中国科学: 信息科学 2024 年 第 54 卷 第 11 期: 2589–2605
SCIENTIA SINICA Informationis
论文

支持容错的轻量级可验证隐私保护传染病监测数据聚合方案

杨小东[1], 杨兰[1*], 魏丽珍[1], 杜小妮[2], 王彩芬[3]

1. 西北师范大学计算机科学与工程学院, 兰州 730070
2. 西北师范大学数学与统计学院, 兰州 730070
3. 深圳技术大学大数据与网络学院, 深圳 518118
* 通信作者. E-mail: 15389057097@163.com, 2022212139@nwnu.edu.cn

收稿日期: 2024–07–27; 修回日期: 2024–09–09; 接受日期: 2024–09–10; 网络出版日期: 2024–11–08

甘肃省重点研发计划 (批准号: 23YFGA0081)、国家自然科学基金 (批准号: 62362059, 62172337) 和甘肃省教育厅产业支持计划项目 (批准号: 2023CYZC-09) 资助

摘要 随着各种流行传染病在全球频繁暴发, 传染病监测在阻止传染病传播方面发挥着至关重要的作用. 隐私保护数据聚合技术常用于避免传染病监测数据传输造成的用户隐私泄露问题. 然而, 现有的数据聚合方案仍然具有一些安全问题, 如聚合节点不可信等. 为了解决这些问题, 本文提出了一个支持容错的轻量级可验证隐私保护传染病监测数据聚合方案. 首先, 使用基于 CRT (Chinese remainder theorem) 改进的 Paillier 同态加密系统和支持批量验证的签名算法分别对传染病数据进行高效加密和签名, 以保护数据传输过程中的数据隐私和数据完整性. 其次, 使用承诺机制解决聚合节点不可信的问题. 此外, 本方案支持容错, 即使某些用户和聚合节点没有按时地上传数据, 聚合工作依然能够继续. 特别地, 本方案能够抵抗合谋攻击, 满足更高的安全需求. 由于本方案没有使用高耗时的计算操作, 如双线性映射等, 仿真实验证明本方案具有优秀的计算和通信开销, 可以安全有效地应用于传染病检测系统.

关键词 传染病监测, 数据聚合, 隐私保护, 同态加密, 轻量级

1 引言

在传染病暴发初期, 由于初步诊断的传染病数据不足, 很容易导致短时间内物资需求突然增加, 从而出现公共卫生资源的严重短缺. 这导致了一些传染病传播早期阶段的高发病率和高死亡率 [1]. 传染病数据的实时传播数据能够诠释传染病的传播轨迹, 为政府公共卫生机构实施相应措施以遏制传染病的进一步传播提供及时的辅助. 为了方便实时展示传染病的传播情况, 需要从诊所、医院、医学检测实

引用格式: 杨小东, 杨兰, 魏丽珍, 等. 支持容错的轻量级可验证隐私保护传染病监测数据聚合方案. 中国科学: 信息科学, 2024, 54: 2589–2605, doi: 10.1360/SSI-2024-0228
Yang X D, Yang L, Wei L Z, et al. Lightweight verifiable privacy-preserving infectious disease surveillance data aggregation scheme with fault tolerance (in Chinese). Sci Sin Inform, 2024, 54: 2589–2605, doi: 10.1360/SSI-2024-0228

验室和 CDC (centers for disease control) 等公共卫生机构获取传染病相关数据. 相应地, 安全和高效地传输传染病监测数据是必要的. 数据聚合是疾病监测系统的重要组成部分. 通过对特定综合征病人的数量进行统计, 可以发现具有重要意义的趋势. 然而, 单个病例数据往往无法检测传染病. 此外, 统计到的一些地理位置信息有助于对传染病进行定位.

传染病数据可能会包含国籍、城市、种类、数据属性和数据提供者等隐私信息. 这些信息通常是敏感的, 可能会暴露病患的私人信息 [2]. 通常来说, 如果传染病监测数据直接进行明文聚合操作处理, 这些数据很有可能泄露或者被恶意用户窃取. 如果数据提供者直接将未受保护的隐私数据上传至边缘服务器, 恶意用户获取到未受保护的私有数据后, 进一步推断出病人的隐私信息, 从而给他们造成心理和经济损失 [3]. 同时, 数据提供者也不愿意让别人知道自己受到传染病暴发的影响. 因此, 对传染病数据的隐私保护是必需的, 以打消数据提供者的顾虑, 增加数据提供者的参与度, 保持传染病监测数据的新鲜度和真实性.

隐私保护数据聚合技术是在传染病检测系统中能够同时实现数据可用性和隐私保护的最有前景的方法. 除了保护用户隐私之外, 一个理想的隐私保护数据聚合方案应当是轻量级的和健壮的. 虽然目前已经有许多隐私保护数据聚合方案, 但是仍有许多问题需要解决. 除了基础的数据隐私保护问题外, 比如数据机密性、数据的完整性和认证, 还需要考虑如何抵抗用户合谋攻击, 从而实现强隐私保护. 目前的许多数据聚合隐私保护方案都假设聚合节点是可信的, 但在实际应用中, 许多聚合节点会为了节省计算资源不诚实地执行聚合操作. 这对后续的分析和统计至关重要. 此外, 在传染病监测这一关键领域内, 面对疫情暴发时所产生的庞大且复杂的数据集, 鉴于传染病数据本身具备的高速传播特性, 这对监测系统提出了极为严苛的性能挑战, 尤其是在计算资源消耗与通信效率方面, 均要求达到极高的标准. 因此, 设计一个轻量级的聚合方案是必要的.

为了解决上述问题, 我们提出了以下解决方案. 在我们提出的方案中, 首先, 考虑到许多方案挨个对签名进行验证, 在数据量较大的时候, 逐个验证签名会耗费大量的时间. 本方案通过使用基于 CRT (Chinese remainder theorem) 改进的 Paillier 同态密码系统和支持批量验证的签名算法, 分别保证了传染病监测数据的机密性和完整性. 其中通过改进的 Paillier 密码系统和批量验证技术, 显著降低了大规模数据验证过程中的计算和通信开销. 其次, 本方案在算法设计上可以规避高耗时的运算操作, 如双线性映射等, 从而实现了运算效率的显著提升. 此外, 由于第三方聚合节点并不都是诚实地执行聚合操作, 可能会因为自身利益或者减少计算开销而不正确地执行聚合操作. 本方案巧妙地融入了承诺机制, 以此来验证聚合节点是否值得信任. 再者, 面对用户无法提交数据的情况, 本方案也可以正确解密与后续处理. 最后, 为了提升系统的健壮性与可靠性, 特别强调了容错机制是聚合方案在现实应用中的关键.

我们的贡献可以总结如下. 本方案使用基于 CRT 改进的 Paillier 同态加密技术和支持批量验证的签名算法, 除了保证传染病监测数据的机密性和完整性之外, 面对疫情暴发时所产生的庞大且实时的数据集的传输, 显著地降低了数据传输过程中的通信和计算开销; 其次, 本方案没有使用高耗时的运算操作, 如双线性映射等, 从而提升了运算效率; 此外, 本方案通过引用承诺机制, 检查聚合数据的准确性, 以解决聚合节点不可信的问题; 最后, 支持数据提供者和聚合节点由于各种原因未按时上传数据时, 聚合工作依然能够继续, 提高了系统的健壮性.

2 相关工作

目前, 有许多基于同态加密的方案用于数据聚合和通信过程中的隐私保护 (privacy-preserving data

aggregation, PPDA). 利用同态加密算法 [4] 的同态性质, 使数据能够以密文形式进行特定的聚合和分析操作. 并且, 其解密结果与直接对明文进行相应操作的结果一致. 在本节中, 我们将从以下 5 个方面来描述相关工作: 医疗系统中的 PPDA、多领域多维 PPDA、可验证 PPDA、健壮的 PPDA、轻量级 PPDA.

2.1 医疗系统中的 PPDA

Han 等 [5] 于 2015 年提出了一种隐私保护和多功能医疗数据聚合方案, 支持多源医疗数据的多功能加性和非加性聚合. 2017 年, He 等 [6] 提出了一种基于 BGN (Boneh-Goh-Nissim) 密码系统的隐私保护数据聚合方案, 并且可以抵抗内部攻击者. Liu 等 [7] 于 2018 年提出了一种不需要依赖于可信第三方的数据聚合方案. 该方案将可信用户组成一个虚拟聚合区域, 在保护用户隐私的同时提高了系统的鲁棒性. Li 等 [8] 提出了一种基于 Paillier 密码系统的隐私保护多子集数据聚合方案, 该方案在满足智能电网控制中心细粒度数据需求的同时, 还能保护用户的隐私安全. Tang 等 [9] 在 2019 年提出了一种安全的多源医疗保健数据聚合方案, 使用 BGN 密码系统和 Shamir 的秘密共享 [10], 并获得公平的用户激励.

2.2 多领域多维 PPDA

目前, 已经有大量应用于各个领域的多维数据聚合方案被提出. Chen 等 [11] 提出了一种基于 Paillier 密码系统的智能电表数据聚合方案. 它使智能电表能够在单个报告消息中报告多种类型的数据, 从而使供应商能够对数据进行方差分析和单向方差分析. 然而, 该方案使用了耗时的操作, 如双线性映射, 因此具有较高的计算开销. Zhang 等 [12] 提出了一种应用于智能电网的可验证的隐私保护多类型数据聚合方案. 具体来说, 使用 Paillier 密码系统和 BLS (Boneh-Lynn-Shacham) 短签名技术 [13], 聚合加密后的多类型数据, 并可以在不知道明文数据的情况下获得聚合数据的统计分析结果 (如均值、方差等). Peng 等 [14] 提出了一种在物联网 (Internet of Things, IoTs) 环境下的隐私保护多维数据聚合方案. 该方案设计了一种基于中国剩余定理的同态加密方案, 能够将一个多维小整数向量加密成一个密文, 并维持各维度的线性同态性. Shang 等 [15] 提出了一种健壮的、隐私保护多维数据聚合方案. 该方案除了满足传统的安全特性 (如机密性、完整性等) 外, 还实现了差分隐私保护. Zhang 等 [16] 提出了一种对智能电网系统中多维聚合数据进行统计分析的隐私保护方案. 然而, 该方案采用了改进的 BGN 算法, 导致计算开销高, 通信效率低. Hu 等 [17] 设计了一种安全增强的数据共享方案, 使用一种自主的、可控的动态假名机制保护车辆隐私.

2.3 可验证 PPDA

上述方案虽然能够处理多维数据, 但都不能验证聚合结果的正确性. Zhang 等 [18] 提出了一个支持区块链分散一致性的验证方案. 监督链可以在不透露任何关于数据的原始信息的情况下检查跨链数据计算的正确性和完整性. Zhao 等 [19] 设计了一个可验证的联邦学习方案. 通过构建一种基于较短的同态签名的聚合结果验证方案, 用户可以验证服务器聚合结果的正确性. Zhang 等 [20] 提出了一种基于细粒度访问控制的分布式数据聚合方案, 并通过零知识证明验证了聚合结果. 但是, 该方案的密文由两部分构成, 并且使用了大量耗时的运算, 因此其时间开销较高. Wang 等 [21] 提出了一种基于联邦学习的高效、强隐私保护数据聚合方案, 利用 Pedersen 承诺验证了云服务器端聚合结果的正确性.

2.4 健壮的 PPDA

除了用户隐私的保护之外, 一个理想的隐私保护数据聚合方案应当是健壮的. 为了提升系统的健壮性, Mei 等 [22] 设计了一个无需可信第三方的高效的隐私保护数据聚合方案. 该方案消除了对可信第三方的依赖, 结合 Shamir 门限秘密共享方案提升了方案的整体容错能力. Zhang 等 [23] 提出了一种具有多功能的细粒度隐私保护数据聚合方案. 该方案将 HMAC (hash-based message authentication code) 和公钥加密与等值测试相结合, 抵抗内部以及外部敌手发起的恶意数据挖掘攻击. Zhong 等 [24] 采用轻量级对称加密算法和数学结构来实现数据聚合, 提出了一种具有隐私增强特性的轻量级容错数据聚合方案. Xu 等 [25] 为了解决隐私泄露和数据聚合效率低下的问题, 设计了基于联邦学习的隐私保护数据聚合协议. 该方案基于 Shamir 秘密共享方案、Paillier 同态密码系统等方法, 保证了模型参数的安全. 利用优化的容错机制, 减少了可信机构的频繁参与.

2.5 轻量级 PPDA

Wu 等 [26] 为了实现隐私保护、轻量级加密和容错机制, 提出了一种具有重加密可验证的同态阈值代理重加密方案. 通过将智能电表和控制中心的计算开销转移到边缘节点, 从而提升加密效率. Gope 等 [27] 构建了一种轻量级的、隐私友好的通过屏蔽的数据聚合方案. Ding 等 [28] 构建了一个高效的基于身份的 PPDA 方案, 该方案支持批量验证技术. Zhang 等 [29] 设计了一种支持在线/离线签名轻量级可验证 PPDA 方案.

现有的隐私保护数据聚合方案能够解决一些方面的问题, 但是它们都有一定的局限性: (1) 许多方案未考虑到聚合节点是否能够信赖的问题, 它们有可能会为了自身利益或者节省计算资源而不诚实地执行聚合操作, 导致最后统计的聚合数据不准确; (2) 大多数方案使用了高耗时的操作, 如: 双线性映射等, 导致许多方案的计算和通信效率低下; (3) 很多聚合方案缺乏健壮性, 不支持容错机制. 在实际应用中, 数据提供者和聚合节点可能没有按时上传数据, 因此, 容错机制是聚合方案在现实生活中应用的关键.

3 预备知识

在本节中, 我们首先定义本方案的系统模型和安全模型并阐明我们方案的安全目标. 然后简单介绍基于 CRT 改进的 Paillier 同态密码系统、秘密共享、支持批量验证的签名算法和承诺机制.

3.1 系统模型

系统模型如图 1 所示, 包括 5 种实体: 数据提供者 (data provider, DP)、聚合机构 (aggregation agency, AGA)、政府部门 (government, GOV)、区块链 (blockchain, BC) 和可信机构 (trust authority, TA).

TA: 可信机构是系统中可靠的实体, 负责初始化系统. 在初始化阶段, TA 生成并发布系统公共参数.

DP: 每个 DP 生成自己的多维传染病数据报告. 使用加密算法保护传染病数据的隐私, 并计算出相应的承诺值. 将传染病数据报告上传到区块链账本.

AGA: 每隔一段时间从对应的区块链账本中提取传染病数据报告密文, AGA 进行聚合操作. 将聚合结果上传到区块链.

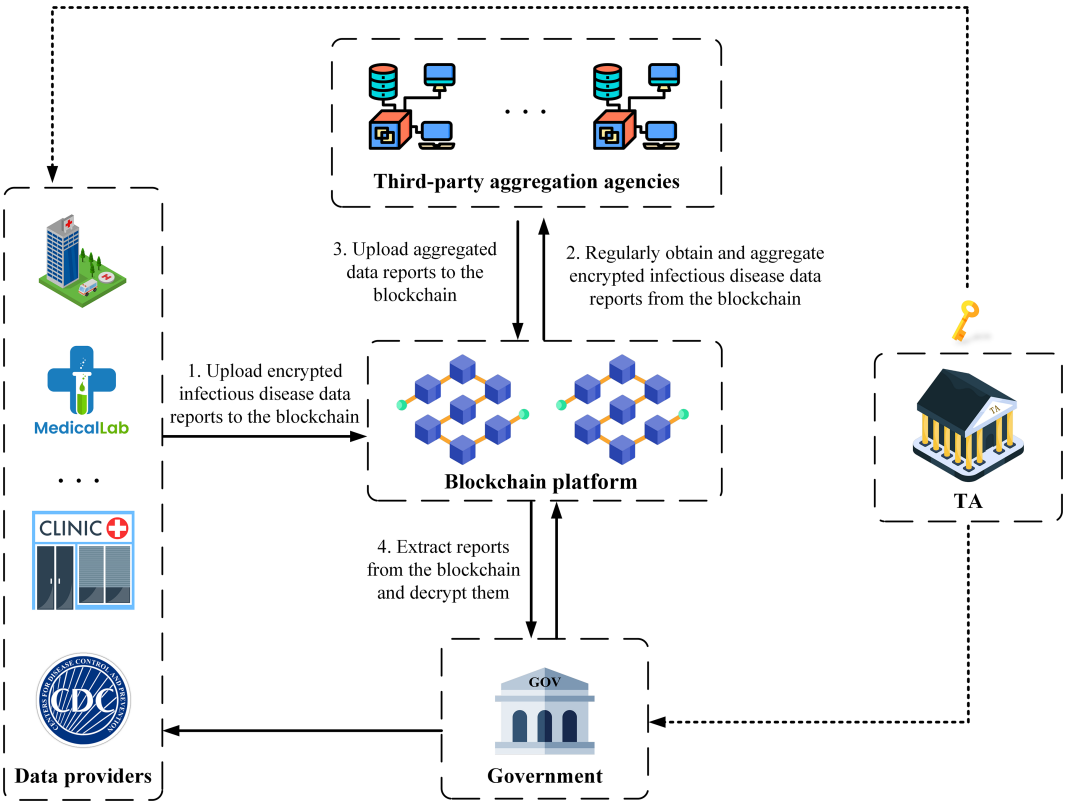

图 1 (网络版彩图) 系统模型

Figure 1 (Color online) System model

GOV: 定期从区块链中提取聚合传染病报告密文, 并对聚合后的数据进行解密并进行进一步的分析.

BC: 由于其去中心化、不变性和透明的特点, 其主要功能是存储, 以保证数据的安全性和透明性.

3.2 安全模型

在我们的安全模型, 为了更接近真实的传染病监测系统. 我们假设 TA 是完全可信的组织, 它们将诚实地执行相应的操作. 此外, GOV 和 AGA 被认为是诚实但好奇的, 也就是会遵循协议执行相应的操作, 但也渴望知道用户的隐私信息. DP 会诚实地上传自己的数据报告, 但是会渴望知道别的用户的隐私信息.

然而, 我们假设公共互联网中的恶意用户可能对传染病数据报告隐私感兴趣. 假设存在一个外部攻击者, 可能会尝试修改或者篡改传输过程中的传染病数据报告. 攻击者的目的是通过窃听网络传输获取传染病数据报告和聚合数据报告, 进而推断用户的隐私信息. 系统的内部攻击者可能试图分析和提取传染病数据报告中蕴含的敏感和私人信息. 此外, 我们允许至多 $(k-1)$ 个 DP 与 GOV 串通, 其中 k 是门限值.

3.3 安全目标

本方案最重要的设计目标, 是在保护用户个人隐私的同时, 增强传染病数据报告的有用性和系统的实用性. 体现在以下几个方面.

强隐私保护: 传染病监测系统中的所有传染病监测报告都应得到保护. 为保护各数据提供者贡献的传染病数据, 除数据提供者自身外, 任何实体不得获取数据提供者的传染病报告. 为了保护聚合报表, 除 AGA 外, 不允许 GOV 以外的所有实体获取聚合数据结果. 利用阈值方法可以增强方案对合谋攻击的抵抗能力. 为了恢复消息, 攻击者需要拥有至少 k 共享. 本文通过抵抗来自至少 $(k-1)$ 数据提供者的共谋尝试实现了强隐私保护.

聚合结果正确性: 在本方案中, 聚合机构 AGA 应当能够正确执行聚合操作. 最后, 能够得到正确的聚合结果.

轻量级: 传染病监测系统的聚合方案应尽可能地减少计算开销, 从而实现轻量级计算, 高效地处理收集到的传染病数据.

低成本容错: 当数据提供者偶尔故障而无法在按时上传传染病数据, 政府部门仍然可以通过容错机制获得正确的全局聚合结果.

3.4 基于 CRT 改进的 Paillier 同态密码系统

在空间层面, CRT 算法描述了两个代数空间的同构. 也就是说, 一个代数空间可以分解为许多相互正交的子空间. 此外, 分解后的空间可以与原始空间保持一对一的映射关系, 就像同一空间的两个表示一样. 特别地, 有 $n=pq$, 其中 p 和 q 是互质的. 它们满足几何空间同构的性质 $\overrightarrow{Z_n} \simeq \overrightarrow{Z_p} \times \overrightarrow{Z_q}$, 这就使得域 $\overrightarrow{Z_n}$ 中的操作可以被转换到域 $(\overrightarrow{Z_p}, \overrightarrow{Z_q})$ 中的操作.

Paillier 同态密码算法包括 3 个部分: 密钥生成 Gen(k)、加密 Enc() 和解密 Dec() 操作. 为了提升加密算法的执行速度, 可以利用 CRT 的空间特性对数据进行预处理. 在 $Z_{N^2}^*$ 下的指数运算可以变换成 $Z_{p^2}^*$ 和 $Z_{q^2}^*$ 下的运算, 从而提高了 Paillier 加密算法的加密速度.

Gen(k). 给定一个安全参数 κ, 随机选择两个大质数 p 和 q, 满足 $|p|=|q|=\kappa$ 和 $\gcd(pq,(p-1)(q-1))=1$, 计算 $N=pq$ 和 $\lambda=\mathrm{lcm}(p-1,q-1)$. 随机选择 $g,R\in Z_{N^2}^*$, 满足 $\gcd(L(R^\lambda \bmod n^2),N)=1$. 定义两个等式: $L(x)=(x-1)/N$ 和 $\mu=(L(g^\lambda \bmod N^2))^{-1} \bmod N$.

Enc(). 计算 p^2 和 q^2 模拟元分别为 $\mu_1=(p^2)^{-1} \bmod p^2$ 和 $\mu_2=(q^2)^{-1} \bmod q^2$, 令 $\ell_1=\mu_1\cdot p^2$, $\ell_2=\mu_2\cdot q^2$. 分别计算 $k_1=g^m \bmod p^2$, $k_2=g^m \bmod q^2$, $k_3=R^N \bmod p^2$, $k_4=R^N \bmod q^2$, $c_1=g^m \bmod N^2=(k_1\cdot\ell_1+k_2\cdot\ell_2) \bmod N^2$ 和 $c_2=R^N \bmod N^2=(k_3\cdot\ell_1+k_4\cdot\ell_2) \bmod N^2$. 计算消息 m 的密文 $c=g^m\cdot R^N \bmod N^2=(c_1\cdot c_2) \bmod N^2$.

Dec(). 密文 c 解密后的明文为 $m=L(c^\lambda \bmod N^2)\cdot\mu \bmod N$.

同态性: Paillier 密码系统具有加法同态性. 对于任意两个明文 $m_1,m_2\in Z_{N^2}^*$ 和 $R_1,R_2\in Z_{N^2}^*$, 对应的密文 c_1,c_2 满足: $c_1\cdot c_2=\mathrm{Enc}(m_1,R_1)\cdot\mathrm{Enc}(m_2,R_2)=g^{m_1+m_2}\cdot(R_1\cdot R_2)^N \bmod N^2=\mathrm{Enc}(m_1+m_2,R)$.

解密后得到 $\mathrm{Dec}(c_1,c_2)=\mathrm{Dec}(\mathrm{Enc}(m_1,R_1)\cdot\mathrm{Enc}(m_2,R_2))=(m_1+m_2) \bmod N^2$.

由此发现, 密文相乘等于明文相加.

3.5 秘密共享

Shamir 秘密共享 [10] 算法涉及两个过程: 密钥分发 SSS.Split 和密钥恢复 SSS.Rec.

SSS.Split. 秘密分发者将秘密值 $S\in Z_q$ 分为 n 个秘密子份额, 为每个参与者分发一个秘密子份额. 通过 $k-1$ 阶多项式 $\mathrm{EK}(x)=\sum_{j=0}^{k-1}(a_j\cdot x^j) \bmod q$ 计算每个子份额 $S_i=\mathrm{EK}(x_i)$, $1\leqslant i\leqslant n$, x_i 代表参与者 P_i 的序号.

SSS.Rec. 拥有 k 个或更多子份额的秘密重构者可以通过计算 $S=\mathrm{EK}(0)=\sum_{i=1}^{k}(\lambda_i\cdot S_i)\bmod q$ 重构秘密值 S, 其中 $\lambda_i=\sum_{j=1,j\neq i}^{k}\frac{x-x_j}{x_i-x_j}\bmod q$.

3.6 支持批量验证的签名算法

我们使用支持批量验证技术的椭圆曲线数字签名算法 [30], 支持对不同签名者的多个签名进行验证, 与普通签名相比, 大大提高了验证效率. 该算法涉及 3 个算法, 具体如下.

ParaGen(). 签名者 S 选择一个大小为 κ_1 的质数 q_1 并随机选择 $a,b\in Z_{q_1}$ 作为方程 $y^2=x^3+ax+b \bmod q_1$ 的系数. 群 $G_{q_1}(a,b)$ 定义为 κ_2 位 n 阶方程上的点, O 是无穷远点, 并令 $G=(x_g,y_g)$ 为群 $G_{q_1}(a,b)$ 的生成元. 假设有 ℓ 位签名者, 每一位签名者 S_i 选择一个随机数 $d_i\in[1,n-1]$ 作为签名密钥并计算相应的公钥 $Q_i=d_i\cdot G$.

Sig(). 签名者 S_i 随机选择 $z_i\in[1,n-1]$ 并计算 $P_i=(x_i,\ y_i)=z_i\cdot G$ 和 $\tau_i=x_i \bmod n$, 其中 $\tau_i\neq 0$. V_i 计算 m_i 的哈希值 $e_i=H(m_i)$, $H\ (\cdot)$ 为 SHA-256 算法. 接着计算 $s_i=z_i^{-1}(e_i+\tau_i\cdot d_i) \bmod n$, $s_i\neq O$. 最后, 得到消息 m_i 的签名 $(m_i,\ P_i,\ s_i)$.

Ver(). 当验证者 V 收到签名 $(m_i,\ P_i,\ s_i)$, $i\in[1,\ \ell]$ 后, 首先检验 d_i 和 s_i 是否为区间 $[1,\ n-1]$ 内的整数. 计算 m_i 的哈希值 $e_i=H(m_i)$, $t_{1i}=e_is_i^{-1}$ 和 $t_{2i}=d_is_i^{-1}$, $i\in[1,n]$. 验证方程 $(\sum_{i=1}^{\ell}t_{1i})\cdot G+(\sum_{i=1}^{\ell}t_{2i})\cdot Q_i=\sum_{i=1}^{\ell}P_i$ 是否成立, 如果成立, 接受签名. 否则 V 拒绝该签名.

3.7 承诺方案

承诺方案是密码学的一个重要原型. 在我们提出的数据聚合方案中, 承诺方案与 Paillier 密码系统结合使用, 即承诺方案需要对 Paillier 密码系统生成的密文执行. 承诺方案在承诺方 C 和验证方 R 之间进行. 承诺方为一个敏感数据计算出相应的承诺, 验证方验证前后承诺是否一致.

Setup(). 承诺方 C 随机选择两个大质数 p 和 q, 计算 $N=pq$ 和欧拉函数 $\phi(N)=(p-1)(q-1)$. C 随机生成满足 $1<E<\phi(N)$ 和 $gcd(\phi(N),E)=1$ 的整数 E 以及 $g\in Z_{N^2}^*$, 并公开 (E,g,N).

GenComm(). 承诺方 C 随机选择 $r\in Z_{N^2}^*$, 计算敏感信息 m 的承诺值 $\mathrm{Com}=\mathrm{Comm}(m,r)=m^E\cdot g^r \bmod N^2$. 发送承诺值 Com 给相应的验证方 R.

OpenComm(). 根据随后公开的 m 和 r, 验证方 R 验证等式 $\mathrm{Com}=m^E\cdot g^r \bmod N^2$ 是否满足.

承诺方案具有乘法同态性, 对于任意两个敏感数据 m_1,m_2 的承诺值 Com_1,Com_2, 证明如下:

$$\begin{aligned}\mathrm{Com}_1,\mathrm{Com}_2 \bmod N^2 =&\mathrm{Comm}(m_1,r_1)\cdot\mathrm{Comm}(m_2,r_2)\\ =&(m_1\cdot m_2)^E\cdot g^{r_1+r_2} \bmod N^2=\mathrm{Comm}(m_1\cdot m_2,r_1+r_2).\end{aligned}$$

4 本文方案

4.1 系统初始化

我们列出了方案中使用的符号和相应的含义, 具体见表 1. 可信机构 TA 设置系统安全参数 κ_1, κ_2, 并生成系统公共参数以及不同实体的私钥, 具体操作如下.

• TA 选择两个大素数 p,q, 计算 $N=pq$ 和欧拉方程 $\phi(N)=(p-1)(q-1)$ 并满足 $\gcd(pq,\phi(N))=1$ 和 $|p|=|q|=\kappa_1$. 接着, TA 随机选择 $R,g\in Z_{N^2}^*$ 满足 $\gcd(L(R^\lambda \bmod n^2),N)=1$. 通过 Paillier 密码系统的密钥生成算法 $\mathrm{Gen}(k)$ 得到参数 (N,R), 并且随机选择一个满足 $1<E<\phi(N)$ 和 $\gcd(\phi(N),E)=1$ 的整数 E.

表 1　符号表

Table 1　Notations

Notations	Descriptions
Dps	Data providers
AGA	Aggregator
$\mathrm{ID}_{\mathrm{DP}_i}$	i-th data provider
$\mathrm{ID}_{\mathrm{AG}}$	Aggregator
GOV	Government
BC	Blockchain ledger
TA	Trust athority
κ_1, κ_2	Two system security parameters
$G_{q_0}(a,b)$	Elliptic curve group
n	Order of the group $G_{q_0}(a,b)$
G	Generator of the group $G_{q_0}(a,b)$
H_1, H_2	Two one-way hash function
$\{\overline{\lambda_i}\}_j$	j-th share of the private key
σ_{DP_i}	$\mathrm{ID}_{\mathrm{DP}_i}$'s signature
σ_{AG}	$\mathrm{ID}_{\mathrm{AG}}$'s signature

• 基于第 3 节提出的支持批量验证的签名算法, 可信机构 TA 首先随机选择一个质数 q_0 和由方程 $y^2 = x^3 + ax + b \bmod q_0$ 定义的椭圆曲线 $G_{q_0}(a,b)$, 其中 $a, b \in Z_{q_0}$ 以及 $4a^3 + 27b^2 \neq 0$. 接着, TA 选择 n 阶群 $G_{q_0}(a,b)$ 的生成元为 G. 最后 TA 选择一个质数 p_0 并设置两个哈希函数: $H_1 : \{0,1\}^* \to Z_{p_0}^*$ 和 $H_2 : \{0,1\}^* \to \{0,1\}^*$, 并且 $|p_0| = \kappa_2$.

• 本方案涉及 $\widehat{n}$ 个 DPs, TA 选择一组质数 $\xi_1, \xi_2, \ldots, \xi_{\widehat{n}}$ 并设置正整数 $\mathbb{Z}$ 和 W 分别代表每个信息向量的上限和每个维度信息的权重. 为了避免数据溢出, 变量设置需满足 $\widehat{n} \cdot \mathbb{Z} \cdot W \leqslant \xi_i, \forall 1 \leqslant i \leqslant \widehat{n}$. TA 计算唯一解模 $\Psi = \xi_1 \times \xi_2 \times \cdots \times \xi_{\widehat{n}}$, 并计算 $\Psi_i = \Psi / \xi_i$ 满足 $\Psi_i \cdot {\Psi_i}^{-1} \equiv 1 (\bmod \xi_i)$, 并令 $\overline{\mathfrak{A}_i} = \Psi_i \cdot {\Psi_i}^{-1}$.

• TA 通过调用 SSS.Split 将每个 DPs 的密钥分为 $\widehat{n}$ 份 $\{\overline{\lambda_i}\}_j, 1 \leqslant j \leqslant \widehat{n}$, 得到 $(\widehat{n}, \{\overline{\lambda_i}\}_j)$. GOV 收集到 k 个份额就可以恢复密钥.

最后, TA 公开系统公钥 $\mathrm{PP} = \{N, g, E, R, G, n, a, b, q_0, \xi_1, \xi_2, \ldots, \xi_{\widehat{n}}, \mathbb{Z}, W, \Psi, \overline{\mathfrak{A}_i}\}$.

4.2　生成用户报告

在一段时间 T_{DP_i} 内, 每个 DPs 的身份标记为 $\mathrm{ID}_{\mathrm{DP}_i}$, 对传染病数据报告进行加密, 并生成相应的承诺. 具体操作如下.

• $\mathrm{ID}_{\mathrm{DP}_i}$ 使用 CRT 将传染病数据报告 $(\widehat{m}_i) = (m_i^{(1)}, m_i^{(2)}, \ldots, m_i^{(\widehat{n})})$ 转换为一个大整数 m_i, 即

$$\widehat{m}_i = m_i^{(1)} \cdot \overline{\mathfrak{A}_1} + m_i^{(2)} \cdot \overline{\mathfrak{A}_2} + \cdots + m_i^{(\widehat{n})} \cdot \overline{\mathfrak{A}_{\widehat{n}}} (\bmod \psi). \tag{1}$$

• $\mathrm{ID}_{\mathrm{DP}_i}$ 分别计算 p^2 和 q^2 的模逆元 $\mu_1 = (p^2)^{-1} \bmod p^2$ 和 $\mu_2 = (q^2)^{-1} \bmod q^2$, 令 $\ell_1 = \mu_1 \cdot p^2$, $\ell_2 = \mu_2 \cdot q^2$. 接着计算 $\{\overline{k_i}\}_1 = g^{m_i} \bmod p^2$, $\{\overline{k_i}\}_2 = g^{m_i} \bmod q^2$, $\{\overline{k_i}\}_3 = R^N \bmod p^2$ 和 $\{\overline{k_i}\}_4 = R^N \bmod q^2$.

• $\mathrm{ID}_{\mathrm{DP}_i}$ 通过使用 CRT 计算 $g^{m_i} \bmod N^2$ 和 $R^N \bmod N^2$, 分别标记为 $\{c_i\}_1$ 和 $\{c_i\}_2$, 其中 $\{c_i\}_1 = g^{m_i} \bmod N^2 = (\{\overline{k_i}\}_1 \cdot \ell_1 + \{\overline{k_i}\}_2 \cdot \ell_2) \bmod N^2$, $\{c_i\}_2 = R^N \bmod N^2 = (\{\overline{k_i}\}_3 \cdot \ell_1 + \{\overline{k_i}\}_4 \cdot \ell_2) \bmod N^2$.

根据模运算法则, 我们可以得到密文

$$c_{\mathrm{DP}_i} = (\{c_i\}_1 \cdot \{c_i\}_2) \bmod N^2 = g^{m_i} \cdot R^N \bmod N^2. \tag{2}$$

• $\mathrm{ID}_{\mathrm{DP}_i}$ 随机选择签名密钥 $d_{\mathrm{DP}_i} \in [1, n-1]$ 并进一步计算对应的公钥 $Q_{\mathrm{DP}_i} = d_{\mathrm{DP}_i} \cdot G$. 接着, 计算签名

$$\sigma_{\mathrm{DP}_i} = (P_{\mathrm{DP}_i}, s_{\mathrm{DP}_i}) = \mathrm{sig}(d_{\mathrm{DP}_i}, H_1(\mathrm{ID}_{\mathrm{DP}_i} || c_{\mathrm{DP}_i} || T_{\mathrm{DP}_i})). \tag{3}$$

• $\mathrm{ID}_{\mathrm{DP}_i}$ 调用签名算法 GenComm() 为密文 c_{DP_i} 产生承诺 $\mathrm{Comm}_{\mathrm{DP_i}}$.

• $\mathrm{ID}_{\mathrm{DP}_i}$ 上传 $\{\mathrm{ID}_{\mathrm{DP}_i}, \sigma_{\mathrm{DP}_i}, c_{\mathrm{DP}_i}, T_{\mathrm{DP}_i}, \mathrm{Comm}_{\mathrm{DP_i}}\}$ 到区块链账本中存储.

4.3 数据聚合

在时间段 T_{AG} 中, 聚合机构 AGA 的身份标记为 $\mathrm{ID}_{\mathrm{AG}}$, 从区块链账本中获得数据提供者加密的传染病数据报告后, 使用批量验证技术来检验收到信息的合法性. 验证通过后, $\mathrm{ID}_{\mathrm{AG}}$ 对加密的传染病数据报告进行数据聚合, 从而提高了数据传输过程的效率.

• 收到 $\mathrm{ID}_{\mathrm{DP}_i}$ 的数据报告后, $\mathrm{ID}_{\mathrm{AG}}$ 首先检查时间戳 T_{EC_i} 的有效性, 使用 U_1 表示 u 个数据提供者集合. 调用算法 $\mathrm{Ver}(\cdot,\cdot)$ 验证签名 $\sigma_{\mathrm{DP}_i} = (P_{\mathrm{DP}_i}, s_{\mathrm{DP}_i})$, 即

$$\sum_{i=1}^{u} H(c_{\mathrm{DP}_i}) \cdot s_{\mathrm{DP}_i}^{-1} \cdot G + \sum_{i=1}^{u} \tau_{\mathrm{DP}_i} \cdot s_{\mathrm{DP}_i}^{-1} \cdot Q_{\mathrm{DP}_i} = \sum_{i=1}^{u} P_{\mathrm{DP}_i}, \tag{4}$$

其中 τ_{DP_i} 由 P_{DP_i} 的 x 坐标值模 N 得到.

• $\mathrm{ID}_{\mathrm{AG}}$ 对收到的密文执行聚合操作

$$c_A = \prod_{i=1}^{u} c_{\mathrm{DP}_i}. \tag{5}$$

• $\mathrm{ID}_{\mathrm{AG}}$ 选择签名密钥 $d_{\mathrm{AG}} \in [1, n-1]$ 对聚合密文数据进行签名

$$\sigma_{\mathrm{AG}} = (P_{\mathrm{AG}}, s_{\mathrm{AG}}) = \mathrm{sig}(d_{\mathrm{AG}}, H_1(\mathrm{ID}_{\mathrm{AG}} || c_{\mathrm{AG}} || T_{\mathrm{AG}})). \tag{6}$$

• $\mathrm{ID}_{\mathrm{AG}}$ 将 $\{\mathrm{ID}_{\mathrm{AG}}, \sigma_{\mathrm{AG}}, c_{\mathrm{AG}}, T_{\mathrm{AG}}, \mathrm{Comm}_{\mathrm{DP}_i}\}$ 存储在区块链账本中.

4.4 数据解密和验证

经过时间 T 后, GOV 从区块链账本中提取 v ($v \geqslant k$) 个相应的密文后, 首先检验时间戳 T_{AG} 是否合法. 验证签名 $\sigma_{\mathrm{AG}} = (P_{\mathrm{AG}}, s_{\mathrm{AG}})$ 是否成立, 即 $\sum_{\mathrm{AG}=1}^{v} H(c_{\mathrm{AG}}) \cdot s_{\mathrm{AG}}^{-1} \cdot G + \sum_{\mathrm{AG}=1}^{v} \tau_{\mathrm{AG}} \cdot s_{\mathrm{AG}}^{-1} \cdot Q_{\mathrm{AG}} = \sum_{\mathrm{AG}=1}^{v} P_{\mathrm{AG}}$. r_{AG} 由 P_{AG} 的 x 坐标值模 N 得到. 若验证通过, 则继续进行后续操作.

• GOV 通过等式

$$\prod_{i \in U_1} \mathrm{Comm}_{\mathrm{EC}_i} = C_{\mathrm{EC}_i}^{E} \cdot g^{\sum_{i \in U_1} r_{\mathrm{EC}_i}}, \tag{7}$$

验证每组 $i \in U_1$ 承诺是否成立. 如果等式成立, 则证明 $\mathrm{ID}_{\mathrm{AG}}$ 的聚合操作是合法的.

• GOV 使用 SSS.Rec 恢复密钥并解密

$$M = \mathrm{Dec}(c_A) = L\left(c^{\sum_{j=1}^{k} \{\overline{\lambda_i}\}_j \cdot y_i} \bmod N^2\right) / \mu \bmod N, \tag{8}$$

其中, $y_i = \sum_{w=1, w \neq j}^{k} \frac{x - x_w}{x_j - x_w} \bmod q$ 和 $\mu = (L(g^{\lambda} \bmod N^2))^{-1} \bmod N$.

- 最终, 通过

$$M_j = M \bmod q, \tag{9}$$

恢复每个维度的数据后, GOV 可获得每个维度的聚合数据 $\{M_1, M_2, \ldots, M_{\hat{n}}\}$.

5 正确性分析

$\mathrm{ID_{AG}}$ 从区块链账本中获取到传染病数据报告 $\{\mathrm{ID_{DP}}_i, \sigma_{\mathrm{DP}_i}, c_{\mathrm{DP}_i}, T_{\mathrm{DP}_i}, \mathrm{Comm_{DP}}_i\}$, 其中 $\sigma_{\mathrm{DP}_i} = (P_{\mathrm{DP}_i}, s_{\mathrm{DP}_i}) = \mathrm{sig}(d_{\mathrm{DP}_i}, H_1(\mathrm{ID_{DP}}_i || c_{\mathrm{DP}_i} || T_{\mathrm{DP}_i}))$, $s_{\mathrm{DP}_i} = Z_{\mathrm{DP}_i}^{-1}(H(c_{\mathrm{DP}_i}) + d_{\mathrm{DP}_i} \cdot \tau_{\mathrm{DP}_i}) \bmod n$, $P_{\mathrm{DP}_i} = Z_{\mathrm{DP}_i} \cdot G = (x_{\mathrm{DP}_i}, y_{\mathrm{DP}_i})$ 和 $\tau_{\mathrm{DP}_i} = x_{\mathrm{DP}_i} \bmod n$.

$$\begin{aligned}
&\sum_{i=1}^{u} H(c_{\mathrm{DP}_i}) \cdot s_{\mathrm{DP}_i}^{-1} \cdot G + \sum_{i=1}^{u} \tau_{\mathrm{DP}_i} \cdot s_{\mathrm{DP}_i}^{-1} \cdot Q_{\mathrm{DP}_i} \\
&= \left(\sum_{i=1}^{u} H(c_{\mathrm{DP}_i}) \cdot s_{\mathrm{DP}_i}^{-1} \right) \cdot G + \left(\sum_{i=1}^{u} \tau_{\mathrm{DP}_i} \cdot s_{\mathrm{DP}_i}^{-1} \cdot d_{\mathrm{DP}_i} \right) \cdot G \\
&= \left(\sum_{i=1}^{u} H(c_{\mathrm{DP}_i}) \cdot s_{\mathrm{DP}_i}^{-1} + \sum_{i=1}^{u} \tau_{\mathrm{DP}_i} \cdot s_{\mathrm{DP}_i}^{-1} \cdot d_{\mathrm{DP}_i} \right) \cdot G \\
&= \sum_{i=1}^{u} s_{\mathrm{DP}_i}^{-1} \cdot (H(c_{\mathrm{DP}_i}) + \tau_{\mathrm{DP}_i} \cdot d_{\mathrm{DP}_i}) \cdot G \\
&= \sum_{i=1}^{u} Z_{\mathrm{DP}_i} \cdot G \\
&= \sum_{i=1}^{u} P_{\mathrm{DP}_i}.
\end{aligned}$$

因此, $\mathrm{ID_{AG}}$ 的批量验证签名操作是正确的.

GOV 从区块链获取到数据报告 $\{\mathrm{ID_{AG}}, \sigma_{\mathrm{AG}}, c_{\mathrm{AG}}, T_{\mathrm{AG}}\}$, 其中 $\sigma_{\mathrm{AG}} = (P_{\mathrm{AG}}, s_{\mathrm{AG}}) = \mathrm{sig}(d_{\mathrm{AG}}, H_1(\mathrm{ID_{AG}} || c_{\mathrm{AG}} || T_{\mathrm{AG}}))$, $P_{\mathrm{AG}} = Z_{\mathrm{AG}} \cdot G = (x_{\mathrm{AG}}, y_{\mathrm{AG}})$, $s_{\mathrm{AG}} = Z_{\mathrm{AG}}^{-1}(H(c_{\mathrm{AG}}) + d_{\mathrm{AG}} \cdot \tau_{\mathrm{AG}}) \bmod n$ 和 $\tau_{\mathrm{AG}} = x_{\mathrm{AG}} \bmod n$.

$$\begin{aligned}
&\sum_{\mathrm{AG}=1}^{v} H(c_{\mathrm{AG}}) \cdot s_{\mathrm{AG}}^{-1} \cdot G + \sum_{\mathrm{AG}=1}^{v} \tau_{\mathrm{AG}} \cdot s_{\mathrm{AG}}^{-1} \cdot Q_{\mathrm{AG}} \\
&= \left(\sum_{\mathrm{AG}=1}^{v} H(c_{\mathrm{AG}}) \cdot s_{\mathrm{AG}}^{-1} \right) \cdot G + \left(\sum_{\mathrm{AG}=1}^{v} \tau_{\mathrm{AG}} \cdot s_{\mathrm{AG}}^{-1} \cdot d_{\mathrm{AG}} \right) \cdot G \\
&= \left(\sum_{\mathrm{AG}=1}^{v} H(c_{\mathrm{AG}}) \cdot s_{\mathrm{AG}}^{-1} + \sum_{\mathrm{AG}=1}^{v} \tau_{\mathrm{AG}} \cdot s_{\mathrm{AG}}^{-1} \cdot d_{\mathrm{AG}} \right) \cdot G \\
&= \sum_{\mathrm{AG}=1}^{v} s_{\mathrm{AG}}^{-1} \cdot (H(c_{\mathrm{AG}}) + \tau_{\mathrm{AG}} \cdot d_{\mathrm{AG}}) \cdot G \\
&= \sum_{i=1}^{v} Z_{\mathrm{AG}} \cdot G \\
&= \sum_{\mathrm{AG}=1}^{v} P_{\mathrm{AG}}.
\end{aligned}$$

因此, GOV 的批量验证签名操作是正确的.

下面的验证表明 GOV 可以正确地执行解密操作.

$$\begin{aligned}
\lambda &= \sum_{j=1}^{k} \{\overline{\lambda_i}\}_j \cdot y_i \\
&= \sum_{j=1}^{k} \{\overline{\lambda_i}\}_j \cdot \sum_{w=1,w\neq j}^{k} \frac{x - x_w}{x_j - x_w} \bmod q, \\
M &= \mathrm{Dec}(c_A) \\
&= \frac{L(c^{\lambda} \bmod N^2)}{\mu} \bmod N \\
&= \frac{L((\prod_{i=1}^{u} g^{m_i} \cdot R_i^N)^{\lambda} \bmod N^2)}{L(g^{\lambda} \bmod N^2)} \bmod N \\
&= \frac{L(g^{\lambda \cdot \sum_{i=1}^{u} m_i} \cdot (\prod_{i=1}^{u} R_i)^{N\cdot\lambda} \bmod N^2)}{L(g^{\lambda} \bmod N^2)} \bmod N \\
&= \frac{L(g^{\lambda \cdot \sum_{i=1}^{u} m_i} \bmod N^2)}{L(g^{\lambda} \bmod N^2)} \bmod N \\
&= \frac{1 + \lambda \cdot N \cdot (\sum_{i=1}^{u} m_i) - 1}{1 + \lambda \cdot N - 1} \bmod N \\
&= \sum_{i=1}^{u} m_i.
\end{aligned}$$

因此, GOV 的解密过程是正确的.

6 安全性分析

6.1 机密性

在本方案中, 我们利用 Paillier 密码系统来加密隐私数据, 并利用其加法同态性来聚合密文数据. 最初, 数据提供者在生成数据报告期间, 使用 Paillier 加密算法将自己的传染病数据报告进行加密, 传染病数据报告在后续的操作中, 一直都是以 Paillier 密文的形式存在.

Paillier 密码系统的安全性可以归约到复合剩余假设困难问题 [31], Paillier 加密系统满足选择明文攻击下的不可区分性. 这便意味着在多项式时间内, 可以保证没有敌手能够在没有解密密钥的情况下获得明文. 因此, 在数据聚合阶段, 当聚合者收到密文数据报告, 聚合者无法在无解密密钥的情况下恢复明文数据. 因此, 即便聚合者是半可信的, 传染病数据报告的机密性和隐私性仍然能够得到保障.

6.2 强隐私保护

在本方案中, 通过使用秘密共享技术将密钥 λ_i 分为密钥子份额 $\{\overline{\lambda_i}\}_j, 1 \leqslant j \leqslant \hat{n}$ 分别发给多个数据提供者, 这种门限机制可以让我们的方案抵抗合谋攻击. 当且仅当至少有门限 k 个数据提供者诚实的上传, GOV 才能最终成功恢复消息. 因此, 当 GOV 与至多 $k-1$ 个数据提供者合谋也不会泄露用户的隐私. 最坏的情况是, 如果 GOV 与 $k-2$ 个数据提供者合谋, 并且得到其 $k-2$ 个密钥子份额, 但其余两个密钥子份额依然有很大的不确定性. 因此, 如果数据提供者和 GOV 合谋, 隐私信息也不会泄露.

6.3 数据完整性和认证

本方案可以保障数据提供者的隐私数据和聚合者的聚合数据的完整性和数据来源.

对于数据完整性, 聚合者定期从区块链账本上取得数据报告 $\{\mathrm{ID}_{\mathrm{EC}_i}, \sigma_{\mathrm{EC}_i}, c_{\mathrm{EC}_i}, T_{\mathrm{EC}_i}\}$, 首先检查时间戳 T_{EC_i} 是否合法, 之后通过本文的批量验证技术对其完整性进行验证. 其中, 由于消息中的每个元素都会涉及到验证, 所以对消息的任何篡改都会导致签名验证失败. 如果没有签名密钥 d_{EC_i}, 敌手无法伪造有效签名. 其次, 得益于区块链的不可篡改性, 这一特性从根本上保障了数据的完整性和可靠性: 一旦交易记录被添加到区块链中, 就无法被修改或删除. 鉴于区块链的上述优势, 我们在本手稿中将加密过后的数据报告以及进行聚合处理过后的数据报告存储在区块链账本中, 数据一旦存放在区块链上, 依据区块链的特性可以保障链上数据的完整性和不可篡改性. 因此, 数据提供者和聚合者之间传输数据的完整性可以得到保障. 同样地, 聚合者和 GOV 之间的数据 $\{\mathrm{ID}_{\mathrm{AG}}, \sigma_{\mathrm{AG}}, c_{\mathrm{AG}}, T_{\mathrm{AG}}\}$ 传输过程中的完整性也能得到保障.

对于身份验证, 数据提供者和聚合者都使用各自的签名私钥 d_{EC_i} 和 d_{AG} 来分别生成签名 σ_{EC_i} 和 σ_{AG}. 敌手无法获得他们的签名私钥, 所以他们无法生成相应的签名并且也无法分辨消息来自哪里. 因此, 本文方案能够保障各个实体数据传输过程中的数据完整性和源认证安全.

7 性能评估

7.1 实验环境设置

本节评估本方案的计算和通信开销方面的性能. 我们使用的是型号为 Intel(R), Xeon(R) Gold 6133 的电脑, 在局域网下配置 Window 10 操作系统及 2.50 GHZ CPU 和 8 GB 内存. 我们使用 PBC 库来执行 Paillier 密码系统以及签名算法的有效性验证过程. 我们将从计算开销和通信开销两个方面对本方案进行评估.

7.2 计算开销分析

7.2.1 比较 Paillier 加密算法的效率

如图 2 和 3 所示, 我们设置明文长度固定为 64 位, 密钥长度分别取 256, 512, 768, 1024 和 2048 位时, 改进的 Paillier 加密算法的加密时间远小于原加密算法的加密时间, 且解密算法的解密时间远小于原解密算法的解密时间.

7.2.2 计算开销

为了更加清晰地分析本方案的通信性能, 我们将分别从用户端和聚合节点端分析各自的时间开销. 将用户数量分别设置为 10, 20, 30, 40, 50, 60, 70, 80, 90 和 100, 以便进行实验分析. 使用表 2 中所示的 4 个基于时间的度量标准来计算和度量本方案的计算开销, 表 3 展示了本方案与现有方案 [17,18] 计算开销的比较. Zhang 等 [18] 提出了一个支持区块链跨链计算分散一致性验证的方案, 利用监督链来检查跨链数据计算的正确性. Hu 等 [17] 设计了一种安全增强的数据共享方案, 利用同态加密在边缘节点聚合数据, 实现数据内容的隐私保护.

图 4 清晰直观地显示了本方案和方案 [17,18] 在用户端的时间开销. 从图 4 中可知, 方案 [17,18] 相较于本方案拥有较高的计算开销. 其中, 方案 [17] 的计算开销最高, 由于其使用了较多的时间开销

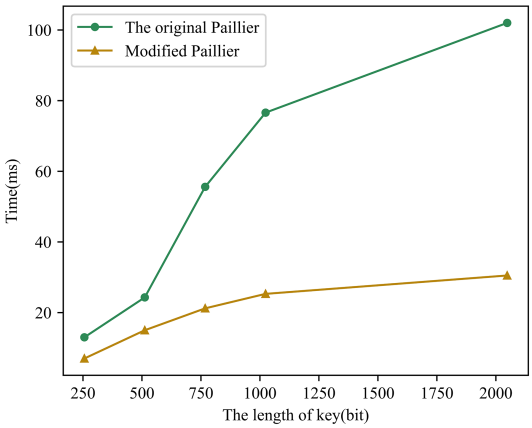

图 2 (网络版彩图) 基于 CRT 改进的算法和原始 Paillier 算法加密速度对比

Figure 2 (Color online) Comparison of the encryption speeds of the improved algorithms based on CRT and the original Paillier algorithm

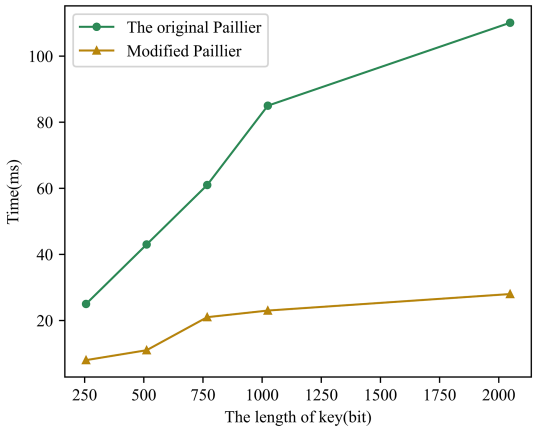

图 3 (网络版彩图) 基于 CRT 改进的算法和原始 Paillier 算法解密速度对比

Figure 3 (Color online) Comparison of the decryption speeds of the improved algorithms based on CRT and the original Paillier algorithm

表 2 时间度量标准的表示

Table 2 Notations of execution time

Notation	Description
T_p	A bilinear pairing operation
T_e	A modular exponentiation operation
T_m	A point multiplication operation
T_h	A hash function operation

表 3 计算开销对比

Table 3 Comparison of computational costs

Scheme	Users	Aggregators
[17]	$2T_e + 6T_m + 2T_h$	$2T_p + 3T_e + 4T_m + 2T_h$
[18]	$4T_e + 3T_m + T_h$	$2T_e + T_m + T_h$
Ours	$2T_e + 2T_m + T_h$	$4T_m + 2T_h$

大的计算操作, 如双线性映射、指数运算等. 本方案具有最优的计算性能, 由于本方案并没有使用时间开销大的计算操作.

从图 5 中能够看到本方案与方案 [17,18] 在聚合节点端的时间开销. 图中显示出方案 [17] 具有最高的时间开销, 方案 [18] 的时间开销略低于本方案. 虽然本方案的时间开销相对较高, 我们对聚合数据进行签名, 对传染病数据的完整性提供了更好的保障.

7.3 通信开销

为了评估本方案的通信开销, 本方案与现有方案 [17,18] 进行对比. 我们将具体的参数设置如下. 椭圆曲线群中的元素长度设置为 128 bytes. 支持批量验证的签名算法中群 $G_{q_0}(a,b)$ 中的元素长度为

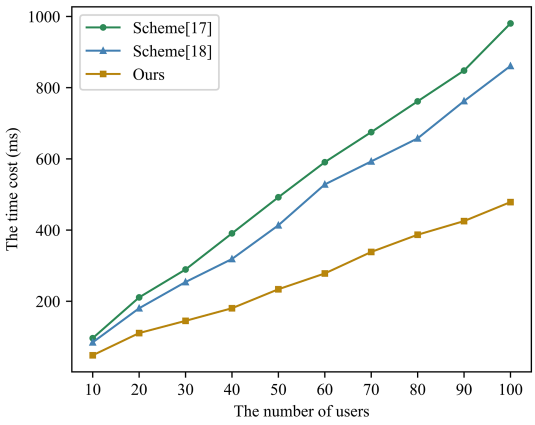

图 4 (网络版彩图) 用户端的计算开销

Figure 4 (Color online) Computational overhead on the user side

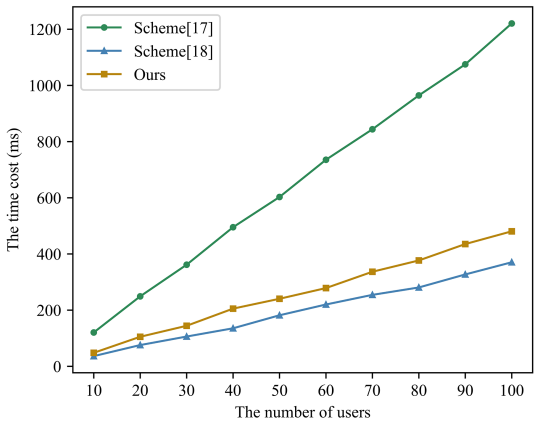

图 5 (网络版彩图) 聚合节点端的计算开销

Figure 5 (Color online) Computational overhead on the aggregator side

表 4 用户端通信开销 (字节)

Table 4 Communication cost comparison on the user side

Scheme	Communication cost (bytes)
[17]	390
[18]	294
Ours	294

32 bytes. Paillier 密文长度为 128 bytes. 哈希函数的长度设置为 32 bytes. 其余信息比如时间戳和身份的长度分别设置为 4 和 2 bytes.

在本方案的生成用户报告阶段, $\mathrm{ID}_{\mathrm{DP}_i}$ 上传 $\{\mathrm{ID}_{\mathrm{DP}_i}, \sigma_{\mathrm{DP}_i}, c_{\mathrm{DP}_i}, T_{\mathrm{DP}_i}, \mathrm{Comm}_{\mathrm{DP}_i}\}$ 到区块链账本中存储, 相应的通信开销为 294 bytes. 方案 [18] 中的数据计算阶段, user_s 将信息 $\{M, \{c_i\}_{i=1}^k, \{\mathrm{name}_i\}_{i=1}^k, \{\beta_i\}_{i=1}^k, h\}$ 存储在 C_{side} 中, 对应的通信开销为 294 bytes. 方案 [17] 在数据报告生成阶段将 $\mathrm{DR}_i = C_{\mathrm{PM}_i}||\mathrm{TS}||\mathrm{PM}_i||\mathrm{Sig}_{\mathrm{PM}_i}$ 传送给 R_j, 相应的通信开销为 390 bytes. 表 4 中展示了用户上传阶段的通信开销对比, 我们可以看出本方案具有优秀的通信开销.

在本方案的数据聚合阶段, 聚合者上传 $\{\mathrm{ID}_{\mathrm{AG}}, \sigma_{\mathrm{AG}}, c_{\mathrm{AG}}, T_{\mathrm{AG}}, \mathrm{Comm}_{\mathrm{DP}_i}\}$ 并存储在区块链账本中相应的通信开销为 294 bytes. 方案 [18] 中的数据写入阶段, C_{side} 将 C 发送给 C_{main}, 并将 $\{M, \{c_i\}_{i=1}^k, \{\mathrm{name}_i\}_{i=1}^k, \{\beta_i\}_{i=1}^k, h\}$ 发送给 SC, 相应的通信开销为 420 bytes. 方案 [17] 在数据报告聚合阶段, R_j 将 $\mathrm{DR}_j = \mathrm{ID}_{R_j}||C||\mathrm{TS}||\mathrm{Sig}_{R_i}$ 发送给 TMC, 对应的通信开销为 390 bytes. 表 5 中展示了聚合阶段的通信开销对比, 其中, 方案 [17,18] 都具有比本方案更高的计算开销.

8 结论

本文提出了一个具有容错机制的轻量级可验证隐私保护传染病监测数据聚合方案. 具体来说, 使用基于 CRT 改进的 Paillier 同态加密系统对传染病数据进行高效的加密操作, 保护数据传输过程中的数据隐私. 支持批量验证的签名算法传输中的数据进行签名, 以保护数据传输过程中的数据完整性. 考虑到聚合节点并不完全可信, 使用承诺机制解决聚合节点不可信的问题. 即使某些用户和聚合节点

表 5 聚合节点端通信开销 (字节)

Table 5 Communication cost comparison on the aggregator side

Scheme	Communication cost (bytes)
[17]	390
[18]	420
Ours	294

没有按时地上传数据, 聚合工作依然能够继续. 安全性分析证明本方案满足机密性、强隐私保护和数据完整性认证. 仿真实验证明本方案具有优秀的计算和通信开销, 可以安全有效地应用于传染病检测系统. 未来我们将继续研究数据聚合工作, 构建一个不需要第三方可信机构的数据聚合隐私保护方案.

参考文献

1 Fang K N, Ren R, Zhu J P, et al. Communicable disease forecasting and policy evaluation based on a dynamic seir model. J Manage Sci, 2022, 25: 114–126 [方匡南, 任蕊, 朱建平, 等. 基于动态 SEIR 模型的传染性疾病预测和政策评估. 管理科学学报, 2022, 25: 114–126]

2 Hu B J, Li Y C, Fang F, et al. Lightweight-blockchain based privacy-preserving data aggregation for epidemic disease surveillance. Sci Sin Inform, 2021, 51: 1885–1899 [胡柏吉, 李元诚, 房方, 等. 基于轻量级区块链的隐私保护传染病监测数据聚合. 中国科学信息科学, 2021, 51: 1885–1899]

3 Huang S. Dynamic analysis of an SEIRS model with nonlinear infectivity on complex networks. Int J Biomath, 2016, 09: 1650009

4 Xun Y, Russell P, Elisa B. Homomorphic Encryption and Applications. Berlin: Springer, 2014

5 Han S, Zhao S, Li Q, et al. PPM-HDA: privacy-preserving and multifunctional health data aggregation with fault tolerance. IEEE TransInformForensic Secur, 2016, 11: 1940–1955

6 He D, Kumar N, Zeadally S, et al. Efficient and privacy-preserving data aggregation scheme for smart grid against internal adversaries. IEEE Trans Smart Grid, 2017, 8: 2411–2419

7 Liu Y, Guo W, Fan C I, et al. A practical privacy-preserving data aggregation (3PDA) scheme for smart grid. IEEE Trans Ind Inf, 2019, 15: 1767–1774

8 Li S, Xue K, Yang Q, et al. PPMA: privacy-preserving multisubset data aggregation in smart grid. IEEE Trans Ind Inf, 2018, 14: 462–471

9 Tang W, Ren J, Deng K, et al. Secure data aggregation of lightweight E-healthcare IoT devices with fair incentives. IEEE Int Things J, 2019, 6: 8714–8726

10 Shamir A. How to share a secret. Commun ACM, 1979, 22: 612–613

11 Chen Y, Martinez-Ortega J F, Castillejo P, et al. A homomorphic-based multiple data aggregation scheme for smart grid. IEEE Sens J, 2019, 19: 3921–3929

12 Zhang X, Huang C, Zhang Y, et al. Enabling verifiable privacy-preserving multi-type data aggregation in smart grids. IEEE Trans Dependable Secure Comput, 2022, 19: 4225–4239

13 Boneh D, Lynn B, Shacham H, Short signatures from the weil pairing. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2001. 514–532

14 Peng C, Luo M, Wang H, et al. An efficient privacy-preserving aggregation scheme for multidimensional data in IoT. IEEE Int Things J, 2022, 9: 589–600

15 Shang S, Li X, Gu K, et al. A robust privacy-preserving data aggregation scheme for edge-supported IIoT. IEEE Trans Ind Inf, 2024, 20: 4305–4316

16 Zhang X, Huang C, Gu D, et al. Privacy-preserving statistical analysis over multi-dimensional aggregated data in edge computing-based smart grid systems. J Syst Architecture, 2022, 127: 102508

17 Hu P, Chu X, Zuo K, et al. Security-enhanced data sharing scheme with location privacy preservation for Internet of vehicles. IEEE Trans Veh Technol, 2024, 73: 13751–13764

18 Zhang Y, Jiang J, Dong X, et al. BeDCV: blockchain-enabled decentralized consistency verification for cross-chain

calculation. IEEE Trans Cloud Comput, 2023, 11: 2273–2284
19 Zhao J X, Su M, Hou J P, et al. A verifiable federated learning scheme based on homomorphic signatures. J Cryptol, 2023, 10: 1019–1034 [赵家雪, 苏铠, 侯金鹏, 等. 一种基于同态签名的可验证联邦学习方案. 密码学报, 2023, 10: 1019–1034]
20 Zhang W, Liu S, Xia Z. A distributed privacy-preserving data aggregation scheme for smart grid with fine-grained access control. J Inf Security Appl, 2022, 66: 103118
21 Wang S, Jin T, Xiao G W, et al. Efficient privacy-preserving secure aggregation scheme for federated learning. Comput Syst Appl, 2023, 32: 175–181 [王珊, 荆桃, 肖淦文, 等. 联邦学习下高效的隐私保护安全聚合方案. 计算机系统应用, 2023, 32: 175–181]
22 Mei X, Wang L, Qin B, et al. EFTA: an efficient and fault-tolerant data aggregation scheme without TTP in smart grid. Comput J, 2024, 67: 2368–2378
23 Zhang J, Wei J. PFDAM: privacy-preserving fine-grained data aggregation scheme supporting multifunctionality in smart grid. IEEE Int Things J, 2024, 11: 25520–25533
24 Wang Z, Zhang F, Zhang A, et al. LFTDA: a lightweight and fault-tolerant data aggregation scheme with privacy-enhanced property in fog-assisted smart grid. Comput Commun, 2024, 220: 35–42
25 Xu Z, Zhang R, Liang W, et al. A privacy-preserving data aggregation protocol for Internet of vehicles with federated learning. IEEE Trans Intell Veh, early access. doi: 10.1109/TIV.2024.3411313
26 Wu L, Fu S, Luo Y, et al. A robust and lightweight privacy-preserving data aggregation scheme for smart grid. IEEE Trans Dependable Secure Comput, 2024, 21: 270–283
27 Gope P, Sikdar B. Lightweight and privacy-friendly spatial data aggregation for secure power supply and demand management in smart grids. IEEE TransInformForensic Secur, 2019, 14: 1554–1566
28 Ding Y, Wang B, Wang Y, et al. Secure metering data aggregation with batch verification in industrial smart grid. IEEE Trans Ind Inf, 2020, 16: 6607–6616
29 Zhang J, Zhao Y, Wu J, et al. LVPDA: a lightweight and verifiable privacy-preserving data aggregation scheme for edge-enabled IoT. IEEE Int Things J, 2020, 7: 4016–4027
30 Kittur A S, Pais A R. A new batch verification scheme for ECDSA* signatures. Sādhanā, 2019, 44: 157
31 Paillier P. Public-key cryptosystems based on composite degree residuosity classes. In: Proceedings of International Conference on the Theory and Applications of Cryptographic Techniques, 1999. 223–238

Lightweight verifiable privacy-preserving infectious disease surveillance data aggregation scheme with fault tolerance

Xiaodong YANG[1], Lan YANG[1*], Lizhen WEI[1], Xiaoni DU[2] & Caifen WANG[3]

1. *College of Computer Science and Engineering, Northwest Normal University, Lanzhou* 730070*, China*;
2. *College of Mathematics and Statistics, Northwest Normal University, Lanzhou* 730070*, China*;
3. *College of Big Data and Internet, Shenzhen Technology University, Shenzhen* 518118*, China*
* Corresponding author. E-mail: 15389057097@163.com, 2022212139@nwnu.edu.cn

Abstract With frequent outbreaks of various epidemic infectious diseases across the globe, infectious disease surveillance plays a vital role in stopping the spread of infectious diseases. Privacy-preserving data aggregation is often used to avoid user privacy leakage caused by the transmission of infectious disease data. However, existing data aggregation schemes still have some security problems, such as untrusted aggregation nodes. To solve above problems, we propose a lightweight verifiable privacy-preserving infectious disease surveillance data aggregation scheme with fault tolerance. First, the improved Paillier homomorphic algorithm based on CRT and the signature algorithm with batch verification are used to efficiently encrypt and sign the infectious disease data to protect the data privacy and data integrity during data transmission. Second, the commitment mechanism is used to solve the problem of untrustworthiness of aggregate nodes. In addition, this scheme supports fault tolerance, and the aggregation work can continue even if some users and aggregation nodes do not upload data on time. In particular, this scheme can resist collusion attacks and meet higher security requirements. Since this scheme does not use time-consuming computational operations, such as bilinear mapping, simulation experiments show that the proposed scheme has excellent computational and communication overhead and can be safely and effectively applied to infectious disease surveillance systems.

Keywords infectious disease surveillance, data aggregation, privacy-preserving, homomorphic, lightweight