软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2022,33(11):4285−4304 [doi: 10.13328/j.cnki.jos.006343]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

可验证医疗密态数据聚合与统计分析方案*

张晓均[1,2], 张经伟[1], 黄 超[1], 谷大武[2], 张 源[3]

[1](西南石油大学 计算机科学学院 网络空间安全研究中心, 四川 成都 610500)

[2](上海交通大学 网络空间安全学院, 上海 200240)

[3](电子科技大学 计算机科学与工程学院, 四川 成都 611731)

通信作者: 谷大武, E-mail: dwgu@sjtu.edu.cn; 张晓均, E-mail: zhangxjdzkd2012@163.com

摘 要: 随着移动通信网络的飞速发展, 越来越多的可穿戴设备通过移动终端接入网络并上传大量医疗数据, 这些医疗数据聚合后具有重要的医学统计分析与决策价值. 然而, 在医疗数据传输和聚合过程中会出现传输中断、信息泄露、数据篡改等问题. 为了解决这些安全与隐私问题, 同时支持高效而正确的医疗密态数据聚合与统计分析功能, 提出了基于移动边缘服务计算的具有容错机制的可验证医疗密态数据聚合方案. 该方案改进了 BGN 同态加密算法, 并结合 Shamir 秘密共享机制, 确保医疗数据机密性、密态数据的可容错聚合. 该方案提出了移动边缘服务计算辅助无线体域网的概念, 结合移动边缘计算和云计算, 实现海量医疗大数据实时处理与统计分析. 该方案通过边缘计算服务器和云服务器两层聚合, 提高聚合效率, 降低通信开销. 同时, 使用聚合签名技术实现医疗密态数据的批量验证功能, 进而保障其在传输与存储过程中的完整性. 性能比较与分析表明, 该方案在计算与通信开销方面都具备突出优势.

关键词: 医疗数据; 边缘服务计算; 密态数据聚合; 同态加密; 聚合签名

中图法分类号: TP311

中文引用格式: 张晓均, 张经伟, 黄超, 谷大武, 张源. 可验证医疗密态数据聚合与统计分析方案. 软件学报, 2022, 33(11): 4285−4304. http://www.jos.org.cn/1000-9825/6343.htm

英文引用格式: Zhang XJ, Zhang JW, Huang C, Gu DW, Zhang Y. Verifiable Encrypted Medical Data Aggregation and Statistical Analysis Scheme. Ruan Jian Xue Bao/Journal of Software, 2022, 33(11): 4285−4304 (in Chinese). http://www.jos.org.cn/1000-9825/6343.htm

Verifiable Encrypted Medical Data Aggregation and Statistical Analysis Scheme

ZHANG Xiao-Jun[1,2], ZHANG Jing-Wei[1], HUANG Chao[1], GU Da-Wu[2], ZHANG Yuan[3]

[1](Research Center for Cyber Security, School of Computer Science, Southwest Petroleum University, Chengdu 610500, China)

[2](School of Cyber Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

[3](School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: Due to the fast development of mobile communication networks, more and more wearable devices access the network through mobile terminals and produce massive data. These aggregated medical data have significant statistical analysis and decision making value. Nevertheless, there are emerge security and privacy issues (e.g., as transmission interruption, information leakage, and data tampering) in medical data transmission and aggregation process. To address those security issues and ensure accurate medical data aggregation and analysis, an efficient verifiable fault-tolerant medical data aggregation scheme is proposed based on mobile edge service computing. The scheme exploits a modified BGN homomorphic encryption algorithm, integrates Shamir secret sharing mechanism to ensure medical data confidentiality, fault tolerance of encrypted aggregated data, simultaneously. The concept of mobile edge-assisted service computing in wireless body area networks is proposed in the scheme. Combined with the advantages of mobile edge computing and cloud computing,

* 基金项目: 国家自然科学基金(U1636217, 61902327, 62072307, 62002050); 中国博士后科学基金(2020M681316); 四川省科技厅重点研发项目(2021YFG0158); 西南石油大学青年科技创新团队项目(2019CXTD05)

收稿时间: 2021-01-20; 修改时间: 2021-02-25; 采用时间: 2021-04-03

the real-time big data processing and statistical analysis of massive medical big data could be conducted. Through edge-level aggregation and cloud-level aggregation, the aggregation efficiency is improved and the communication overhead is reduced. Besides, the scheme designs an aggregate signature algorithm to conduct batch verification on medical encrypted data, and guarantee the integrity during transmission and storage process. The comprehensive performance evaluation demonstrates that the proposed scheme has outstanding advantages in terms of computational costs and communication overhead.

Key words: medical data; edge service computing; encrypted data aggregation; homomorphic encryption; aggregate signature

随着传感技术和微电子技术的发展, 越来越多的传感器通过物联网、无线体域网和车联网等技术接入到互联网[1,2], 构建出各种智能应用, 如智能电网、智能医疗和无人驾驶等[3−5]. 随着这些应用的出现, 传输在互联网上的数据也随之激增, 同时, 云计算技术在计算资源和存储资源方面可以提供有效的保障. 然而, 据统计:在不久的将来, 将有数十亿的智能设备接入互联网并产生大量的数据[6], 这些数据需要以安全有效的方式进行分析和处理. 特别地, 在智能医疗领域, 无线体域网的广泛应用要求智能设备具有低延迟、高速率、快速的数据访问能力, 用于实时数据处理、分析和决策. 因此, 单一的云计算架构无法满足这些需求, 迫切需要引入各种智能边缘计算设备[7,8], 辅助云计算实现海量医疗大数据实时处理与计算分析.

在医疗数据的传输过程中, 由于无线体域网采集的医疗数据涉及用户的隐私数据, 较为敏感. 在开放的无线网络传输过程中, 会采用数据加密技术来保障医疗数据机密性和用户隐私安全[9]. 由于解密密钥在某些特殊情况下可能因安全保护措施不够而泄露, 甚至可能会被敌手窃取, 从而解密单个用户的医疗密态数据, 对用户的隐私安全产生威胁. 此外, 在开放的无线网络环境中, 可能存在外部敌手对通信信道进行窃听, 拦截, 甚至替换、篡改用户传输的医疗数据, 导致医生使用错误数据产生临床误诊, 因此需要采用数字签名等技术保障密态数据传输的完整性.

事实上, 医疗数据加密传输, 将会丧失不同程度的数据可用性. 近年来, 许多学者基于同态加密算法设计了各种密态数据聚合协议[10−14]. 由于同态加密算法所具有的保持加法或者乘法运算的特性, 数据被加密后, 能够被高效地进行聚合, 同时, 医疗数据分析中心可以利用解密私钥直接对聚合密文进行解密, 得到一些核心统计指标, 为精确的诊断决策提供隐私保护的深度数据统计分析. 整个过程无需对单个用户的密态数据进行解密, 因此有效保护了用户隐私和数据机密性.

基于无线体域网的医疗数据可以为医疗数据分析中心提供重要的医疗信息挖掘和决策价值, 然而由于医疗数据通常在无线体域网中被使用各种方法进行加密处理以保障数据机密性和用户隐私安全, 导致密态数据经过聚合后, 医疗分析中心通过对聚合结果解密只能够获取有限的统计信息. 因此, 数据聚合必须在保障数据机密性、完整性和用户隐私安全这些安全需求的同时, 为医疗数据分析中心提供尽可能多的统计分析结果.

在某些情况下, 终端用户可能非常注重自己的医疗隐私数据, 并不愿意按照要求实时通过互联网分享自己的敏感医疗数据. 另一方面, 用户在传输自己加密医疗数据过程中, 可能因为网络传输问题, 或者恶意敌手的截断等行为, 导致医疗密态数据传输失败. 文献[15]提出了支持容错机制的医疗密态数据聚合方案, 但此方案采用的 Shamir 的秘密共享技术并不能达到最终聚合密态数据的容错特性.

本文针对医疗密态数据设计了一种支持边缘服务计算和完整性验证功能的聚合与统计分析方案. 本方案将移动边缘服务计算集成到传统的云计算框架, 提出移动边缘服务计算辅助无线体域网的概念, 改进 BGN 同态加密算法, 并结合 Shamir 的秘密共享技术确保数据传输过程中的机密性, 以及聚合密态数据的容错特性. 本方案设计了一种基于身份的聚合签名算法, 实现医疗密态数据在传输和存储过程中的完整性验证. 特别地, 本方案中的移动边缘服务计算辅助无线体域网分为 4 层结构: 第 1 层是部署在无线体域网中的传感器设备, 负责采集用户医疗数据, 然后对医疗数据进行加密盲化并签名上传; 第 2 层是部署在云计算网络边缘的边缘计算服务器, 负责收集本区域无线体域网上传的医疗数据, 然后进行聚合运算, 并去除盲化提交给云服务器; 第 3 层是云服务器, 负责聚合存储医疗数据, 并对医疗数据分析中心提供数据分析服务; 第 4 层是医疗数据分析中心, 负责对云服务器发送挑战,然后对其返回的聚合数据的完整性和正确性进行验证, 并用私钥解密聚合密文, 最后对解密数据进行统计分析.

本文的主要贡献如下:

(1) 移动边缘服务计算辅助无线体域网: 在移动边缘服务计算的第 2 层中, 边缘服务计算服务器所具备的计算、存储和通信资源远高于传感器设备, 并且在物理位置上比云服务器更靠近无线体域网. 因此, 与传统的云计算相比, 边缘计算大大减少了响应时间和能耗.

(2) 医疗数据机密性: 用户在使用医疗数据分析中心的公钥对医疗数据加密后, 需要使用边缘服务器分发的秘密参数对密文进行盲化运算, 生成最终密文值, 使得即使敌手获得医疗数据分析中心的私钥, 在没有得到边缘服务器的控制权和足够多的密文数据时, 也无法独立解密出用户的医疗数据.

(3) 密态数据统计分析多功能性: 本方案改进了 BGN 同态加密算法, 将原 BGN 密码体制扩展到双消息加密, 可以同时对医疗数据的两种形态进行加密, 在保障医疗数据机密性和用户隐私安全的基础上, 为医疗数据分析中心提供均值、方差等多功能统计分析结果.

(4) 支持传输容错机制: 考虑到有些用户不愿意上传敏感的医疗数据或者医疗数据在传输过程中被截断, 本方案在注册阶段要求边缘服务计算服务器采用 Shamir 秘密共享技术, 将秘密参数分发给本区域的各个用户. 这样, 用户使用秘密参数对加密后的医疗密态数据进行盲化后上传给边缘服务器, 边缘计算服务器只需要接收到超过一定样本数量的可验证密态数据信息, 就可以正确地聚合医疗密态数据并去除盲化.

(5) 批量验证: 为了实现医疗密态数据可批量验证的功能, 本方案设计了基于身份的聚合签名算法, 在用户上传医疗密态数据到边缘服务器、边缘服务器上传聚合密态数据到云服务器以及云服务器上传聚合密态数据到医疗数据分析中心的过程中, 发送方都需要对发送的数据、发送者的身份以及发送时间进行聚合签名再上传, 并且接收方在收到消息后, 根据签名值和密态数据以及发送方公钥执行批量验证, 从而可以得知密态数据在处理和传输过程中有没有遭到篡改、替换或销毁.

本文第 1 节总结现有的数据聚合与统计分析领域的研究现状, 对已有的加密聚合方案的优劣进行讨论与分析. 第 2 节介绍双线性对技术、BGN 同态加密算法和 Shamir 秘密共享技术的相关理论和概念, 并详细描述系统模型. 第 3 节给出新型 BGN 同态加密算法的设计、可验证医疗密态数据聚合与统计分析方案的具体实施步骤. 第 4 节对方案的正确性进行证明. 第 5 节对方案的安全性进行分析. 第 6 节将本设计方案与同类方案进行性能对比与分析. 第 7 节对全文进行总结.

1 相关工作

数据聚合技术为信息系统提供全面而准确的数据分析, 基于数据聚合技术的统计分析方法可有效提高系统的可靠性和准确性. 考虑到移动终端设备的多样性以及用户的隐私性, 用户的敏感数据往往需要加密传输, 特别是隐私保护要求更高的医疗系统、政务系统、国防系统. 因此, 开放式物联网环境中的密态数据聚合已成为当前学术界和产业界的研究热点. 近年来, 许多密态数据聚合方案已被提出. Lu 等人[16]采用 Paillier 同态加密体制, 提出了一种具有隐私保护的用户数据聚合系统, 同时采用批量验证技术降低了认证成本. Zhang 等人[17]开发了一种具有多级可信机构的框架, 并采用基于身份的聚合签名技术, 在车联网中实现了隐私保护的数据聚合. Kang 等人[18]利用联盟区块链和智能合约技术, 在车载边缘网络中实现了安全数据存储和授权共享. 此外, 他们还提出了一种新型数据共享方案, 以确保智能车辆之间的高质量数据共享. 最近, Wang 等人[19]提出了一种基于身份的智能电网数据聚合协议, 以抵御恶意篡改攻击.

为了减少网络延迟, 合理管理数据流, 保证网络安全运行, Li 等人[20]提出了基于移动边缘计算框架的聚合方案, 方案可以有效地保护数据隐私, 提高网络的数据传输能力. 然而该方案没有容错功能, 任意一个移动终端上传的数据无效, 都会导致数据聚合的失败. 事实上, 基于数据聚合的网络传输过程, 往往会出现一部分数据由于网络原因传输失败, 或者终端用户为了自己更高的隐私安全不愿意上传数据, 因此更实际的系统应该是当参与聚合的有效数据量大于等于门限值时, 就能完成正确的数据聚合, 实现后续的隐私保护的统计分析. Tang 等人[15]基于秘密共享技术提出了具有容错机制的轻量级电子医疗物联网设备的安全数据聚合方案,

而事实上, 该方案由于 BGN 密码算法的固有特性, 还不能真正做到数据传输容错机制. Chan 等人[21]提出了一种新的机制, 能够有效地支持数据动态连接和离开. 在不可信聚合器场景中, Benhamouda 等人[22]考虑让一个不可信的聚合器周期性地计算由一组用户贡献的密态数据的聚合值. 在移动边缘计算辅助无线体域网的环境下, 边缘服务器在计算能力和通信时延上的优势可以有效地降低云服务器的带宽压力[23,24], 同时, 对医疗数据进行聚合可以进一步节省边缘服务器与云服务器之间的通信开销和云服务器的计算资源. 为了保证医疗数据聚合的安全性, 用户的医疗数据在提交前需要加密, 边缘服务器应该能够以密态的形式聚合数据. 显然, 传统的加密算法无法实现这一功能, 而文献[25,26]中同态加密算法则允许我们对密态数据进行特殊的代数运算, 其结果与对明文进行相同的运算后再进行加密是一样的. 信息技术时代, 隐私越来越受到人们的关注[27,28]. 移动边缘计算辅助无线体域网虽然减少了通信开销, 但对于如何保护用户隐私仍然是一个挑战, 即终端用户通过无线体域网上传医疗数据, 而这些医疗数据可能会导致隐私泄露. 因此, 为实现医疗数据隐私保护统计分析, 数据聚合方案[29,30]应保证终端用户的隐私性, 即云服务器不能从聚合数据中检索到用户的医疗数据.

2 预备知识

2.1 双线性对技术

在本节中我们主要对双线性对技术的定义进行介绍.

定义 1. 对于 p 阶乘法循环群 G_1 和 p 阶乘法循环群 G_2 的双线性对映射关系: $e:G_1\times G_1\to G_2$, 应该满足以下 3 个性质.

(1) 双线性: 对任意两个元素$\forall g_1,g_2\in G_1$ 以及 $\forall a,b\in Z_p^*$, 使得 $e(g_1^a,g_2^b)=e(g_1,g_2)^{ab}$;

(2) 非退化性: 存在两个元素$\forall g_1,g_2\in G_1$ 使得 $e(g_1,g_2)\neq 1\in G_2$, 这里的 1 是 G_2 的单位元;

(3) 可计算性: $\forall g_1,g_2\in G_1$, 对任何的输入都能找到一个有效的算法来计算 $e(g_1,g_2)$.

2.2 BGN同态加密算法

BGN 同态加密算法[25]属于半全同态加密技术,具有加法同态和一次乘法同态的特性,其主要由以下 3 个子算法构成.

- 密钥生成: 根据安全参数κ, 合数阶双线性对映射产生函数 $\mathcal{G}(\kappa)$ 生成一个数组 $(n,\tilde{g},\tilde{G}_1,\tilde{G}_2,\tilde{e})$, 在数组中, n 是由两个长度为κ比特的大素数 p 和 q 的乘积构成的; $\tilde{G}_1$, $\tilde{G}_2$ 是两个合数阶循环群,其阶为 n; $\tilde{g}$ 是 $\tilde{G}_1$ 循环群的生成元; $\tilde{e}:\tilde{G}_1\times\tilde{G}_1\to\tilde{G}_2$ 为循环群 $\tilde{G}_1$ 和循环群 $\tilde{G}_2$ 的双线性对映射. 计算 $\tilde{G}_1$ 的 q 阶循环子群的生成元 $v=\tilde{g}^p$. 公钥是 $pk=(n,\tilde{G}_1,\tilde{G}_2,\tilde{e},\tilde{g},v)$, 私钥 $sk=q$.
- 加密: 假设明文 m 的空间是一个整数集$\{1,2,\ldots,T\}$, 其中, $T<p$. 随机选取一个数字 $r\in Z_n$, 计算 m 的密文:

$$c=Enc(m,r)=\tilde{g}^m v^r\in\tilde{G}_1.$$

- 解密: 使用私钥 q 对密文 $c=\tilde{g}^m v^r$ 进行解密运算: $c^q=(\tilde{g}^m v^r)=(\tilde{g}^q)^m$. 然后令 $\hat{g}=\tilde{g}^q$, 若要恢复 m 只需要计算以 $\hat{g}$ 为底 c^q 的离散对数. 由于 0≤m≤T, 因此使用 Pollard's lambda 方法[25]只需在 $o(\sqrt{T})$ 的时间复杂度内完成离散对数运算恢复出明文 m.

2.3 Shamir秘密共享技术

Shamir 秘密共享技术[31]是基于多项式的拉格朗日插值公式, 提出的一个门限秘密共享方案. 其主要思想是将一个秘密值拆分成若干个子秘密, 并将子秘密交由不同的成员保管, 只有当协作的成员数量达到门限值时才能恢复秘密. Shamir 秘密共享技术主要包括秘密共享和秘密恢复这两个算法.

- 秘密共享: 假设现在有一个秘密值λ要与 n 个用户进行分享, 首先确定一个 $k-1$ 次多项式 $EK(x)=\lambda+a_1x+a_2x^2+\ldots+a_{k-1}x^{k-1}$, 其中, $a_1,a_2,\ldots,a_{k-1}\in Z_q$是对应的多项式系数, $k<n$. 然后计算$\lambda_j=EK(x_j)$, 其中, $x_j=1$,

2,…,n, 并把λ_j作为子秘密分发给第 j 个用户.

- 秘密恢复: 在恢复秘密λ时, 至少需要 k 个成员参与并用各自的子秘密一起构建拉格朗日插值公式才可以恢复秘密 λ. 假设$\{(x_1,\lambda_1),(x_2,\lambda_2),\ldots,(x_k,\lambda_k)\}$是任意 k 个成员的子秘密, 重构如下公式:

$$EK(x)=\sum_{z=1}^{k}\lambda_z\prod_{\substack{j=1\\ j\neq z}}^{k}\frac{x-x_j}{x_z-x_j}.$$

因此, 当 x=0 时, $EK(0)=\lambda$. 即恢复出原始秘密值λ.

2.4 系统模型

支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案的系统模型主要包含 5 类通信实体: 可信中心、用户、边缘服务器、云服务器以及医疗数据分析中心, 如图 1 所示.

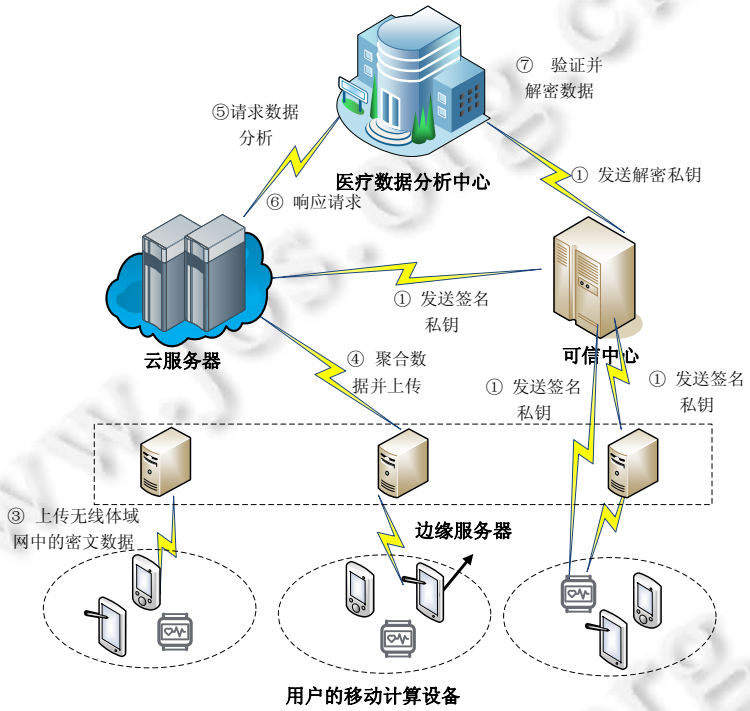

图 1 支持边缘辅助计算的医疗密态数据聚合和统计分析系统模型

- 可信中心: 负责产生整个系统所需的密码参数, 并为各个通信实体计算公私钥对, 最后发布需要公开的参数.
- 用户: 使用可穿戴设备采集自身的生命体征数据, 并通过无线体域网将采集的数据传输到移动计算设备中进行加密、盲化和数字签名. 最后再将这些密态医疗数据和签名值上传到所在区域的边缘计算服务器上.
- 边缘服务器: 边缘服务器是部署在特定区域内的服务器, 其为该区域用户提供计算服务并且由云服务器进行维护. 由于边缘服务器所具备的计算、存储和通信资源远高于用户设备, 并且在物理位置上比云服务器更靠近用户, 因此边缘服务器能够有效地降低用户端的计算开销和云服务器的带宽压力. 主要功能是聚合该区域用户传输的密态数据并去除盲化, 然后在有效周期内上传可验证的边缘级聚合密态数据信息到云服务器进行长期存储.
- 云服务器: 主要用于存储和验证边缘服务器上载的边缘级聚合密态数据信息, 并向医疗数据分析中心提供数据支持. 当医疗数据分析中心向其发送挑战请求时, 云服务器会对挑战的边缘级聚合密态

数据信息进行云级聚合运算和数字签名, 并向医疗数据分析中心返回所有的计算结果.

- 医疗数据分析中心: 在接收到云级聚合密态数据后, 首先对其进行完整性验证, 然后使用私钥对密态数据进行解密, 获得不同区域用户生命体征数据的聚合值、均值和方差. 最后, 医疗数据分析中心对这些医疗数据进行进一步隐私保护地统计分析.

3 支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案

为了实现医疗密态数据聚合以及均值、方差等多种统计分析功能, 本节首先需要改进和设计新型 BGN 同态加密算法. 之后, 本节将移动边缘服务计算集成到云计算架构之中, 进一步详细介绍支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案.

3.1 新型BGN同态加密算法

本方案中, 我们将原 BGN 密码体制扩展到双消息加密, 即用 4 个素数构造群的合数阶 n, 并引入两个不同的解密私钥对两类明文进行解密. 改进的新型 BGN 同态加密算法的细节如下.

- 密钥生成: 根据安全参数 κ, 合数阶双线性对映射产生函数 $\mathcal{G}(\kappa)$ 生成一个数组 $(n,\tilde{g},\tilde{G}_1,\tilde{G}_2,\tilde{e})$, 其中, $n=q_0q_1q_2q_3$, q_0, q_1, q_2, q_3 是长度为 κ 比特的大素数; $\tilde{G}_1$, $\tilde{G}_2$ 是两个合数阶循环群, 其阶为 n; $\tilde{g}$ 是循环群 $\tilde{G}_1$ 的生成元, $\tilde{e}:\tilde{G}_1\times\tilde{G}_1\to\tilde{G}_2$ 是循环群 $\tilde{G}_1$ 和循环群 $\tilde{G}_2$ 的双线性对映射. 设置 $v_1=\tilde{g}^{q_1q_2}$, $v_2=\tilde{g}^{q_0q_2}$, $v_3=\tilde{g}^{q_0q_1}$, 其中, v_1 是 $\tilde{G}_1$ 的 q_0q_3 阶循环子群的生成元, v_2 是 $\tilde{G}_1$ 的 q_1q_3 阶循环子群的生成元, v_3 是 $\tilde{G}_1$ 的 q_2q_3 阶循环子群的生成元. 公钥是 $pk=(n,\tilde{G}_1,\tilde{G}_2,\tilde{e},\tilde{g},v_1,v_2,v_3)$, 私钥 $sk_1=q_1q_2q_3$, $sk_2=q_0q_2q_3$.
- 加密: 对于两个需要加密的明文 $m_1\in[0,T_1]$和 $m_2\in[0,T_2]$, 其中, $T_1<q_0q_3$, $T_2<q_1q_3$. 选择一个随机数 $r\in Z_n$, 然后计算密文 $c=E(\{m_1,m_2\},r)=v_1^{m_1}v_2^{m_2}v_3^r\in\tilde{G}_1$.
- 解密: 给定密文 $c=v_1^{m_1}v_2^{m_2}v_3^r$, 对应的明文 m_1 和 m_2 可以分别通过私钥 sk_1 和 sk_2 解密. 计算 $c^{sk_1}=(v_1^{m_1}v_2^{m_2}v_3^r)^{sk_1}=(v_1^{m_1}v_2^{m_2}v_3^r)^{q_1q_2q_3}=(v_1^{q_1q_2q_3})^{m_1}$. 然后根据文献[25]中 Pollard 的 lambda 解密方法计算以 $\tilde{g}^{q_1^2q_2^2q_3}$ 为底 c^{sk_1} 的离散对数, 即可恢复 m_1. 然后计算 $c^{sk_2}=(v_1^{m_1}v_2^{m_2}v_3^r)^{sk_2}=(v_1^{m_1}v_2^{m_2}v_3^r)^{q_0q_2q_3}=(v_2^{q_0q_2q_3})^{m_2}$, 同理, 根据 Pollard 的 lambda 解密方法计算以 $\tilde{g}^{q_0^2q_2^2q_3}$ 为底 c^{sk_2} 的离散对数, 即可恢复 m_2.

新型 BGN 同态加密算法的加法同态性质描述如下.

对于 4 个消息 $m_1,m_2\in[0,T_1]$, $m_3,m_4\in[0,T_2]$:

$$E(\{m_1,m_3\},r_1)\cdot E(\{m_2,m_4\},r_2)=v_1^{m_1}v_2^{m_3}v_3^{r_1}\cdot v_1^{m_2}v_2^{m_4}v_3^{r_2}=v_1^{m_1+m_2}v_2^{m_3+m_4}v_3^{r_1+r_2}=E(\{m_1+m_2,m_3+m_4\},r_1+r_2).$$

3.2 方案具体步骤

支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案分为 7 个阶段: 系统初始化、系统注册、用户的医疗数据加密和签名上传、边缘计算服务器数据聚合去盲化、云服务器存储有效数据、云服务器数据聚合、可验证的聚合密态数据解密与统计分析. 其中, 密态数据的聚合、完整性验证和统计分析的流程如图 2 所示.

- 系统初始化阶段

此阶段由可信中心 TA 为各通信实体颁发私钥和秘密参数, 并发布系统公开参数.

(1) TA 设置一个合数阶 $n=q_0q_1q_2q_3$ 的双线性对映射 $\tilde{e}:\tilde{G}_1\times\tilde{G}_1\to\tilde{G}_2$, 其中, $\tilde{G}_1$, $\tilde{G}_2$ 为 n 阶乘法循环群. TA 选取 $\tilde{G}_1$ 的生成元 $\tilde{g}$, 分别计算 $v_1=\tilde{g}^{q_1q_2}$, $v_2=\tilde{g}^{q_0q_2}$, $v_3=\tilde{g}^{q_0q_1}$, $f=\tilde{g}^{q_1q_2q_3}$, 并计算 $sk_1=q_1q_2q_3$ 和 $sk_2=q_0q_2q_3$.

(2) TA 生成另一个非退化的双线性对映射 $e:G_1\times G_1\to G_2$, 其中, G_1, G_2 为具有相同素数阶 p 的乘法循环群, 选取 G_1 中的生成元 g_1. TA 设置两个抗碰撞的哈希函数 $H:\{0,1\}^*\to G_1$ 和 $h:\{0,1\}^*\to Z_p^*$.

(3) TA 为云服务器 ID_{PCC} 选取签名私钥 $u\leftarrow Z_p^*$, 计算其签名公钥 $U=g_1^u$. 对于每一个边缘计算服务器

ID_{ES_i} (i=1,2,…,N), TA 为 ID_{ES_i} 选取签名私钥 $u_i \leftarrow Z_p^*$, 计算对应的签名公钥 $U_i = g_1^{u_i}$. 同时, TA 为 ID_{ES_i}.

所辖区域的每一个用户 $ID_{PS_{i_j}}$ (j=1,2,…,η)选取签名私钥 $u_{i_j} \leftarrow Z_p^*$, 计算对应的签名公钥 $U_{i_j} = g_1^{u_{i_j}}$.

最后, TA 通过安全信道将解密私钥 $sk_1=q_1q_2q_3$ 和 $sk_2=q_0q_2q_3$ 发送给医疗数据分析中心 ID_{DAC}, 将签名私钥 u 发送给 ID_{PCC}, 将签名私钥 u_i 发送给对应的 ID_{ES_i} (i=1,2,…,n), 以及将签名私钥 u_{i_j} 发送给对应的用户 $ID_{PS_{i_j}}$ (j=1,2,…,η). TA 发布公开参数 $para_1 = (n, \tilde{G}_1, \tilde{G}_2, \tilde{e}, \tilde{g}, v_1, v_2, v_3, f)$ 以及公开参数:

$$para_2 = (g_1, G_1, G_2, e, H, h, U, \{U_i\}_{1<i\leqslant N}, \{U_{i_j}\}_{1<i\leqslant \eta}).$$

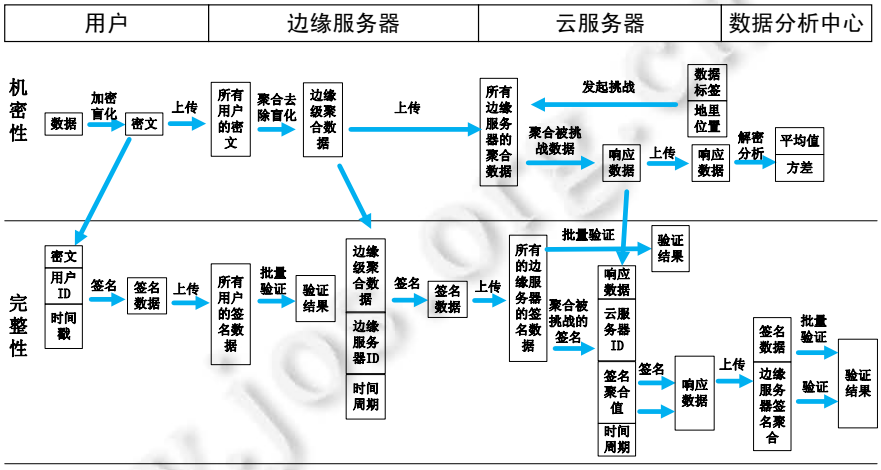

图 2 医疗密态数据聚合、完整性验证和统计分析流程图

- 系统注册阶段

当用户 $ID_{PS_{i_j}}$ 想要在对应的边缘计算服务器 ID_{ES_i} 注册时, $ID_{PS_{i_j}}$ 选取随机数 $v_{i_j} \leftarrow Z_p^*$, 计算数字签名 $sig_{i_j} = (sig_{i_j,1}, sig_{i_j,2}) = \left(g_1^{v_{i_j}}, H(ID_{ES_i})^{h(ID_{PS_{i_j}}\|t_{reg})v_{i_j}+u_{i_j}}\right)$, 其中, t_{reg} 是当前时间戳.

当收到身份注册信息 $(ID_{PS_{i_j}}, sig_{i_j}, t_{reg})$, ID_{ES_i} 检测时间戳 t_{reg} 是否失效, 验证以下方程:

$$e(sig_{i_j,2}, g_1) = e\left(H(ID_{ES_i}), sig_{i_j,1}^{h(ID_{PS_{i_j}}\|t_{reg})} U_{i_j}\right) \tag{1}$$

如果验证方程通过, 边缘计算服务器 ID_{ES_i} 采用 Shamir 秘密共享技术为合法用户 $ID_{PS_{i_j}}$ 分享秘密参数.

(1) ID_{ES_i} 设置两个秘密参数 λ_i, γ_i, 满足 $\lambda_i+\gamma_i=0(\mathrm{mod}\ q_0)$.

(2) ID_{ES_i} 设置 k–1 次多项式 $EK_i(x)=\lambda_i+a_{i,1}x+a_{i,2}x^2+\ldots+a_{i,k-1}x^{k-1}$, 其中, $a_{i,1}, a_{i,2}, \ldots, a_{i,k-1} \in Z_{q_0}$ 是对应的多项式系数.

(3) ID_{ES_i} 为用户 $ID_{PS_{i_j}}$ 计算秘密值 $\lambda_{i_j} = EK_i(j)$, 同时秘密保存 $(ID_{PS_{i_j}}, \lambda_{i_j})$.

- 医疗数据加密和签名上传阶段

(1) 对于医疗数据 m_{i_j}, 用户 $ID_{PS_{i_j}}$ 选择随机数 r_{i_j} ($1 \leqslant r_{i_j} \leqslant n$), 利用秘密参数 λ_{i_j}, 根据新型 BGN 同态加密算法, 计算盲化的密态数据 $c_{i_j} = f^{\lambda_{i_j}} v_1^{m_{i_j}} v_2^{m_{i_j}^2} v_3^{r_{i_j}} \in \tilde{G}_1$.

(2) $ID_{PS_{i_j}}$ 选取随机数 $w_{i_j} \leftarrow Z_p^*$, 利用私钥 u_{i_j} 产生 c_{i_j} 的数字签名:

$$\sigma_{i_j} = (\sigma_{i_j,1}, \sigma_{i_j,2}) = \left(g_1^{w_{i_j}}, H(tag)^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})w_{i_j}+u_{i_j}}\right).$$

其中, tag 是医疗数据的属性类型, t_{i_j} 是当前时间戳.

(3) $ID_{PS_{i_j}}$ 将可验证密态数据信息 $Auth_{i_j}=\{c_{i_j},\sigma_{i_j},ID_{PS_{i_j}},t_{i_j},tag\}$ 发送给对应的边缘计算服务器 ID_{ES_i}.

- 边缘计算服务器数据聚合去盲化阶段

一旦成功接收到不同用户 $ID_{PS_{i_j}}$ 发送的 $Auth_{i_j}=\{c_{i_j},\sigma_{i_j},ID_{PS_{i_j}},t_{i_j},tag\}$, ID_{ES_i} 将这些有效身份 $ID_{PS_{i_j}}$ 信息形成数据集 $PST_i=\{ID_{PS_{i_j}}\}$, 当数据集中的有效样本容量 l_i 满足 $l_i\geq k(k$ 为门限值), ID_{ES_i} 对这 l_i 个不同的可验证密态数据信息进行如下批量验证:

$$e\left(\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,2},g_1\right)=e\left(H(tag),\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,1}^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})}U_{i_j}\right) \tag{2}$$

如果验证未通过, 则表示至少有一个用户 $ID_{PS_{i_j}}$ 上传的 $Auth_{i_j}$ 已经被替换或者篡改, ID_{ES_i} 终止后续操作. 如果以上方程验证通过, ID_{ES_i} 计算拉格朗日插值系数 $\beta_{i_j}=\prod_{ID_{PS_{i_z}}\in PST_i,z\neq j}\frac{z}{z-j}$, 进行如下边缘级密态数据聚合去盲化操作:

$$\begin{aligned}
c_i&=f^{\gamma_i}\prod_{ID_{PS_{i_j}}\in PST_i}f^{\lambda_{i_j}(\beta_{i_j}-1)}c_{i_j}\\
&=f^{\gamma_i}\prod_{ID_{PS_{i_j}}\in PST_i}f^{\lambda_{i_j}(\beta_{i_j}-1)}f^{\lambda_{i_j}}v_1^{m_{i_j}}v_2^{m_{i_j}^2}v_3^{r_{i_j}}\\
&=f^{\gamma_i}\prod_{ID_{PS_{i_j}}\in PST_i}f^{\lambda_{i_j}\beta_{i_j}}v_1^{m_{i_j}}v_2^{m_{i_j}^2}v_3^{r_{i_j}}\\
&=f^{\gamma_i}f^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}\lambda_{i_j}\beta_{i_j}}v_1^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}\\
&=f^{\gamma_i+\lambda_i}v_1^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}\\
&=v_1^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{\sum\limits_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}.
\end{aligned}$$

ID_{ES_i} 选取随机数 $w_i\leftarrow Z_p^*$, 利用私钥 u_i 产生 c_i 的数字签名 $\sigma_i=(\sigma_{i,1},\sigma_{i,2})=(g_1^{w_i},H(tag)^{h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i+u_i})$, 其中, Tim 是有效周期. 最后, ID_{ES_i} 向云服务器 ID_{PCC} 上载可验证的边缘级聚合密态数据信息:

$$\{c_i,\sigma_i,ID_{ES_i},tag,l_i\}.$$

- 云服务器存储有效数据

当云服务器 ID_{PCC} 接收到边缘计算服务器 ID_{ES_i} (i=1,2,…,N)的 $\{\sigma_i,ID_{ES_i},c_i,tag,l_i\}$, 云服务器 ID_{PCC} 对这 N 个可验证的边缘级聚合密态数据信息进行如下批量验证:

$$e\left(\prod_{i=1}^{N}\sigma_{i,2},g_1\right)=e\left(H(tag),\prod_{i=1}^{N}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}U_i\right) \tag{3}$$

如果验证未通过, 则表示至少有一个边缘计算服务器 ID_{ES_i} 上传的信息是无效的, 然后, ID_{PCC} 逐个执行如下验证: $e(\sigma_{i,2},g_1)=e(H(tag),\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}U_i)$. 当所有信息验证通过之后, ID_{PCC} 存储有效的边缘级聚合密态数据信息 $\{\sigma_i,ID_{ES_i},c_i,Tim,tag,l_i\}_{1\leqslant i\leqslant N}$.

- 云服务器数据聚合阶段

在有效周期 Tim 内, 当医疗数据分析中心 ID_{DAC} 需要对特定区域的 tag 属性类型的医疗数据进行统计分析时, 选择这些区域的边缘计算服务器的身份信息的集合 EST, 然后发送挑战信息 $\{EST,tag\}$ 给云服务器 ID_{PCC}.

按照集合 EST 中的所有身份信息提取出 tag 属性类型对应的可验证的边缘级聚合密态数据信息, ID_{PCC} 产生如下云级聚合密态数据:

$$c=\prod_{ID_{ES_i}\in EST}c_i=v_1^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}.$$

ID_{PCC} 产生如下云级聚合数字签名: $\sigma_{Agg}=(\sigma_{Agg,1},\sigma_{Agg,2})=\left(\prod_{ID_{ES_i}\in EST}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)},\prod_{ID_{ES_i}\in EST}\sigma_{i,2}\right)$.

ID_{PCC} 计算所有用户数量 $L=\sum_{ID_{ES_i}\in EST}l_i$.

最后, ID_{PCC} 选取随机数 $w\leftarrow Z_p^*$, 并用私钥 u 产生如下数字签名:

$$\sigma_{PCC}=(\sigma_{PCC,1},\sigma_{PCC,2})=(g_1^w,H(tag)^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)w+u}).$$

ID_{PCC} 返回可验证的云级聚合数据信息$\{\sigma_{Agg},ID_{PCC},c,L,tag,\sigma_{PCC},Tim\}$给 ID_{DAC}.

- 可验证的聚合密态数据解密与统计分析

当收到 ID_{PCC} 发送的可验证的云级聚合数据信息$\{\sigma_{Agg},ID_{PCC},c,L,tag,\sigma_{PCC},Tim\}$, 医疗数据分析中心 ID_{DAC} 执行如下验证:

$$e(\sigma_{PCC,2},g_1)=e(H(tag),\sigma_{PCC,1}^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)}U) \tag{4}$$

如果验证未通过, 则表示数据无效(数据被替换或者篡改), ID_{DAC} 重新发起挑战; 如果通过, 则执行第 2 个验证方程:

$$e(\sigma_{Agg,2},g_1)=e\left(H(tag),\sigma_{Agg,1}\prod_{ID_{ES_i}\in EST}U_i\right) \tag{5}$$

如果验证通过, 则表明 ID_{PCC} 是严格按照挑战信息$\{EST,tag\}$进行云级密态数据聚合.

最后, ID_{DAC} 利用私钥 sk_1, 计算:

$$\begin{aligned}SC&=c^{sk_1}\\&=v_1^{(q_1q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{(q_1q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{(q_1q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}\\&=\tilde{g}^{(q_1^2q_2^2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}\tilde{g}^{(q_0q_1q_2^2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}\tilde{g}^{(q_0q_1^2q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}\\&=\tilde{g}^{(q_1^2q_2^2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}.\end{aligned}$$

ID_{DAC} 私钥 sk_2 计算:

$$\begin{aligned}QSC&=c^{sk_2}\\&=v_1^{(q_0q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{(q_0q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{(q_0q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}\\&=\tilde{g}^{(q_0q_1q_2^2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}\tilde{g}^{(q_0^2q_2^2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}\tilde{g}^{(q_0^2q_1q_2q_3)\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}\\&=\tilde{g}^{(q_0^2q_2^2q_3)\sum_{D_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}.\end{aligned}$$

根据新型 BGN 同态加密算法的解密步骤, ID_{DAC} 可有效求解 $\log_{\tilde{g}^{q_1^2q_2^2q_3}}^{SC}$ 和 $\log_{\tilde{g}^{q_0^2q_2^2q_3}}^{QSC}$, 即可恢复 tag 属性类型医疗数据的统计和 $\sum_{D_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}$ 以及平方和 $\sum_{D_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2$. 据此, ID_{DAC} 可计算出该类型医疗数据的平均值和方差:

平均值: $\bar{m}=\dfrac{\sum_{D_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}}{L}$; 方差: $\mathrm{var}=\dfrac{\sum_{D_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}^2}{L}-\left(\dfrac{\sum_{D_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}}{L}\right)^2$.

最后, ID_{DAC} 可进一步在确保用户医疗数据安全的情况下进行大数据统计与深度分析.

4 正确性证明

在本节, 我们首先进行方案中所涉及的密态数据完整性验证的各个方程正确性证明.

身份信息注册验证方程(1) $e(sig_{i_j,2},g_1)=e\left(H(ID_{ES_i}),sig_{i_j,1}^{h(ID_{PS_{i_j}}\|t_{reg})}U_{i_j}\right)$ 的正确性推导如下:

$$\begin{aligned}
e(sig_{i_j,2},g_1)&=e\left(H(ID_{ES_i})^{h(ID_{PS_{i_j}}\|t_{reg})v_{i_j}+u_{i_j}},g_1\right)\\
&=e\left(H(ID_{ES_i})^{h(ID_{PS_{i_j}}\|t_{reg})v_{i_j}}H(ID_{ES_i})^{u_{i_j}},g_1\right)\\
&=e\left(H(ID_{ES_i})^{h(ID_{PS_{i_j}}\|t_{reg})v_{i_j}},g_1\right)\cdot e(H(ID_{ES_i})^{u_{i_j}},g_1)\\
&=e\left(H(ID_{ES_i}),g_1^{h(ID_{PS_{i_j}}\|t_{reg})v_{i_j}}\right)\cdot e(H(ID_{ES_i}),g_1^{u_{i_j}})\\
&=e\left(H(ID_{ES_i}),sig_{i_j,1}^{h(ID_{PS_{i_j}}\|t_{reg})}\right)\cdot e(H(ID_{ES_i}),U_{i_j})\\
&=e\left(H(ID_{ES_i}),sig_{i_j,1}^{h(ID_{PS_{i_j}}\|t_{reg})}U_{i_j}\right).
\end{aligned}$$

批量验证方程(2) $e\left(\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,2},g_1\right)=e\left(H(tag),\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,1}^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})}U_{i_j}\right)$ 的正确性推导如下:

$$\begin{aligned}
e\left(\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,2},g_1\right)&=e\left(\prod_{ID_{PS_{i_j}}\in PST_i}H(tag)^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})w_{i_j}+u_{i_j}},g_1\right)\\
&=e\left(\prod_{ID_{PS_{i_j}}\in PST_i}H(tag)^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})w_{i_j}}H(tag)^{u_{i_j}},g_1\right)\\
&=e\left(H(tag)^{\sum_{ID_{PS_{i_j}}\in PST_i}h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})w_{i_j}},g_1\right)\cdot e\left(H(tag)^{\sum_{ID_{PS_{i_j}}\in PST_i}u_{i_j}},g_1\right)\\
&=e\left(H(tag),g_1^{\sum_{ID_{PS_{i_j}}\in PST_i}h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})w_{i_j}}\right)\cdot e\left(H(tag),g_1^{\sum_{ID_{PS_{i_j}}\in PST_i}u_{i_j}}\right)\\
&=e\left(H(tag),\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,1}^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})}\right)\cdot e\left(H(tag),\prod_{ID_{PS_{i_j}}\in PST_i}U_{i_j}\right)\\
&=e\left(H(tag),\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,1}^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})}U_{i_j}\right).
\end{aligned}$$

批量验证方程(3) $e\left(\prod_{i=1}^{N}\sigma_{i,2},g_1\right)=e\left(H(tag),\prod_{i=1}^{N}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}U_i\right)$ 的正确性推导如下:

$$\begin{aligned}
e\left(\prod_{i=1}^{N}\sigma_{i,2},g_1\right)&=e\left(\prod_{i=1}^{N}H(tag)^{h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i+u_i},g_1\right)\\
&=e\left(\prod_{i=1}^{N}H(tag)^{h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i}H(tag)^{u_i},g_1\right)\\
&=e\left(H(tag)^{\sum_{i=1}^{N}h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i},g_1\right)\cdot e\left(H(tag)^{\sum_{i=1}^{N}u_i},g_1\right)
\end{aligned}$$

$$
\begin{aligned}
&= e\left(H(tag), g_1^{\sum_{i=1}^{N} h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i}\right)\cdot e\left(H(tag), g_1^{\sum_{i=1}^{N} u_i}\right)\\
&= e\left(H(tag), \prod_{i=1}^{N}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}\right)\cdot e\left(H(tag), \prod_{i=1}^{N} U_i\right)\\
&= e\left(H(tag), \prod_{i=1}^{N}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)} U_i\right).
\end{aligned}
$$

验证方程(4) $e(\sigma_{PCC,2}, g_1) = e(H(tag), \sigma_{PCC,1}^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)}U)$ 的正确性推导如下:

$$
\begin{aligned}
e(\sigma_{PCC,2}, g_1) &= e(H(tag)^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)w+u}, g_1)\\
&= e(H(tag)^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)w}H(tag)^{u}, g_1)\\
&= e(H(tag)^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)w}, g_1)\cdot e(H(tag)^{u}, g_1)\\
&= e(H(tag), g_1^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)w})\cdot e(H(tag), g_1^{u})\\
&= e(H(tag), \sigma_{PCC,1}^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)})\cdot e(H(tag), U)\\
&= e(H(tag), \sigma_{PCC,1}^{h(ID_{PCC}\|\sigma_{Agg}\|c\|L\|Tim)}U).
\end{aligned}
$$

验证方程(5) $e(\sigma_{Agg,2}, g_1) = e\left(H(tag), \sigma_{Agg,1}\prod_{ID_{ES_i}\in EST} U_i\right)$ 的正确性推导如下:

$$
\begin{aligned}
e(\sigma_{Agg,2}, g_1) &= e\left(\prod_{ID_{ES_i}\in EST}\sigma_{i,2}, g_1\right)\\
&= e\left(\prod_{ID_{ES_i}\in EST} H(tag)^{h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i+u_i}, g_1\right)\\
&= e\left(H(tag)^{\sum_{ID_{ES_i}\in EST} h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i} H(tag)^{\sum_{ID_{ES_i}\in EST} u_i}, g_1\right)\\
&= e\left(H(tag)^{\sum_{ID_{ES_i}\in EST} h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i}, g_1\right)\cdot e\left(H(tag)^{\sum_{ID_{ES_i}\in EST} u_i}, g_1\right)\\
&= e\left(H(tag), g_1^{\sum_{ID_{ES_i}\in EST} h(ID_{ES_i}\|c_i\|l_i\|Tim)w_i}\right)\cdot e\left(H(tag), g_1^{\sum_{ID_{ES_i}\in EST} u_i}\right)\\
&= e\left(H(tag), \prod_{ID_{ES_i}\in EST}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}\right)\cdot e\left(H(tag), \prod_{ID_{ES_i}\in EST} U_i\right)\\
&= e\left(H(tag), \sigma_{Agg,1}\prod_{ID_{ES_i}\in EST} U_i\right).
\end{aligned}
$$

接下来, 我们对支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案的进行安全性分析.

5 安全性分析

定理 1. 支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案可确保各阶段用户医疗数据的机密性.

证明: 移动终端用户 $ID_{PS_{i_j}}$ 产生 m_{i_j} 的医疗密态数据 $c_{i_j} = f^{\lambda_{i_j}} v_1^{m_{i_j}} v_2^{m_{i_j}^2} v_3^{r_{i_j}}$, 本质上是新型 BGN 同态加密算法生成密文的盲化值.

此外, 在边缘计算服务器对数据集 $Auth_i$ 的有效性进行批量验证通过后, 对数据集 $Auth_i$ 中所有医疗密态数据 $c_{i_j} = f^{\lambda_{i_j}} v_1^{m_{i_j}} v_2^{m_{i_j}^2} v_3^{r_{i_j}}$ 进行边缘级去盲化的密态数据聚合, 最终得到:

$$
c_i = v_1^{\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}} v_2^{\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}^2} v_3^{\sum_{ID_{PS_{i_j}}\in PST_i} r_{i_j}}.
$$

本质上, c_i 是 $\sum_{j=1}^{l} m_{i_j}$ 和 $\sum_{j=1}^{l} m_{i_j}^2$ 的新型 BGN 同态加密算法去盲化的密文.

一旦接收到医疗数据分析中心 ID_{DAC} 的挑战信息, ID_{PCC} 云服务器产生如下的云级密态聚合数据:

$$c=\prod_{ID_{ES_i}\in EST} c_i = v_1^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}} v_2^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}^2} v_3^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} r_{i_j}}.$$

本质上, c 是 $\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}$ 和 $\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i} m_{i_j}^2$ 的新型 BGN 的同态加密算法的密文.

根据文献[25]安全性分析可知: 新型BGN同态加密算法的安全性本质上与原型BGN加密算法一样, 都是基于子群判定困难性问题, 满足选择明文安全的语义安全性. 因此, 即使敌手在整个医疗信息传输或存储阶段截获到相关密文, 也不能恢复用户的原始医疗数据及其相关统计信息. □

定理 2. 支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案可确保各阶段医疗密态数据聚合的完整性.

证明: 支持边缘服务计算的可验证医疗密态数据聚合中设计了一个基于身份的聚合签名算法来确保各阶段医疗密态数据聚合的完整性. 具体来说, 在医疗数据加密和签名上传阶段, 用户 $ID_{PS_{i_j}}$ 使用自己的私钥 u_{i_j} 对盲化的密态数据进行数字签名, 得到签名值 $\sigma_{i_j}=(\sigma_{i_j,1},\sigma_{i_j,2})=(g_1^{w_{i_j}}, H(tag)^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})w_{i_j}+u_{i_j}})$; 在边缘计算服务器数据聚合去盲化阶段, 边缘服务器 ID_{ES_i} 利用自己的私钥 u_i 对聚合后的密态数据进行数字聚合签名, 得到签名值 $\sigma_i=(g_1^{w_i}, H(tag)^{h(ID_{ES_i}\|c_i\|t_i'\|Tim)w_i+u_i})$. 事实上, 本方案基于身份聚合签名算法的设计是基于文献[32]中构造思想, 文献[32]已经对基于身份的聚合签名算法存在不可伪造性的可证明安全论证详细归约到 CDH 困难性问题, 因此, 本方案中的聚合签名算法同样可以达到存在不可伪造性.

本方案将侧重证明在各阶段医疗密态数据传输与聚合的完整性, 即在各阶段, 敌手试图篡改或者替换密态数据来通过完整性验证是计算不可行的. 接下来, 我们将从 Game 1, Game 2, Game 3 和 Game 4 能够满足数据完整性保证的方案设计目标.

- Game 1

首先, 在本方案中, 在医疗数据加密和签名上传阶段, 不同于按照方案步骤生成正确的可验证密态数据信息, 我们假设至少有一个用户 $ID_{PS_{i_\tau}}$ 在传输可验证密态数据信息 $Auth_{i_\tau}=\{c_{i_\tau},\sigma_{i_\tau},ID_{PS_{i_\tau}},t_{i_\tau},tag\}$ 到边缘计算服务器 ID_{ES_i} 的过程中, 其密态数据 c_{i_τ} 被敌手 $\mathcal{A}_1$ 以不可忽略的优势篡改或者替换为 $c_{i_\tau}^*$, 通过以下批量验证方程:

$$e\left(\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,2}, g_1\right)=e\left(H(tag), \sigma_{i_\tau,1}^{h(ID_{PS_{i_\tau}}\|c_{i_\tau}^*\|t_{i_\tau})}\cdot\prod_{ID_{PS_{i_j}}\in PST_i\setminus\{ID_{PS_{i_\tau}}\}}\sigma_{i_j,1}^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})}U_{i_j}\right).$$

对于不同的真实的可验证密态数据信息需要满足如下批量验证方程:

$$e\left(\prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,2}, g_1\right)=e\left(H(tag), \prod_{ID_{PS_{i_j}}\in PST_i}\sigma_{i_j,1}^{h(ID_{PS_{i_j}}\|c_{i_j}\|t_{i_j})}U_{i_j}\right).$$

结合以上两个方程得到: $\sigma_{i_\tau,1}^{h(ID_{PS_{i_\tau}}\|c_{i_\tau}^*\|t_{i_\tau})}=\sigma_{i_\tau,1}^{h(ID_{PS_{i_\tau}}\|c_{i_\tau}\|t_{i_\tau})}$.

设置 $\beta=\sigma_{i_\tau,1}^{h(ID_{PS_{i_\tau}}\|c_{i_\tau}\|t_{i_\tau})}$, 于是, 如果敌手可以在多少项时间内篡改或者替换密态数据并通过完整性验证, 则其必在多项式时间内求解到 $\sigma_{i_\tau,1}$ 与 β 之间的离散对数 $h(ID_{PS_{i_\tau}}\|c_{i_\tau}^*\|t_{i_\tau})$. 这与求解离散对数困难性相矛盾. 因此, 敌手 $\mathcal{A}_1$ 以不可忽略的优势篡改或者替换某些用户的密态数据, 通过边缘服务器的批量验证来赢得 Game 1 是计算不可行的.

- Game 2

在边缘计算服务器数据聚合去盲化阶段, 由于 $c_i=v_1^{\sum_{Auth_{i_j}\in Auth_i} m_{i_j}} v_2^{\sum_{Auth_{i_j}\in Auth_i} m_{i_j}^2} v_3^{\sum_{Auth_{i_j}\in Auth_i} r_{i_j}}$ 是边缘服务器对密态

数据聚合去盲化后产生的, 不同于按照方案步骤生成正确的边缘级聚合信息, 我们假设至少有一个边缘服务器 ID_{ES_τ} 在上传边缘级聚合数据 c_τ到云服务器的过程中, 敌手$\mathcal{A}_2$ 以不可忽略的优势替换或者篡改 c_τ为 c_τ^*, 即替换或者篡改后的边缘级聚合信息 $\{\sigma_\tau, ID_{ES_\tau}, c_\tau^*, tag, l_\tau\}$ 要通过云服务器的如下批量验证方程:

$$e\left(\prod_{i=1}^{N}\sigma_{i,2}, g_1\right)=e\left(H(tag), \sigma_{\tau,1}^{h(ID_{ES_\tau}\|c_\tau^*\|l_\tau\|Tim)}\cdot\prod_{i\in\{1,\ldots,N\}\setminus\{\tau\}}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}U_i\right).$$

对于 N 个不同的边缘级聚合信息 $\{\sigma_i, ID_{ES_i}, c_i, tag, l_i\}$, 应满足如下批量验证方程:

$$e\left(\prod_{i=1}^{N}\sigma_{i,2}, g_1\right)=e\left(H(tag), \prod_{i=1}^{N}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}U_i\right).$$

根据以上两个批量验证方程得知: $\sigma_{i,1}^{h(ID_{ES_i}\|c_i^*\|l_i\|Tim)}=\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}$.

设置 $\alpha=g_1^{w_1h(ID_{ES_i}\|c_i^*\|l_i\|Tim)}=\sigma_{i,1}^{h(ID_{ES_i}\|c_i^*\|l_i\|Tim)}$, 我们可以求解 g_1 和α之间的离散对数 $w_1h(ID_{ES_i}\|c_i^*\|l_i\|Tim)$, 这与离散对数困难问题假设是矛盾的. 因此, 敌手$\mathcal{A}_2$ 以不可忽略的优势篡改或者替换某些边缘级聚合数据, 并通过云服务器的批量验证来赢得 Game 2 是计算不可行的.

- Game 3

在云服务器数据聚合阶段, ID_{PCC}产生如下云级聚合密态数据:

$$c=\prod_{ID_{ES_i}\in EST}c_i=v_1^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}}v_2^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}m_{i_j}^2}v_3^{\sum_{ID_{ES_i}\in EST}\sum_{ID_{PS_{i_j}}\in PST_i}r_{i_j}}$$

以及云级聚合数字签名:

$$\sigma_{Agg}=(\sigma_{Agg,1},\sigma_{Agg,2})=\left(\prod_{ID_{ES_i}\in EST}\sigma_{i,1}^{h(ID_{ES_i}\|c_i\|l_i\|Tim)}, \prod_{ID_{ES_i}\in EST}\sigma_{i,2}\right).$$

我们假设敌手$\mathcal{A}_3$ 以不可忽略的优势篡改或者替换最终挑战的聚合密态数据 c 为 c^*, 通过以下验证方程:

$$e(\sigma_{PCC,2}, g_1)=e(H(tag), \sigma_{PCC,1}^{h(ID_{PCC}\|\sigma_{Agg}\|c^*\|L\|Tim)}U).$$

用同样的方法可以分析得到敌手$\mathcal{A}_3$ 必须要求解离散对数困难问题. 因此, 敌手$\mathcal{A}_3$ 以不可忽略的优势篡改或者替换最终挑战的聚合密态数据, 并通过医疗数据分析中的完整性验证来赢得 Game 3 是计算不可 行的.

- Game 4

由于σ_{Agg} 是云服务器按照集合 EST 中的所有身份信息提取出 tag 属性类型对应的可验证的边缘级聚合密态数据的签名信息进行聚合后的云级聚合数字签名, 我们假设敌手$\mathcal{A}_4$ (恶意云服务器)以不可忽略的优势在聚合过程中是按照非法集合 EST^*对边缘级聚合密态数据的签名信息进行聚合得到非法的云级聚合数字签名 σ_{Agg}^*, 那么非法的云级聚合数字签名 σ_{Agg}^* 要通过医疗数据分析中心的验证, 并满足以下方程:

$$e(\sigma_{Agg,2}^*, g_1)=e\left(H(tag), \sigma_{Agg,1}^*\prod_{ID_{ES_i}\in EST}U_i\right).$$

而根据方程(5)的正确性验证可知, σ_{Agg}^* 满足 $e(\sigma_{Agg,2}^*, g_1)=e\left(H(tag), \sigma_{Agg,1}^*\prod_{ID_{ES_i}\in EST^*}U_i\right)$.

根据以上两个方程得知:

$$e\pi\left(H(tag), \sigma_{Agg,1}^*\prod_{ID_{ES_i}\in EST^*}U_i\right)=e\left(H(tag), \sigma_{Agg,1}^*\prod_{ID_{ES_i}\in EST}U_i\right).$$

即 $\prod_{ID_{ES_i}\in EST^*}U_i=\prod_{ID_{ES_i}\in EST}U_i$. 显然是不成立的. 因此, 敌手$\mathcal{A}_4$ 以不可忽略的优势按照非法集合 EST^*对边缘级聚合密态数据的签名信息进行聚合, 并通过医疗数据分析中的验证来赢得 Game 4 是计算不可行的.

因此, 根据以上 Game 1, Game 2, Game 3 和 Game 4 安全性分析过程得知, 支持边缘服务计算的可验证医

疗密态数据聚合与统计分析方案可确保各阶段医疗密态数据聚合的完整性. □

6 性能分析与评估

6.1 安全与统计特性比较

本节我们首先将支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案和现有的方案[15,20,29,30]在安全与统计特性功能方面进行比较, 具体见表 1. 结果表明, 支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案同时具备多源密态数据聚合、分层级聚合、支持传输容错机制、支持方差统计特征, 同时确保数据在整个系统传输与存储中的机密性与完整性验证的安全功能. 因此, 本方案可有效部署在无线医疗大数据处理与隐私保护领域.

表 1 安全与统计特性比较

方案	多源	可容错	机密性	完整性验证	分层级聚合	支持方差统计
文献[30]	√	√	√	√	×	×
文献[29]	√	√	√	√	×	×
文献[15]	√	√	√	×	√	×
文献[20]	√	×	√	√	√	×
本方案	√	√	√	√	√	√

6.2 计算与通信开销比较

现在, 我们将支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案与可验证的密态数据聚合方案[20,29,30]进行性能比较与分析. 在性能比较和分析时, 硬件环境为处理器: Inter (R) Core (TM) I5-2320 3.00 GHz 和内存条: DDR4 2 666 MHz 8 GB 组成的主机, 软件环境为操作系统: Windows 10 和 c 语言密码算法基础函数库 MIRACL. 其中, q_0, q_1, q_2, q_3 都是 512 位的大素数, 群 $\tilde{G}_1$, $\tilde{G}_2$ 中的元素长度是 2 048 位, 群 G_1, G_2 中的元素长度是 512 位, 椭圆曲线上点元素是 320 位. T_{pa} 是一次双线性对运算所需时间, T_{Mu} 是椭圆曲线中一次倍点运算所需时间, T_{mu} 是一次普通模乘法运算所需时间, T_{ex} 是一次普通模指数运算所需时间, T_{Ha} 是一次映射到循环群中的椭圆曲线点坐标的运算所需时间, T_{ha} 是一次普通哈希函数运算所需时间, T_{add} 是基于椭圆曲线的加法循环群中一次加法运算所需时间. 具体的算法实验仿真数值见表 2.

表 2 不同算法仿真的执行时间

符号	运行时间(ms)
T_{pa}	5.427 0
T_{Mu}	2.165 2
T_{mu}	0.000 9
T_{ex}	1.17
T_{Ha}	5.493 0
T_{ha}	0.007 87
T_{add}	0.013 2

支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案与可验证的密态数据聚合方案[20,29,30]在用户端、边缘服务器端、数据分析中心处理端的计算开销对比结果见表 3.

在我们提出的系统模型中, 云服务器和医疗数据分析中心所执行的功能和其他方案的模型中的数据中心实体所执行的功能类似, 因此, 我们将云服务器和医疗数据分析中心统称为数据中心和其他方案进行对比. 接下来, 我们进行计算效率分析. 根据文献[30]中所提出的 ASAS 方案, 我们分析得知: 在用户端, 需要执行 1 次模乘法运算和 3 次普通模指数运算才能构造出密文, 同时需要执行 1 次映射到循环群的哈希运算和 1 次倍点运算才能获得签名值, 最后执行 2 次双线性对运算、1 次映射到循环群的哈希运算用来验证数据完整性. 用户端总共的计算开销为 $3T_{ex}+T_{mu}+2T_{Ha}+T_{Mu}+2T_{pa}$. 边缘服务器需要执行 $2l$ 次双线性对运算和 l 次映射到循环群的哈希运算验证所有查询, 同时需要执行 2 次双线性对运算、$2(l-1)$次椭圆曲线上的加法运算和 l 次映射到循环群的哈希运算验证 l 个消息的有效性, 然后执行 $2(l-1)$次模乘法运算对数据进行聚合, 最后执行 1 次映射到

循环群的哈希运算和 1 次模乘法运算构造签名. 边缘服务器总共的计算开销为 $(2l+1)T_{Ha}+T_{Mu}+2(l+1)T_{pa}+2(l-1)T_{mu}+2(l-1)T_{add}$. 数据中心需要执行 $2N$ 次双线性对运算和 N 次映射到循环群的哈希运算验证聚合消息的有效性, 并执 N 次模指数运算和 N 次模乘法运算解密出明文. 数据中心总共的计算开销为 $NT_{Ha}+NT_{mu}+2NT_{pa}+NT_e$.

表 3 计算开销比较

方案	用户端	边缘服务器端	数据分析中心端
文献[30]	$3T_{ex}+T_{mu}+2T_{Ha}+T_{Mu}+2T_{pa}$	$(2l+1)T_{Ha}+T_{Mu}+2(l+1)T_{pa}+2(l-1)T_{mu}+2(l-1)T_{add}$	$NT_{Ha}+NT_{mu}+2NT_{pa}+NT_{ex}$
文献[29]	$2T_{ex}+T_{Ha}+3T_{Mu}+T_{ha}+3T_{mu}+T_{add}$	$(l+2)T_{pa}+2(l-1)T_{add}+lT_{ex}+lT_{ha}+6lT_{mu}+T_{Ha}+T_{Mu}$	$2NT_{pa}+NT_{Ha}+NT_{ex}+2NT_{mu}$
文献[20]	$2T_{mu}+T_{Ha}+4T_{ex}$	$(l+1)T_{pa}+2T_{ex}+(3l-2)T_{mu}+(l+1)T_{Ha}$	$(N+1)T_{pa}+NT_{Ha}+3(N-1)T_{mu}+T_{ex}$
本方案	$T_{Ha}+T_{ha}+6T_{ex}+4T_{mu}$	$2T_{Ha}+(6l-1)T_{mu}+(2l+3)T_{ex}+2T_{pa}+(l+1)T_{ha}$	$3T_{Ha}+(7N-3)T_{mu}+(N+5)T_{ex}+6T_{pa}+(N+2)T_{ha}$

根据文献[29]中所提出的 PHDA 方案, 我们分析得知: 在用户端, 需要执行 2 次模指数运算和 2 次乘法运算构造密文, 然后执行 1 次射到循环群的哈希运算、3 次倍点运算、1 次乘法运算、1 次椭圆曲线上加法循环群的加法运算以及 1 次普通哈希运算完成对密文的签名计算. 用户端总共的计算开销为 $2T_{ex}+T_{Ha}+3T_{Mu}+T_{ha}+3T_{mu}+T_{add}$. 边缘服务器需要执行$(2l-1)$次椭圆曲线上加法循环群中的加法运算、$l+2$ 次双线性对运算、l 次模指数运算、l 次普通哈希运算以及 l 次乘法运算验证用户数据完整性. 同时需要执行 $5l$ 次乘法运算聚合用户的密文数据. 最后执行 1 次映射到循环群的哈希运算和 1 次倍点运算构造聚合值的签名值. 边缘服务器总共的计算开销为$(l+2)T_{pa}+2(l-1)T_{add}+lT_{ex}+lT_{ha}+6lT_{mu}+T_{Ha}+T_{Mu}$. 数据中心需要执行 $2N$ 次双线性对运算和 N 次映射到循环群的哈希运算验证边缘服务器的数据完整性. 然后执行 N 次模指数运算和 $2N$ 次乘法运算对数据进行解密. 数据中心总共的计算开销为 $2NT_{pa}+NT_{Ha}+NT_{ex}+2NT_{mu}$.

根据文献[20]中的方案, 我们分析得知: 在用户端, 需要执行 3 次模指数运算和 2 次模乘法运算构造密文, 然后执行 1 次映射到循环群的哈希运算和 1 次普通模指数运算完成对密文的签名计算. 用户端总共的计算开销为 $2T_{mu}+T_{Ha}+4T_{ex}$. 边缘服务器需要执行$(l+1)$次双线性对运算、$2(l-1)$次乘法运算和 l 次映射到循环群的哈希运算验证用户数据完整性. 然后执行 l 次模乘法运算和 1 次模指数运算完成数据的聚合运算. 最后执行 1 次映射到循环群的哈希运算和 1 次普通的模指数运算构造聚合值的签名值. 边缘服务器总共的计算开销为 $(l+1)T_{pa}+2T_{ex}+(3l-2)T_{mu}+(l+1)T_{Ha}$. 数据中心需要执行$(N+1)$次双线性对运算、$N$ 次映射到循环群的哈希运算以及 $2(N-1)$次模乘法运算验证用户数据完整性. 然后执行$(N-1)$次模乘法运算和 1 次模指数运算完成聚合解密. 数据中心总共的计算开销为$(N+1)T_{pa}+NT_{Ha}+3(N-1)T_{mu}+T_{ex}$.

在本方案中, 用户端需要执行 4 次模指数运算和 3 次模乘法运算构造密文, 然后执行 1 次映射到循环群的哈希运算、1 次普通哈希运算、1 次模乘法运算以及 2 次模指数运算得到密文数据的签名值. 用户端总共的计算开销为 $T_{Ha}+T_{ha}+6T_{ex}+4T_{mu}$. 边缘服务器需要执行 1 次映射到循环群的哈希运算、$3l-2$ 次模乘法运算、l 次模指数运算、2 次双线性对运算和 l 次普通哈希运算验证用户数据的完整性. 然后执行$(l+1)$次模指数运算和 $3l$ 次模乘法运算聚合用户密文数据并去除盲化. 最后执行 2 次模指数运算、1 次映射到循环群的哈希运算、1 次模乘法运算和 1 次普通哈希运算对边缘级聚合数据进行签名. 边缘服务器总共的计算开销为 $2T_{Ha}+(6l-1)T_{mu}+(2l+3)T_{ex}+2T_{pa}+(l+1)T_{ha}$. 在本方案中, 数据中心被分为云服务器和医疗数据分析中两个部分, 其中,

- 云服务器需要执行 2 次双线性对运算、$3N-2$ 次模乘法运算、N 次模指数运算、N 次普通哈希运算和 1 次映射到循环群的哈希运算验证边缘服务器上传数据的完整性, 然后执行 $3(N-1)$次模乘法运算得到云级聚合数据, 最后执行 2 次模指数运算、1 次普通哈希运算、1 次模乘法运算和 1 次映射到循环群的哈希运算对聚合数据进行签名. 云服务器总共的计算开销为 $2T_{Ha}+(6N-4)T_{mu}+(N+2)T_{ex}+2T_{pa}+(N+1)T_{ha}$.
- 医疗数据分析中心需要执行 1 次映射到循环群的哈希运算、1 次普通哈希运算、$N+1$ 次模乘法运算、1 次模指数运算和 4 次双线性对运算验证云服务器上传数据的完整性, 最后执行 2 次模指数运算解密

得到明文. 医疗数据分析中心的总开销为 $T_{Ha}+(N+1)T_{mu}+3T_{ex}+4T_{pa}+T_{ha}$.

所以数据中心的总开销为 $3T_{Ha}+(7N-3)T_{mu}+(N+5)T_{ex}+6T_{pa}+(N+2)T_{ha}$. 具体的性能分析对比如图 3–图 5 所示.

图 3 表明, 本设计方案在用户端的计算开销方面低于文献[30]和文献[29]中方案, 略高于文献[20]中方案. 这是由于本设计方案在功能方面增加了传输容错机制, 并且可以同时保持对密态数据的和、方差等统计特征; 而文献[20]中方案不能同时满足这两点特征. 图 4 和图 5 表明, 本设计方案在边缘服务器和数据中心端的计算开销明显低于其他方案. 特别地, 本设计方案在边缘服务器端和数据中心端进行了两层聚合, 随着移动终端用户数量和边缘服务器数量的增加, 其计算开销增幅非常小. 因此, 本设计方案在分层密态数据聚合与统计分析过程中具有明显的计算效率优势.

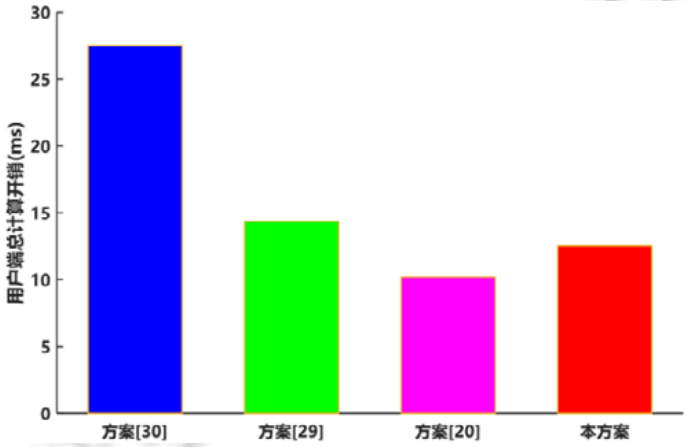

图 3 用户端的计算开销

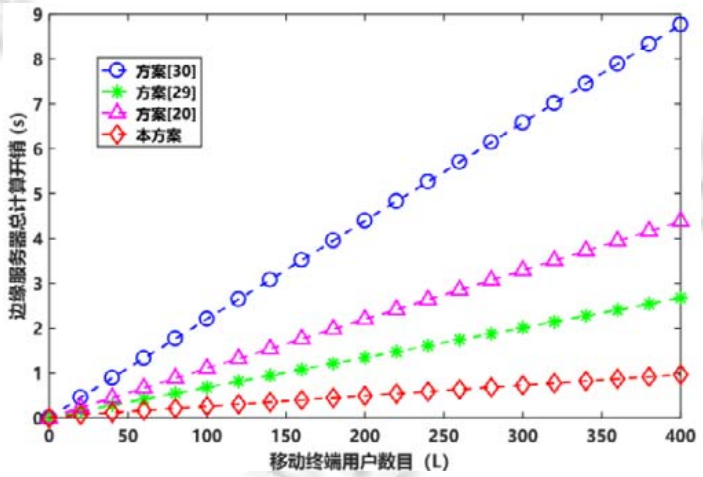

图 4 边缘服务器端的计算开销

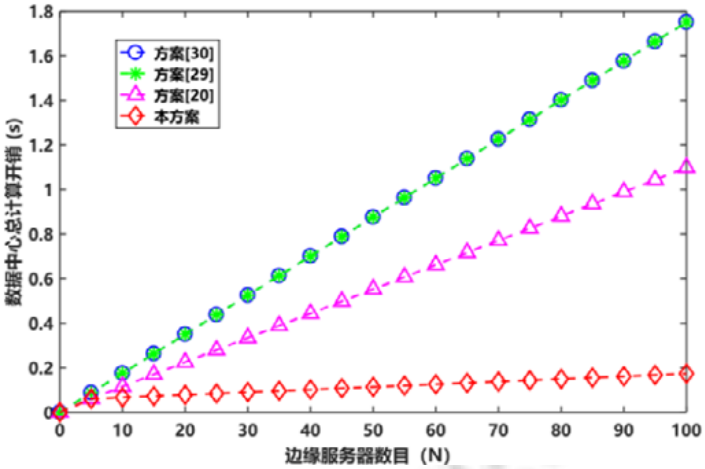

图 5 数据中心端的计算开销

最后, 在通信开销方面, 我们将支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案与可验证的密态数据聚合方案[20,29,30]进行分析与比较. 首先统一实体的身份信息(ID)长度为 64 位, 时间戳(t)的长度为 32 位, 数据标签(tag)的长度为 32 位, 用户数量(l)的长度为 32 位. 其余群中元素的长度与本方案统一.

本方案的通信包括两个部分——用户端上传密文数据到边缘服务器和边缘服务器上传聚合数据到数据中心: 在用户数据加密和上传阶段, 用户端上传可验证密态数据消息 $\{c_{i_j},\sigma_{i_j},ID_{PS_{i_j}},t_{i_j},tag\}$ 到边缘服务器的通信开销为 2048+1024+64+32+32=3200 比特; 在边缘服务器聚合和上传阶段, 边缘服务器上传可验证的边缘级聚合密态数据信息 $\{c_i,\sigma_i,ID_{ES_i},tag,l_i\}$ 到数据中心的通信开销为 2048+1024+64+32+32=3200 比特.

在方案[30]中, 用户端需要给边缘服务器发送 32+32+512=576 比特的查询信息, 边缘服务器也需要回复用户端 32+32+512=576 比特的信息完成身份认证. 此外, 用户端需要上传 4096+32+64+512=4704 比特的密态数据到边缘服务器. 因此在用户端, 加密和上传阶段用户端和边缘服务器的总通信开销为 576+576+4704=5856 比特; 在数据聚合和上传阶段, 边缘服务器上传聚合数据到数据中心的通信开销为 4096+32+32+64+512=4736 比特.

在文献[29]的方案中, 用户端上传加密数据到边缘服务器的通信开销为 1024+2048=3072 比特, 边缘服务器上传聚合数据到数据中心的通信开销为 512+2048=2560 比特.

最后, 在文献[20]的方案中, 用户端上传加密数据到边缘服务器的通信开销为 64+32+2048+2048=4192 比特, 边缘服务器上传聚合数据到数据中心的通信开销为 64+32+2048+2048=4192 比特. 具体的通信开销对比如图 6 所示.

图 6 表明: 设计方案在用户端上传数据阶段的通信开销方面均低于文献[30]和文献[20]中方案, 略高于文献[29]中方案. 根据前面计算开销方面得知: 本设计方案在边缘服务器端和数据中心端方面的计算开销都明显低于文献[29]中方案, 而且文献[29]中方案不能支持方差统计. 整体上, 本设计方案在各层数据传输之中具备合理的通信开销优势.

因此, 从以上的性能分析与比较结果可知, 本设计方案不仅同时达到了多源密态数据聚合、分层级聚合、支持传输容错机制、支持方差统计特征, 确保数据在整个系统传输与存储中的机密性与完整性验证的安全功能, 而且在计算与通信开销方面也占有优势, 非常有利于部署在分层级医疗密态数据聚合与统计分析系统.

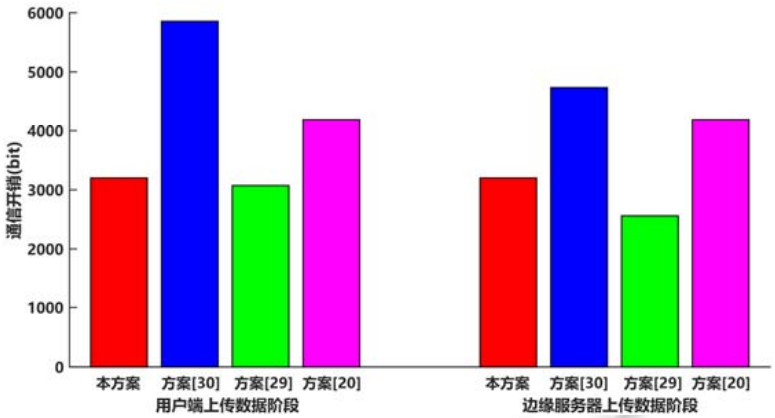

图 6 通信开销比较

7 结束语

本文设计了支持边缘服务计算的可验证医疗密态数据聚合与统计分析方案. 方案集成了云服务器与边缘服务器, 改进了 BGN 同态加密技术, 构建了密态数据分层聚合模型, 使得医疗数据分析中心只需通过两次解密就可以进行具有隐私保护的均值和方差等统计分析. 方案结合 Shamir 秘密共享技术与改进的 BGN 同态加密技术, 保证了移动终端用户医疗数据的机密性, 同时支持传输容错机制. 这样, 即使部分数据由于网络不稳定导致数据丢失或者用户不愿意上传自己敏感的医疗数据, 边缘级密态数据聚合仍能正常进行; 即使敌手获得医疗数据中心的解密私钥, 在没有得到边缘服务器的控制权和足够多的密文数据时, 也无法独立解密出用户的医疗数据. 此外, 方案设计了基于身份的聚合签名算法确保各层级医疗密态数据在传输和存储过程中的完整性. 性能分析表明: 与现有类似方案相比, 本方案在性能方面具有明显优势. 特别是随着参与密态数据聚合的用户数量增加, 医疗数据中心只需要几乎恒定量的计算开销就可以判断出密态数据在传输和存储过程中是否遭到篡改、替换或销毁.

References:

[1] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Computer Networks, 2010, 54(15): 2787−2805.

[2] Stankovic JA. Research directions for the Internet of things. IEEE Internet of Things Journal, 2014, 1(1): 3−9.

[3] Ni MX, Zhang Q, Tan HY, Luo WM, Tang XX. Smart healthcare: From IoT to cloud computing. Scientia Sinica Informationis, 2013, 43(4): 515−528 (in Chinese with English abstract).

[4] Yu YX, Qin C. Expatiation on the basic ideas of smart grid. Scientia Sinica Informationis, 2014, 44(6): 694−701 (in Chinese with English abstract).

[5] Huang QY, Li ZY, Xie WT, Zhang Q. Edge computing in smart homes. Journal of Computer Research and Development, 2020, 57(9): 1800−1809 (in Chinese with English abstract).

[6] Nordrum A. The Internet of fewer things. IEEE Spectrum, 2016, 53(10): 12−13.

[7] Su M, Wu B, Fu AM, Yu Y, Zhang GX. Assured update scheme of authorization for cloud data access based on proxy re-encryption. Ruan Jian Xue Bao/Journal of Software, 2020, 31(5): 1563−1572 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5676.htm [doi: 10.13328/j.cnki.jos.005676]

[8] Abbas N, Zhang Y, Taherkordi A, Skeie T. Mobile edge computing: A survey. IEEE Internet of Things Journal, 2018, 5(1): 450−465.

[9] Chu W. Application of data encryption technology in computer network security. Journal of Physics: Conf. Series, 2019, 1237(2): 1−5.

[10] Fu S, Jiang Q, Ma JF. A privacy-preserving data aggregation scheme in wireless sensor networks. Journal of Computer Research and Development, 2016, 53(9): 2030−2038 (in Chinese with English abstract).

[11] Zhang XJ, Zhang JW, Huang C, Tang W. Verifiable statistical analysis scheme for encrypted medical data in cloud storage. Computer Engineering, 2021, 47(6): 32−37 (in Chinese with English abstract).

[12] Han S, Zhao S, Li QH, Ju CH, Zhou WL. PPM-HDA: Privacy preserving and multifunctional health data aggregation with fault tolerance. IEEE Trans. on Information Forensics and Security, 2016, 11(9): 1940−1955.

[13] Li RN, Sturtivant C, Yu JG, Cheng XZ. A novel secure and efficient data aggregation scheme for IoT. IEEE Internet of Things Journal, 2018, 6(2): 1551−1560.

[14] Ara A, Al-Rodhaan M, Tian Y, Al-Dhelaan A. SPPDA scheme based on bilinear ELGamal cryptosystem. IEEE Access, 2017, 5: 12601−12617.

[15] Tang WJ, Ren J, Deng K, Zhang YX. Secure data aggregation of lightweight e-healthcare IoT devices with fair incentives. IEEE Internet of Things Journal, 2019, 6(5): 8714−8725.

[16] Lu RX, Liang XH, Li X, Lin XD, Shen XM. EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications. IEEE Trans. on Parallel and Distributed Systems, 2012, 23(9): 1621−1631.

[17] Zhang L, Wu QH, Domingo-Ferrer J, Qin B, Hu CY. Distributed aggregate privacy-preserving authentication in VANETs. IEEE Trans. on Intelligent Transportation Systems, 2017, 18(3): 516−526.

[18] Kang JW, Yu R, Huang XM, Wu MQ, Maharjan S, Xie SL, Zhang Y. Blockchain for secure and efficient data sharing in vehicular edge computing and networks. IEEE Internet of Things Journal, 2019, 6(3): 4660−4670.

[19] Wang ZW. An identity-based data aggregation protocol for the smart grid. IEEE Trans. on Industrial Informatics, 2017, 13(5): 2428−2435.

[20] Li X, Liu SP, Wu F, Kumari S, Rodrigues RJ. Privacy preserving data aggregation scheme for mobile edge computing assisted IoT applications. IEEE Internet of Things Journal, 2019, 6(3): 4755−4763.

[21] Chan THH, Shi E, Song D. Privacy-preserving stream aggregation with fault tolerance. In: Proc. of the Int'l Conf. on Financial Cryptography and Data Security. 2012. 200−214.

[22] Benhamouda F, Joye M, Libert B. A new framework for privacy preserving aggregation of time-series data. ACM Trans. on Information & System Security, 2016, 18(3): 1−21.

[23] Abbas N, Zhang Y, Taherkordi Y, Skeie T. Mobile edge computing: A survey. IEEE Internet of Things Journal, 2018, 5(1): 450−465.

[24] Shi WS, Cao J, Zhang Q, Li YHZ, Xu LY. Edge computing: Vision and challenges. IEEE Internet of Things Journal, 2016, 3(5): 637−646.

[25] Boneh D, Goh EJ, Nissim K. Evaluating 2-DNF formulas on ciphertexts. In: Proc. of the Theory of Cryptography. 2005. 325−341.

[26] Abdallah A, Shen XM. A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid. IEEE Trans. on Smart Grid, 2016, 9(1): 396−405.

[27] Luo ET, Wang GJ, Liu Q, Meng DC, Tang YY. Privacy preserving friend discovery of matrix confusion encryption in mobile social networks. Ruan Jian Xue Bao/Journal of Software, 2019, 30(12): 3798−3814 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5601.htm [doi: 10.13328/j.cnki.jos.005601]

[28] Li ZY, Gui XL, Gu XJ, Li XS, Dai HJ, Zhang XJ. Survey on homomorphic encryption algorithm and its application in the privacy-preserving for cloud computing. Ruan Jian Xue Bao/Journal of Software, 2018, 29(7): 1830−1851 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5354.htm [doi: 10.13328/j.cnki.jos.005354]

[29] Zhang K, Liang XH, Baura M, Lu RX, Shen XM. PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs. Information Sciences, 2014, 284: 130−141.

[30] Wang HQ, Wang ZW, Domingo-Ferrer J. Anonymous and secure aggregation scheme in fog-based public cloud computing. Future Generation Computer Systems, 2018, 78: 712−719.

[31] Shamir A. How to share a secret. Communications of the ACM, 1979, 22(11): 612−613.

[32] Gentry C, Ramzan Z. Identity-based aggregate signatures. In: Proc. of the Practice and Theory in Public Key Cryptography (PKC), Vol.3958. 2006. 257−273.

附中文参考文献:

[3] 倪明选, 张黔, 谭浩宇, 罗吴蔓, 汤小溪, 智慧医疗——从物联网到云计算. 中国科学: 信息科学, 2013, 43(4): 515−528.

[4] 余贻鑫, 秦超, 智能电网基本理念阐释. 中国科学: 信息科学, 2014, 44(6): 694−701.

[5] 黄倩怡, 李志洋, 谢文涛, 张黔. 智能家居中的边缘计算. 计算机研究与发展, 2020, 57(9): 1800−1809.

[7] 苏铓, 吴槟, 付安民, 俞研, 张功萱. 基于代理重加密的云数据访问授权确定性更新方案. 软件学报, 2020, 31(5): 1563−1572. http://www.jos.org.cn/1000-9825/5676.htm [doi: 10.13328/j.cnki.jos.005676]

[10] 付帅, 姜奇, 马建峰. 一种无线传感器网络隐私保护数据聚合方案. 计算机研究与发展, 2016, 53(9): 2030−2038.

[11] 张晓均, 张经伟, 黄超, 唐伟. 可验证的云存储医疗加密数据统计分析方案. 计算机工程, 2021, 47(6): 32−37.

[27] 罗恩韬, 王国军, 刘琴, 孟大程, 唐雅媛. 移动社交网络中矩阵混淆加密交友隐私保护策略. 软件学报, 2019, 30(12): 3798−3814. http://www.jos.org.cn/1000-9825/5601.htm [doi: 10.13328/j.cnki.jos.005601]

[28] 李宗育, 桂小林, 顾迎捷, 李雪松, 戴慧珺, 张学军. 同态加密技术及其在云计算隐私保护中的应用. 软件学报, 2018, 29(7): 1830−1851. http://www.jos.org.cn/1000-9825/5354.htm [doi: 10.13328/j.cnki.jos.005354]

张晓均(1985－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为密码学, 信息安全.

谷大武(1970－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为密码学, 计算机安全.

张经伟(1995－), 男, 硕士, 主要研究领域为密码学, 医疗数据安全.

张源(1991－), 男, 博士, 研究员,主要研究领域为网络空间安全, 数据安全, 区块链.

黄超(1995－), 男, 硕士, 主要研究领域为密码学, 智能电网数据安全.