举有助于客户体验更加快捷的规范化注册流程。KYC和AML是针对同一组数据进行的，所以两项流程都能更高效的开展。然而，客户隐私监管法规通常禁止机构间共享个人身份信息（在某些情况下，甚至企业各事业部间也不得共享）。此外，为防止竞争对手对自身客户开展定向广告营销，机构通常对客户数据库严格保密。

对隐私增强技术应用的设想：

零知识证明（ZKP）可用于解决此类问题并且让各机构减少重复劳动。通过共享程序获取必需文件，各银行无须各自开展CIP。用户仅须在注册该共享程序时一次性提供一般KYC流程所需全部资料（如家庭住址、身份证号），通过ZKP技术，当用户后续在各金融机构针对某产品进行注册时，该共享程序会同相应机构仅共享该产品所需的CIP数据。



通过该程序，金融机构无须储存个人身份识别数据副本，也避免生成重复信息（最终成为过时信息）。该共享程序还能为注册以外的流程带来好处：除了简化CIP，还能支持KYC/AML规定的持续监控——例如，家庭住址的变更只需上传至CIP共享程序中，当各金融机构要求提供或需要使用住址信息时，旧信息会立即通过CIP共享程序得到更新。

需注意，单一集中的共享程序并非理想的端对端数字身份解决方案。该示例仅用于说明如何通过ZKP解决CIP所产生的重复工作问题。如需了解更多金融服务中数字身份有关资讯以及其他可行的金融服务数字身份解决方案，请参见世界经济论坛于2016年发布的关于数字身份蓝图的报告。[50]

为监管机构解锁新价值

用例六：分析整个生态系统的金融风险敞口

使用安全多方计算或同态加密

背景：

系统性风险（会对整个金融系统和市场造成威胁）的防范是一件复杂的工作，并且很难事先预防。按规定，各机构对很多类型的风险（如信用风险、流动性风险）自行管理，但因掌握的信息不全面，无法了解整个金融生态系统所面临的风险状况；各机构自身的流程或尚属稳健，但单个机构与其他市场参与者的互动或对整个金融系统带来无法预料的后果。因所掌握的数据呈碎片化，无法纵观全局，单个金融机构或监管机构难以预测或主动监测此类风险。

数据共享：

对整个生态系统的数据进行前瞻性分析或能对金融系统中形成的系统性风险发出提前预警，例如，引发2008年全球经济衰退的系统性风险。举例而言，如果汇集美国共同基金全行业数据就有可能发现开放式基金因持雷曼债券所形成的集中风险。但由于此类数据属高度敏感信息，对其进行共享会使机构在战略上面临重大竞争威胁，所以至今未能将其纳入金融系统数据共享范围内。但同时，及时掌握该等信息对于预测威胁以确保金融生态系统的安全稳健又至关重要。

对隐私增强技术应用的设想：

Emmanuel A. Abbe、Amir E. Khandani和Andrew W. Lo在合著的论文《金融风险共担：隐私保护新方法》（Privacy-Preserving Methods for Sharing Financial Risk Exposures）[51]中详细介绍了：安全多方计算能够对机构所面临的风险进行综合分析，并可确保不侵犯个人隐私，也不将机构战略泄露给竞争对手。

运用与第五项技术中安全多方计算类似的技术，可在不暴露企业敏感专有数据——自身信贷组合的情况下，通过计算垂直行业（例如，房屋建筑业vs工业vs汽车业）贷款总额得出经济对利率变化的敏感度。

理论上，同态加密也能用于类似分析，但在进行更为复杂的分析时（如均值、方差），使用同态加密技术提供及时有用的洞察可能导致计算成本过高，从而限制了实际应用。随着同态加密技术渐趋成熟，今后或可直接替代安全多方计算系统。

结语

现今，金融机构之间的竞争主要集中于价格（以最低总成本提供产品和服务）和客户体验（推出独特的价值主张）。在科技变革的推动下，基于价格和客户体验的竞争推动了全球金融体系的发展。然而，注重隐私和安全作为支持行业发展的新动能，正成为优秀机构的重要特征。对于客户和监管机构而言，能够安全储存并管理数据的机构才值得信赖，而一些行业频频爆出丑闻，严重动摇了社会各界对相关行业的信任。

在客户和监管双重要求的推动下，数据掌控权愈发向客户倾斜，越来越多的金融业人士认为：金融机构将失去利用自身所持数据为客户、自身、乃至社会创造价值的能力。

了解隐私增强技术

正如本文中对各类用例的探讨所示，一整套新兴的“隐私增强技术”（PET）看起来具有创造价值的潜力，却由于各界对数据隐私的担忧而难以实现。差分隐私、联合分析、同态加密、零知识证明以及安全多方计算技术的结合运用能够实现上述用例以外的其他更多应用，包括：

- 共享对各机构交易数据分析所得的模式和洞察而无须共享交易数据本身以避免内幕交易。
- 通过算法代替中介，确保匿名性、透明性，杜绝行贿受贿，从而避免围标。
- 分析公司的购销发票以检测税收欺诈，同时保持交易数据的机密性。

技术实施所面临的挑战

隐私增强技术呈现巨大潜力并且发展迅速，但*值得注意的是：成功使用隐私增强技术要求机构除了解和部署技术本身以外，还需采取以下步骤。*

研发投资：此类技术大多处于初始阶段，由于近几年得到巨大发展，使其应用于金融服务成为可能；为使这些技术更便于业务应用，各机构仍需投入大量资金。迄今为止，隐私增强技术的发展主要由学术研究所推动，技术开发者较少考虑金融服务中的实施情况，企业用户亦少有思考如何将其为己所用。因此，很多此类技术系统难以在企业中实施，为解决理论与实践的差距，一些提供过渡服务的公司应运而生，助力金融机构更为轻松地利用隐私增强技术所带来的好处。从近几年的发展形势来看，无论是通过与此类第三方公司合作还是投资新的研发项目，各机构需为维持隐私增强技术创新不断投入。

与公共部门合作：因为技术处于初始阶段，我们不确定世界各地的隐私监管法规将如何规范隐私增强技术。例如，某些地区不允许数据跨境共享，理论上联合分析和安全多方计算应能帮助企业实现跨地区的数据分析。然而，为避免受处罚或遇到其它监管风险，确保此作法合规对企业至关重要，而在很多情况下，监管存在不确定性。为推动监管机构认可该技术，各界需加强对隐私增强技术的了解，公共部门与企业需就如何在金融领域安全使用隐私增强技术进行公开探讨。

增进客户对隐私增强技术的认知：这些技术大多很抽象，而以往产生的风险经历会导致客户认为这些技术不能保护其隐私、也并不安全。为获得客户信任并被客户接纳，金融机构在实施隐私增强技术时，须同时注意保护客户数据并帮助客户了解其数据已被妥善保护的这一事实。

克服其他阻碍：除与隐私增强技术直接相关的问题，金融机构还需克服其他阻碍以完全实现本文所探讨的机遇以及更多的目标。这些阻碍包括：

- 数据质量差：以往的数据集存在诸多问题（例如，人工输入错误、缺少细化信息，和/或难以清理与编排），不便于计算机处理。
- 技术滞后：老旧的核心系统不能支持持续数据共享和分析所需的数据存取方式（例如：实时、通过应用程序接口）。
- 数据结构碎片化：企业的数据分散于不同数据库中，难以整合形成洞察。
- 缺乏数据互操作性：因数据格式不同，机构间数据共享往往无法实现深层次的运用或造成数据质量打折。
- 地区差异：各司法管辖区因政策差异而在数据的使用和管理方面存在不同的要求，进一步增加了跨国企业使用这些技术的难度。

各种技术的出现，也使很多上述问题得以解决。例如，各类银行核心系统的现代化解决方案提供商推出模块化且灵活的系统以替代金融机构的老旧系统，便于机构将隐私增强技术等新功能顺利整合入系统。

总结

尽管充满挑战，隐私增强技术潜力巨大并且发展迅速。金融机构目前无法完全看清整个行业共同面临的最主要并且最迫切的问题。隐私增强技术可促进机构间的全面沟通共享，且不会影响各机构为保持优势所依赖的竞争性信息的保密性，也不会辜负客户对其数据守卫者这一角色的期望。

附录

技术优势和限制

第一项技术：差分隐私

优势：

- **允许在隐私保护和数据准确性间取得平衡**：添加噪声的重点不在于添加与否，添加的数量取决于机构愿为实用目的放弃多少隐私。在“原理揭示”的示例中，调查员可使用两个及以上个人数据代替单人数据输入，从而为计算引入更多噪声。这能更大程度确保第十人的隐私，然而因为房间里的其他人不确信平均工资值是否准确，因此这对他们判断自己薪水高低并无太大帮助。
- **或将能对隐私泄露概率进行数理计量**：在实际操作中，“噪声”通过严谨的数学公式进行添加（并计算其在汇总统计结果中的占比），根据“差分隐私”算法进行计量。这有可能导致统计计量隐私泄露。金融机构可就所得数值判断隐私泄露概率是否在可接受范围内。通过添加受控噪音，该技术能够根据具体数据的敏感程度作灵活调整。
- **低计算成本**：与传统的直接传输数据相比，在数据共享下添加噪声不需要额外增加大量计算。在“原理揭示”的示例中，中介能轻易的在分析中添加随机噪声因子，随后照常计算出平均值。

限制：

- **只能用于大型数据集**：若对小规模数据集添加噪声，无法在保护数据提供者隐私的同时，为汇总统计提供足够精确的信息。在“原理揭示”的示例中，如果房间里只有三人，那么替换其中一人的数据就会给平均值结果带来巨大影响，各人均无法判断自己的薪资高低。
- **限制准确性**：添加噪声以后，对输入或输出的计算最终都会导致分析的准确性下降。因此，差分隐私不能用于对准确性要求严格的特定情形（例如，异常检测，因其需要检测数值之间微小却具有统计显著性的差别）。

第二项技术：联合分析

优势：

- **沟通成本最小化**：在某些涉及大量数据的情况时，数据共享本身的成本可能极高。而联合分析允许对更简洁的洞察进行共享而不涉及大量数据。在“原理揭示”的示例中，只需共享反垃圾邮件引擎的洞察结果，而无须将所有垃圾邮件都复制到中央数据库中。

限制：

- **每个数据集的数据须达到一定规模**：因为联合分析假定能够从单独的数据集中得出有意义的洞察：在某些情况下，数据尚未达到一定规模，从而导致联合分析所得数值受限。在“原理揭示”的示例中，如果每个邮件提供商缺少足够的数据来建立垃圾邮件防范模型，联合分析也就无法得到有价值的结果。
- **分布式系统更复杂**：管理联合生态系统远比管理传统集中式数据库更为复杂。在“原理揭示”的示例中，当各邮件提供商分别开发反垃圾邮件引擎时，会得到三套分析数据（分别来自三个公司），而三个公司没有就此进行交流。当三个公司建立集中式数据库时，就会出现三套通信数据（因为要将三方数据合并到一个数据库中）和一套分析数据（对集中式数据库中的数据进行分析）。而在联合生态系统中，同时有三套分析数据（三个公司均自行分析）以及三套通信数据（三个公司要共享内部分析所得洞察）。尽管如“优势”部分所述，联合分析所传输数据的量成本较低，但通信则更加复杂。

第三项技术：同态加密

优势：

- **不用依赖第三方保证隐私保护程度**：大多数情况下隐私保护的实现需要借助第三方（如认证机构）。然而使用同态加密，将无须依赖第三方，且数据能与更多参与者共享。使用同态加密，供应商不用为进入市场和吸引顾客而进行各项繁琐的认证，这将为市场释放更多的竞争力和创新力。在“原理揭示”的示例中，若同态加密运用得当，Susan在寻找健康数据分析公司时，只需根据分析质量来做选择，而不用考虑HealthAnalytics Co. 能否妥善管理其数据或是否拥有完善的安全协议。

限制：

- **技术限制**：支持同态加密的现有技术不够简单高效
  - 分析全同态加密的数据会比分析底层的加密数据慢几个数量级（取决于计算的复杂程度）。增加了计算成本，同时意味着全同态加密只适用于时间成本要求不高的用例。在“原理揭示”的示例中，HealthAnalytics Co.要花大量时间才能从Susan提供的健康记录中得到有价值的分析，而使用传统方式进行数据共享，耗时将大大减少。
  - 另一种同态加密能加快分析速度，但只能对底层数据进行一种或几种运算（例如，加法或乘法，而不是两者同时进行）。这种加密是同态加密 (HE)，与之相对是全同态加密（FHE）。因为同态加密不能同时进行多种运算，所以分析速度会更快，但无法从数据中获得更有深度的洞察。在“原理揭示”的示例中，如HealthAnalytics Co.仅能进行简单的运算，分析将会受到限制，得出的洞察意义也不大。
- **结果核验**：大多数HE和FHE方案无法被验证，意味着没有证据证明计算结果的准确性。因此，使用同态加密需要保证其方案的准确性以及未被干扰。可验证的（完全）同态加密正在开发中，但在技术方面会受到更大的限制。

第四项技术：零知识证明

优势：

- **易实施**：零知识证明在数学运算方面并不复杂，能够轻松与现有系统相结合。在“原理揭示”的示例中，Peggy和Victor并不需要复杂的数学运算就能完成信息交换。
- **在不过多影响客户体验的情况下，提高安全性**：许多其他安全和隐私保护措施都会使客户体验打折扣。例如，零售支付中的双因素认证会降低购买效率（客户需要收到手机短信才能进行身份验证），使信用卡支付变得更麻烦。将零知识证明与支付相结合，客户将无须进行复杂操作。在“原理揭示”的示例中，“品尝/调换杯子/再次品尝”这个过程需要多次重复，Victor才能确定Peggy的确知道两种品牌苏打水的区别而不是在猜测，实际操作中，这种交互过程会由高速的计算机完成。

限制：

- **计算成本高**：客户使用传统的交互式证明时无须进行过多交流，而零知识证明需要做更多的工作。如在“原理揭示”示例的初次证明中，Peggy可直接告诉Victor区分苏打水的方式。而采用零知识证明时，Victor多次调换杯子，而每次调换前后都让Peggy品尝两杯苏打水，这表明了零知识证明需要做更多的工作。Peggy和Victor在采用非交互式的零知识证明时无须反复交流，但是证明者（Peggy）需在零知识证明系统上耗费更多精力。

第五项技术：安全多方计算

优势：

- **无须依赖第三方：**大多数情况下安全和隐私的实现需要借助第三方（如数据分析公司）。安全多方计算通过使目标相同的个人互相协作来摆脱对第三方的依赖。通过多方共识，各方无须在数据共享协作中信赖其他参与者。在“原理揭示”的示例中，每个人都无须信赖其他参与者，而只需相信合作的最终目的是达成预期目标。阈值是可进行设定的，其高低取决于数据共享协作参与者对彼此的信任程度（以及其他因素）。
- **计算成本低：**与同态加密不同，安全多方计算不进行复杂加密，亦不对加密数据进行分析，使得分析本身更加容易进行。在“原理揭示”的示例中，仅就数学运算而言，相较于应用设想中采用可信中介的情形，采用安全多方计算几乎无需增加额外的工作。

限制：

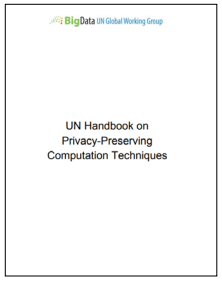
- **通信成本高：**与同态加密不同，安全多方计算的通信成本明显更高。在“原理揭示”的示例中，分析得到的结果虽简单却需各方完成多个步骤，而采用可信的中介能为各参与方省去很多反复工作。
- **设置成本高：**安全多方计算系统需根据具体用例单独进行设计和定制。因此，系统设置成本高且耗时长。相比而言，无论共享的数据和所作的分析如何，差分隐私均采用标准化且通用的数学公式。

相关阅读资料

Protecting Privacy in Practice（英国皇家学会）

Is Privacy Privacy?（伯克曼克莱恩互联网及社会研究中心）

UN Handbook on Privacy Preserving Computation Techniques（联合国全球大数据工作组）



鸣谢

项目团队：

Jesse McWaters 世界经济论坛金融业创新项目主管
jesse.mcwaters@weforum.org

Matthew Blake 世界经济论坛金融和货币体系未来部门负责人
matthew.blake@weforum.org

Rob Galaski 德勤管理咨询银行业及资本市场全球领导人
rgalaski@deloitte.ca

Hemanth Soni 德勤加拿大德勤摩立特高级顾问
hemasoni@deloitte.ca

Ishani Majumdar 德勤加拿大Omnia AI高级顾问
ismajumdar@deloitte.ca

指导委员会：

Josh Bottomley 汇丰银行全球数字主管

Pierre-Olivier Bouée 瑞信首席运营官

Nick Cafferillo 标普全球首席数据与技术官

Vanessa Colella 花旗风投首席创新官兼主管

Juan Colombas 劳埃德银行集团执行董事兼首席运营官

Robert Contri 德勤金融服务行业全球领导合伙人

David Craig 路孚特创始人、首席执行官及董事会成员

Tony Cyriac 蒙特利尔银行首席数据分析官

Rob Goldstein 贝莱德首席运营官兼Blackrock Solutions全球主管

Greg Jensen 桥水联合基金联席首席投资官

Axel P. Lehmann 教授，博士，瑞士联合银行瑞士个人与企业银行业务总裁

Lena Mass-Cresnik 博士，美驰集团首席数据官

Max Neukirchen 摩根大通企业战略负责人

Kush Saxena 万事达卡首席技术官

Nicolas de Skowronski 瑞士宝盛理事会成员、咨询解决方案负责人

Michael Zerbs 丰业银行集团总裁兼首席技术官

工作组：

Sami Ahmed 蒙特利尔银行金融集团副总裁兼数字、分析与人工智能转型业务主管

Secil Arslan Yapi Kredi银行研发与特殊项目主管

Tim Baker 路孚特全球应用创新主管

Beth Devin 花旗创新网络及新兴技术常务总监兼主管

Roland Fejfar 摩根士丹利欧洲中东和非洲地区/亚太区技术业务开发与创新主管

Gero Gunkel 苏黎世保险集团人工智能主管

James Harborne 汇丰银行数字公共政策主管

Milos Krstajic 德国安联集团索赔业务人工智能解决方案主管

Wei-Lin Lee 贝宝战略与发展高级总监

Juan Martinez 环球同业银行金融电讯协会开放程序接口、身份识别及连接业务全球主管

Michael O' Rourke 纳斯达克机器智能与全球信息服务主管

Jennifer Peve 存托及结算机构业务发展与金融科技战略常务总监

Jim Psota 磐聚网联合创始人兼首席技术官（标普全球）

Nicole Sandler 巴克莱全球创新政策主管

Annika Schröder 瑞士联合银行人工智能卓越中心执行董事

Chadwick Westlake 丰业银行企业生产力与加拿大银行金融业务执行副总裁

特别鸣谢：

Pavle Avramovic 金融行为监管局专员

Jordan Brandt Inpher联合创始人兼首席执行官

Andrew Burt Immuta首席隐私官兼法务工程师

Anton Dimitrov Inpher高级软件工程师

Mariya Georgieva Inpher安全创新总监

Dimitar Jetchev Inpher联合创始人兼首席技术官

Iraklis Leontiadis Inpher加密与安全研究高级工程师

Kevin McCarthy Inpher副总裁

Bhaskar Medhi Ziroh Labs联合创始人兼首席执行官

Alfred Rossi Immuta研究科学家

额外鸣谢：

Derek Baraldi

Mary Emma Barton

Andre Belelieu

Kerry Butts

Alexandra Durbak

Natalya Guseva

Kai Keller

Courtney Kidd Chubb

Abel Lee

Nicole Peerless

Denizhan Uykur

Han Yik

尾注

1. https://www.thomsonreuters.com/en/press-releases/2018/june/thomson-reuters-expands-sentiment-da- ta-to-track-top-100-cryptocurrencies.html (访问于2019年9月9日)
2. https://www.newswire.ca/news-releases/wealthsimple-announces-partnership-with-mint-577957751.html (访问于2019年8月5日)
3. https://www.nordea.com/en/press-and-news/news-and-press-releases/press-releases/2019/07-05-14h00-thecol-laboration-of-six-nordic-banks-results-in-a-joint-kyc-company.html (访问于2019年8月5日)
4. https://www.gov.uk/cma-cases/review-of-banking-for-small-and-medium-sized-businesses-smes-in-the-uk (访问于2019年8月5日)
5. https://www.lenddo.com/products.html#creditscore (访问于2019年8月5日)
6. https://marketingland.com/survey-58-will-share-personal-data-under-the-right-circumstances-242750 (访问于2019年8月5日)
7. 注：此处并非表示监管机构要求金融机构共享数据 (如PSD2) 的作法与保护客户隐私的相关法规 (如GDPR) 是矛盾的。两种监管形式的目的是统一的，即让客户在如何管理自身数据方面有更多掌控权，并知晓自身数据是如何被使用的。然而，这二者的确为金融机构带来了难题：在让数据共享变得更为复杂的同时又要求各机构对某些数据集 (如交易数据) 进行共享。
8. https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act (访问于2019年8月5日)
9. https://newsroom.ibm.com/Cybersecurity-and-Privacy-Research (访问于2019年8月5日)
10. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (访问于2019年8月5日)
11. http://press.capitalone.com/phoenix.zhtml?c=251626&p=irol-newsArticle&ID=2405043 (访问于2019年8月5日)
12. https://www.blog.google/technology/safety-security/project-strobe (访问于2019年8月5日)
13. https://techcrunch.com/2019/01/31/aadhaar-data-leak (访问于2019年8月5日)
14. 注：重要的是，如果相关信息被滥用而助长歧视或侵害，那么使用客户过多信息可能无意间对其造成伤害。在世界经济论坛即将推出的一篇关于人工智能的报告中我们会对此作详细探讨。
15. 注：这些技术不会相互制约，可结合使用。如用例中所示，几类技术是相互联系的，可一并采用。将一些技术引入中间环节，会有助于相关技术的实施 (如零知识证明对安全多方计算的作用) 。此外还有诸如可信执行环境等其他有关技术，在本文中不作探讨。
16. https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html (访问于2019年8月5日)
17. Cynthia Dwork, Frank McSherry, Kobbi Nissim及Adam Smith (参见注释19)
18. https://www.microsoft.com/en-us/research/publication/calibrating-noise-to-sensitivity-in-private-data-analysis (访问于2019年8月5日)
19. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf (访问于2019年8月5日)
20. https://github.com/OpenMined/PySyft (访问于2019年8月5日)
21. https://medium.com/tensorflow/introducing-tensorflow-privacy-learning-with-differential-privacy-for-training-da- ta-b143c5e801b6 (访问于2019年8月5日)
22. https://techscience.org/a/2015092903 (访问于2019年8月5日)
23. 世界经济论坛一篇题为《联合数据系统：在敏感数据使用时平衡创新与信任》(Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data) 的白皮书围绕健康相关数据对这一概念作了详细探讨。
24. https://www.groundai.com/project/applied-federated-learning-improving-google-keyboard-query-suggestions (访问于2019年8月5日)
25. https://medium.com/tensorflow/introducing-tensorflow-federated-a4147aa20041 (访问于2019年8月5日)
26. https://thenextweb.com/security/2017/12/05/personal-info-31-million-people-leaked-popular-virtual-keyboard-ai-type (访问于2019年8月5日)

27. 注：单靠联合分析并不足以防止敏感个人信息被转化形成集中式主模型，还需结合其他隐私保护措施（如过滤器）。
28. https://venturebeat.com/2019/06/03/how-federated-learning-could-shape-the-future-of-ai-in-a-privacy-obsessed-world (访问于2019年8月5日)
29. https://www.techrepublic.com/article/is-homomorphic-encryption-ready-to-deliver-confidential-cloud-comput- ing-to-enterprises (访问于2019年8月5日)
30. https://crypto.stanford.edu/craig (访问于2019年8月5日)
31. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (访问于2019年8月5日)
32. https://dl.acm.org/citation.cfm?id=22178 (访问于2019年8月5日)
33. https://www.wthr.com/article/kohls-cashier-accused-stealing-customers-credit-card-numbers-and-using-them (访问于2019年8月5日)
34. https://www.itwire.com/security/85702-warning,-card-fraud-could-cost-retailers-us130-billion-over-next-5-years.html (访问于2019年8月5日)
35. https://www.americanbanker.com/news/how-zcash-tries-to-balance-privacy-transparency-in-blockchain (访问于2019年8月5日)
36. https://coinjournal.net/blockchain-security-platform-nucypher-raises-us4-3m-cryptofunds-vcs (访问于2019年8月5日)
37. https://www.unlock-bc.com/news/2019-04-30/nuggets-selected-for-fca-innovation-sandbox-for-blockchain-solution (访问于2019年8月5日)
38. https://www.ingwb.com/themes/distributed-ledger-technology-articles/ing-launches-major-addition-to-block- chain-technology (访问于2019年8月5日)
39. 与多方计算涉及多个实体的情况不同，安全两方计算仅涉及两个实体的数据共享，是安全多方计算的一个特例。本文将集中讨论多方计算。
40. https://apps.dtic.mil/dtic/tr/fulltext/u2/a066331.pdf (访问于2019年8月5日)
41. https://ercim-news.ercim.eu/en73/special/trading-sugar-beet-quotas-secure-multiparty-computation-in-practice (访问于2019年8月5日)
42. https://www.space.com/5542-satellite-destroyed-space-collision.html (访问于2019年8月5日)
43. https://www.fbi.gov/stats-services/publications/insurance-fraud (访问于2019年8月5日)
44. Federal Deposit Insurance Company, https://research.fdic.gov/bankfind/index.html (访问于2019年8月5日)
45. https://www.bbva.com/en/bbva-launches-first-baas-platform-in-the-u-s (访问于2019年8月5日)
46. 注：相较于采用应用程序接口的系统，安全多方计算系统所需设置更为复杂，也更耗时，且需针对参与计算的各方定制系统，故可能只适用于涉及高度敏感信息的重要用例。

    关于该技术在零售银行业务以外的应用，对冲基金评估第三方数据提供商的情形可作为一个示例。对冲基金想了解行将采购的数据能否有助于改良其模型的质量，而数据提供商在未收到全款前又不愿向其披露自身的专有数据集。通过小样本数据或历史数据无法精确反映对冲基金模型中数据的实际效用。借助安全多方计算系统，双方能算出该专有数据集能够给对冲基金模型带来的效用而无需披露数据本身。
47. https://news.gallup.com/poll/162872/one-three-americans-prepare-detailed-household-budget.aspx (访问于2019年8月5日)
48. https://benefittrends.metlife.com/us-perspectives/work-redefined-a-new-age-of-benefits (访问于2019年8月5日)
49. https://www.fraud-magazine.com/article.aspx?id=4294990598 (访问于2019年8月5日)
50. https://www.weforum.org/reports/disruptive-innovation-in-financial-services-a-blueprint-for-digital (访问于2019年8月5日)
51. https://arxiv.org/abs/1111.5228 (访问于2019年8月5日)

WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

世界经济论坛是推动公私合作的国际组织，致力于改善世界状况。

论坛汇聚政界、商界等社会各界重要领袖，共同制定全球、区域和行业议程。

世界经济论坛
地址：91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

电话：+41 (0) 22 869 1212
传真：+41 (0) 22 786 2744

邮箱：contact@weforum.org
网址：www.weforum.org