基于分布式存储的外包EHR隐私保护分类审计方案

张晓旭[1,2]，陈宇辰[1,2]，哈冠雄[1,2]，贾春福[1,2]
（1.南开大学网络空间安全学院，天津 300350；2.天津市网络与数据安全技术重点实验室，天津 300350）

摘　要：随着电子医疗领域的发展，电子健康记录（EHR）常被外包到雾节点上进行分布式存储以提升可靠性。EHR中包含大量隐私信息，然而数据外包易造成安全隐患，可能破坏EHR的完整性与隐私性。为确保EHR的安全存储，提出一种高效的基于EHR分类的分布式数据完整性审计方案。该方案将分类标签与布隆过滤器相结合以提升审计效率，利用Shamir秘密共享完成分布式审计，采用属性基可搜索加密以保护EHR所属类别信息的隐私。实验结果表明，所提方案的通信开销和计算开销较低。

关键词：分布式存储；数据审计；EHR；Shamir秘密共享；属性基可搜索加密
中图分类号：TP309.2
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024169

Classification auditting scheme for privacy protection of outsourced EHR based on distributed storage

ZHANG Xiaoxu[1,2], CHEN Yuchen[1,2], HA Guanxiong[1,2], JIA Chunfu[1,2]
1. College of Cyber Science, Nankai University, Tianjin 300350, China
2. Tianjin Key Laboratory of Network and Data Security Technology, Tianjin 300350, China

Abstract: With the development of the electronic medical field, Electronic health record (EHR) are often outsourced to fog nodes for distributed storage to enhance reliability. EHR contains a large amount of private information, however, data outsourcing can create security risks, potentially compromising the integrity and privacy of EHR. In order to ensure the secure storage of EHR, an efficient distributed data integrity auditing scheme based on EHR classification was proposed. The scheme combined classification labels with Bloom filters to enhance auditing efficiency, Shamir's secret sharing was used for distributed auditing, and attribute-based searchable encryption was employed to protect the privacy of EHR' category information. Experimental results show that the proposed scheme has low communication and computation costs.

Keywords: distributed storage, data auditing, EHR, Shamir secret sharing, attribute-based searchable encryption

0　引言

近几年来，分布式存储技术发展迅速，并成为一种新型的存储数据的方式。它不仅为用户提供了方便快捷的数据存储服务，同时兼具使用成本低、可扩展、灵活等优点[1-2]。在电子医疗系统中，电子健康记录（EHR, electronic health record）主要用于追踪用户的身体数据以判断其健康状况或用于医学研究[3-4]。鉴于医院每天有很多患者，许多医疗机构面临着如何管理并存储好大量电子

收稿日期：2024–06–18；修回日期：2024–09–09
通信作者：贾春福，cfjia@nankai.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2018YFA0704703）；国家自然科学基金资助项目（No.61972215, No.61972073, No.62172238）
Foundation Items：The National Key Research and Development Program of China (No.2018YFA0704703), The National Natural Science Foundation of China (No.61972215, No.61972073, No.62172238)

病历的挑战。一些医院将其本地电子病历外包到分布式雾存储服务器端，以减轻本地存储管理的负担[5]。

然而，大量的电子病历中包含高度敏感的数据，例如患者的身份证号、手机号、病情等信息[6]。这些隐私信息一旦泄露，患者的人身安全将会遭到威胁。此外，雾服务器可能会恶意泄露电子病历，暴露患者的隐私信息以获得盈利。因此，保护EHR的隐私至关重要。为了降低网络时延与网络带宽成本，基于雾节点的分布式场景受到广泛关注[7]。雾计算作为云计算的扩展，具有异构性、低时延、位置感知、可移植性等优点。然而，数据所有者将电子病历外包给雾节点时，必须检查电子病历的完整性，并保证电子病历的隐私性。因此，现阶段迫切需要设计一种在分布式场景中保护外包电子病历的完整性和隐私性的方案。文献[8]提出一种利用匹配令牌机制来保护类别信息的方法，然而该方法会将电子病历类别信息泄露给恶意云。

数据审计常被用于检测数据完整性，然而，现有的大多数据审计方案在效率上有待提高。数据分类具有节省存储空间和提高检索效率等优点，与此同时，在数据审计方面，数据分类管理将大大提升审计效率。在大型基于云-雾的EHR系统中，如何将大量患者的病例信息分类存储并实行高效的分布式数据审计是值得探究的。

与此同时，在大型EHR系统中，为了提高数据所有者（DO, data owner）的满意度和忠诚度，实现个性化服务，系统应根据数据所有者历史行为进行预测，使得数据所有者在网络拥塞时可以提前访问已经被加载到雾节点中的相关数据。由此，系统可以为数据所有者提供更加便捷的服务。

本文的主要贡献如下。

1) 提出了一种分布式数据收集场景中的EHR分类审计模型，提升了审计效率，保护了电子病历类别信息的隐私性。利用密文策略属性基可搜索公钥加密方法完全保护了类别信息。

2) 利用推荐系统中的奇异值分解（SVD, singular value decomposition），根据数据所有者的历史行为进行预测，数据所有者可以预先访问已被存储到雾节点中的相关数据。在网络流量高峰期时，可降低数据块吞吐量。

3) 本文方案同时支持数据的动态更新，利用带有哈希链表的计数式布隆过滤器存储数据，实现数据所有者对数据块的动态操作。

4) 在现有的基于分布式存储的数据审计方案中，恶意审计员因一己私利而窃取患者隐私数据的情况屡见不鲜。鉴于此，本文方案利用密文策略属性基加密方法实现对恶意审计员的双重访问控制机制，以防止其窃取EHR中的隐私数据及EHR类别信息。

1　相关工作

在当今互联网飞速发展的时代，数据管理与共享为人们的生活带来了极大的便捷，人们的生活方式也因此变得丰富。数据管理是指对数据进行合理规划、组织、存储、处理的过程[9]。数据共享可以促进跨领域合作，减少数据资源的浪费，充分利用信息资源的同步性。目前，电子医疗应用领域中已经提出了许多数据管理与共享的系统，Bera等[10]提出了一种基于区块链的访问控制方案，该方案实现了跨区域数据共享；Yu等[11]提出了一种基于轻量级数据管理和密钥管理的细粒度访问控制方案；He等[12]提出将边缘计算与可扩展访问控制标记语言模型相结合来提供细粒度的访问控制。然而，上述方案没有针对EHR中患者的个人隐私信息进行隐私保护，例如身份证号、手机号、病情等重要信息，个人隐私信息一旦遭到泄露，极有可能威胁到患者的财产及人身安全。

鉴于此，电子医疗系统中的数据隐私保护受到了广泛关注。近些年，一些保护远程数据存储安全的方案相继被提出。为了验证远程存储数据的完整性，Ateniese等[13]首先提出了可证明数据占有（PDP, provable data possession），该模型能检查远程云服务器是否完整地保留一个数据文件。PDP允许审计人员在不下载云数据的情况下审计云数据的完整性，大大节省了系统的通信开销。Juels等[14]定义了可检索性证明（PoR, proof of retrievability），该模型可以检索存储在云中的数据，并保证数据的完整性。与PDP不同，PoR侧重于在数据被验证为不完整时进行数据恢复。王子园等[15]结合在线/离线签名思想，提出了一种在分布式环境下基于无证书公钥密码的审计方案。

雾计算于2012年被首次提出，并引起了业界的广泛关注，进一步推动了5G网络的蓬勃发展。

在基于雾计算的EHR系统中，雾节点拥有和云服务器一样强大的计算资源，并连接到靠近患者的移动设备[16]。雾节点上也可以相应缓存数据，为用户提供更加高效的服务。同时，与在云服务器中存储数据类似，在高度分布式、不稳定的雾计算环境中，数据同样可能由于人为或自身因素遭到破坏[17]，从而不再完整。因此，提出一种高效的基于分布式场景的EHR审计方案是十分必要的。

在每日需要大量数据动态更新的EHR系统中，单单是静态审计方案已无法满足现状。与此同时，许多方案采用合适的数据存储结构来提高动态数据更新的效率。Wang等[18]提出了一种利用默克尔哈希树（MHT, Merkle hash tree）结构进行数据动态更新公共云审计方案（DPDP-MHT）。然而，该方案不能完全保护数据的机密性，因此该方案不适用于审计隐私性较强的EHR数据。Ke等[19]提出了基于分布式存储的动态数据审计方案。Zhu等[20]提出了一种基于索引哈希表的公共PDP方案，该方案利用对每个数据块的更新操作代替对整个数据文件的更新操作，大大提升了通信效率。然而，查找和更新操作效率较低。Yang等[21]基于双链表和位置阵列组成的动态结构提出了一种动态云审计（DAP）方案，该方案可以充分保证数据块的完整性和机密性。遗憾的是，该结构与索引哈希表类似，仍存在数据动态更新时效率较低的问题。Zhang等[22]提出了一种将基于保护患者身份隐私的公共审计（CIPPPA）方案，任何第三方无法将患者的身份与其EHR对应，充分维护了患者的权益。然而，该方案无法支持数据动态更新操作。

但是，上述方案没有将数据分类存储功能与一个可以高效更新的数据结构联系在一起，如果这样，就可以同时提升数据存储、数据完整性检测、数据更新的效率。实际上，分类存储使得EHR管理更加便捷，不仅提高了存储效率，同时提升了数据查找效率。然而，当大量分类EHR存储在多个雾节点中时，如何充分保护EHR的类别信息及EHR自身的隐私是一个值得思考的课题。

文献[8]实现了在云端场景中的EHR分类管理并审计，然而其在保护类别信息的隐私上存在安全漏洞，针对此，本文提出了一种在分布式场景中分类存储EHR并审计的方案，安全性分析表明，本文方案充分保护了EHR类别信息及其自身的隐私。文献[8]虽然考虑了上述问题，但并没有充分保护EHR类别信息，由于双线性对的性质，恶意云还有可能获取类别信息，一旦泄露，将造成严重的影响。本文方案将分类存储和带有哈希链表的计数式布隆过滤器结构相结合，同时利用属性基可搜索加密方法充分保护EHR类别信息及其自身的隐私。

2 预备知识

2.1 计数式布隆过滤器

布隆过滤器可以看成一个固定大小的二进制向量。然而，最初的布隆过滤器版本无法实现数据动态操作。文献[23]提出了计数式布隆过滤器（CBF, counting Bloom filter）。通过插入和删除元素，每个位值将动态更新。如图1所示，计数式布隆过滤器$\boldsymbol{E}$是长度为h的整数向量，可以看作由多个布隆过滤器组成，即$\boldsymbol{E}[c]=\sum_{x} b_i[c]$，其中$1 \leqslant c \leqslant l$。每个位上的counter值即待插入元素哈希后映射的计数值，$\boldsymbol{E}[c]$就是计数式布隆过滤器$\boldsymbol{E}$在位置c上的counter值。给定x个布隆过滤器b_i，可以通过使用向量加法运算形成$\boldsymbol{E}$，即$\boldsymbol{E}=\sum_{i} b_i$。本文方案使用带有哈希链表的计数式布隆过滤器（CBCH, counting Bloom filter and chained hash table）。其中，$A_t=H(\mathrm{kk}_j)(1 \leqslant t \leqslant \mathrm{oj})$用来存储EHR所属类别信息（假设共有oj个类别），将其置于CBF头部，代表该CBF隶属于同一个类别。链表被附加到每个桶中，桶用于存储待插入的元素。CBCH支持3种操作：插入、删除和查询。将待插入元素的类别信息kk_j与A_t匹配，确定其隶属于哪个CBF，如果不存在可以匹配的A_t，则根据其类别kk_j计算$A_t=H(\mathrm{kk}_j)(1 \leqslant t \leqslant \mathrm{oj})$。数据插入具体操作如下。

匹配类别信息A_t，如果A_t已经存在，则将EHR数据块哈希后分别映射到对应的位上，查询数据块是否存在于该位上的哈希链表中，不存在则counter值+1，在该位相应的哈希链表的桶中插入该数据块元素；如果A_t不存在，则计算$A_t=H(\mathrm{kk}_j)(1 \leqslant t \leqslant \mathrm{oj})$，创建新的CBF，重复之前的元素插入操作。

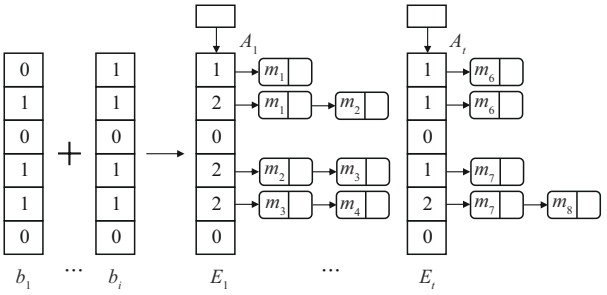

图1　CBF原理及动态操作

2.2　双线性对

假设群G_0、G_1和G_T是素数p阶的循环群，其中，群G_0中的运算为加法，G_1中的运算为乘法。如果满足以下条件，则双线性映射$e{:}G_0 \times G_1 \to G_T$成立。

1) 双线性：对于所有$g_x \in G_0$，$g_y \in G_1$，$(b,r) \in Z_p^*$，有$(g_x^b, g_y^r) = (g_x, g_y)^{br}$，其中$g_x \neq g_y$。

2) 非退化性：存在$g_x \in G_0$，$g_y \in G_1$，使得$e(g_x, g_y) \neq 1$。

3) 可计算性：对于任意$g_x \in G_0$，$g_y \in G_1$，存在一个多项式时间算法来计算$e(g_x, g_y)$，称e为双线性映射。

2.3　访问结构

密文策略属性基加密是一种基于属性的加密方案，它允许数据所有者定义访问结构，只有满足特定属性的用户才能解密数据。具体来说，将密文与访问结构关联起来，只有拥有与访问结构匹配的属性集的用户才能成功解密密文。

假设有一组参与方$F = \{F_1, F_2, \cdots, F_n\}$，对于$\forall G, H$，集合$T \subseteq 2^F$是单调的。如果$G \in T$且$G \subseteq H$，那么存在$H \in T$。则$\{F_1, F_2, \cdots, F_n\}$的非空子集$T$（均为单调集合）为访问结构（均为单调访问结构），即$T \subseteq 2^{\{F_1, F_2, \cdots, F_n\}} \setminus \{p\}$。其中，$T$是一个被授权的集合，其他集合代表未被授权的集合[24]。

2.4　线性秘密共享结构

如果有关参与方集合$F = \{F_1, F_2, \cdots, F_n\}$中的一个秘密共享方案V满足以下条件，则称该方案为在Z_p上的线性秘密共享方案[24]。

1) 每个实体的秘密共享份额构成Z_p上的一个向量。

2) 对于秘密共享方案V，首先生成一个矩阵$\boldsymbol{M}^{d \times j}$，矩阵中的每行为$i\,(i = 1,2,\cdots,d)$。定义映射$\varepsilon{:}\{1,2,\cdots,d\}$，将$\boldsymbol{M}^{d \times j}$中的每一行映射到$\varepsilon(i)$，每行向量为一个属性。设向量$\boldsymbol{v} = (s, y_2, \cdots, y_j)$，其中，$s \in Z_p$是秘密值，$y_2, \cdots, y_j \in Z_p$是随机值，$\boldsymbol{M}_v$是一个包含$d$个秘密份额的向量。$\boldsymbol{\lambda}_i = (\boldsymbol{v}\boldsymbol{M}_i)$代表新$\varepsilon(i)$的秘密共享份额，即访问策略的矩阵，其中，$\boldsymbol{M}_i$指矩阵的第$i$行。

由于V是访问结构T的秘密共享，设$G \in A$是一个访问授权集合，定义为$I = (i{:}\varepsilon(i) \in G)$。如果$\lambda_i$是秘密$s$的有效份额，属性密钥由一组可在多项式时间内获得的常向量$\boldsymbol{w}_i \in Z_p$组成，即式$\sum_{i \in I} \boldsymbol{w}_i \boldsymbol{\lambda}_i = G$成立则解密成功。

2.5　CDH (computation Diffie-Hellman) 假设

设$G_0 = \langle g_0 \rangle$是一个素数$p$阶循环群，其中$g_0$是群$G_0$的生成元。对于未知的$a, b \in Z_p$，给出$ag$和$bg$，如果敌手可以获得$abg$，则可以打破CDH困难问题[18]。

对于任意概率多项式时间（PPT）敌手$\mathcal{A}$，其打破CDH问题的概率可以忽略不计，即

$$\Pr\left[\mathcal{A}_{\mathrm{CDH}}(g_0, g_0^a, g_0^b \in G_0 \to g_0^{ab} \in G_0 {:} \forall a,b \in {}_R Z_p^*)\right] \leqslant \varepsilon$$

2.6　判定性q-BDHE假设

选择一个p阶的双线性群G，其中g和h是群G的2个生成元。选择一个随机数$\alpha \in Z_p$，定义$y_{g,a,l} = \left(g_1, g_2, \cdots, g_l, g_{l+2}, \cdots, g_{2l}\right) \in G^{2l-1}$，其中$g_i = g^{(\alpha^i)}$，敌手根据输出值$Z \in \{0,1\}$进行猜测[24]。

定义ε表示求解判定性q-BDHE困难问题的优势。如果存在多项式时间算法B，使得$\left|\Pr\left[\mathrm{B}\left(g, h, y_{g,a,l}, e\left(g_l, h\right)\right)=0\right]-\Pr\left[\mathrm{B}\left(g, h, y_{g,a,l}, Z\right)=0\right]\right| \geqslant \varepsilon$，那么敌手在群$G$和$G_T$中求解判定性q-BDHE问题的优势是可忽略的。

3　方案流程

3.1　系统架构

本文的系统架构如图2所示，其中每个实体拥有不同的职责。

密钥生成中心（KGC, key generation center）负责发布系统公共参数，同时为数据所有者分发公钥和私钥，为第三方审计员（TPA, third-party auditor）分发其属性密钥。

TPA与雾节点交互，完成整个审计过程。TPA

向雾节点发送挑战，属性满足密文访问结构的TPA可以成功解密，与密文中的类别信息关键字进行匹配。等待云发送proof后，TPA验证proof，并将审计结果返回给数据所有者。

雾节点与雾节点之间共享信息，并负责存储盲化后的数据块、加密的类别信息、标签集合等。雾节点收到TPA发送的挑战请求后，将包含类别信息关键字的密文返回给TPA，以验证TPA是否拥有访问权限。雾节点根据挑战信息将生成的proof发送给合法的TPA，然后根据数据所有者的历史行为为其推荐可以预访问的内容。

数据所有者（DO）即用户，可以是医院的工作人员，其将盲化因子通过秘密共享分发给各个雾节点，接着将盲化后的数据块、加密的类别信息、标签集合等上传到雾节点。雾节点根据DO的历史行为为其推荐相关访问数据，便于DO可以提前访问。

3.2 方案定义

本文方案主要分为以下几个步骤。

1) 系统建立$(1^{\lambda}) \to P_{\text{pub}}$。该算法由KGC执行，以安全参数$\lambda$作为输入，输出系统公共参数$P_{\text{pub}}$。

2) 密钥生成$\left(\alpha,b,\text{Hdd}_j,j\right) \to$PK,MSK,ASK。该算法由KGC执行，输入参数$\alpha,b \in Z_p$，然后为DO输出PK和MSK；输入类别信息Hdd_j和属性j，然后为TPA输出属性密钥ASK。

3) 准备阶段$\left(m_{l,w}^{(j)},\boldsymbol{M},\Re\right) \to m_{l,w}^{(j)'}$,CT。该算法由DO执行，输入EHR数据块明文$m_{l,w}^{(j)}$以及访问策略$(\boldsymbol{M},\Re)$，输出盲化后的明文$m_{l,w}^{(j)'}$以及根据访问策略加密后的类别信息密文CT。

4) 生成标签$\left(\text{Hdd}_j,\text{ID}^{(j)},\left\{E_l = m_{l,w}^{(j)}\right\}_{w=1}^{n},\text{MSK}\right) \to$ Cf,CTT。该算法由DO执行，输入参数类别信息Hdd_j、EHR的ID信息$\text{ID}^{(j)}$、$\left\{E_l = m_{l,w}^{(j)}\right\}_{w=1}^{n}$以及私钥MSK，输出云标签Cf以及TPA标签CTT。

5) 生成chal$(t_j) \to$ chal。该算法由TPA执行，输入随机值t_j，输出挑战信息chal。

6) 匹配验证$(\text{CT},\text{ASK}) \to$ T/F。该算法由TPA执行，输入类别信息密文CT和TPA属性密钥ASK，返回验证结果T/F。

7) 生成proof (chal) $\to$ proof。该算法由雾节点执行，输入挑战信息chal，输出证明信息proof。

8) 验证verify (proof) $\to$ T/F。该算法由TPA执行，输入证明信息proof，输出proof的验证结果T/F。

9) 用户推荐。根据用户的历史行为，为其推荐可以预先访问的内容，避免网络拥塞导致访问时延。

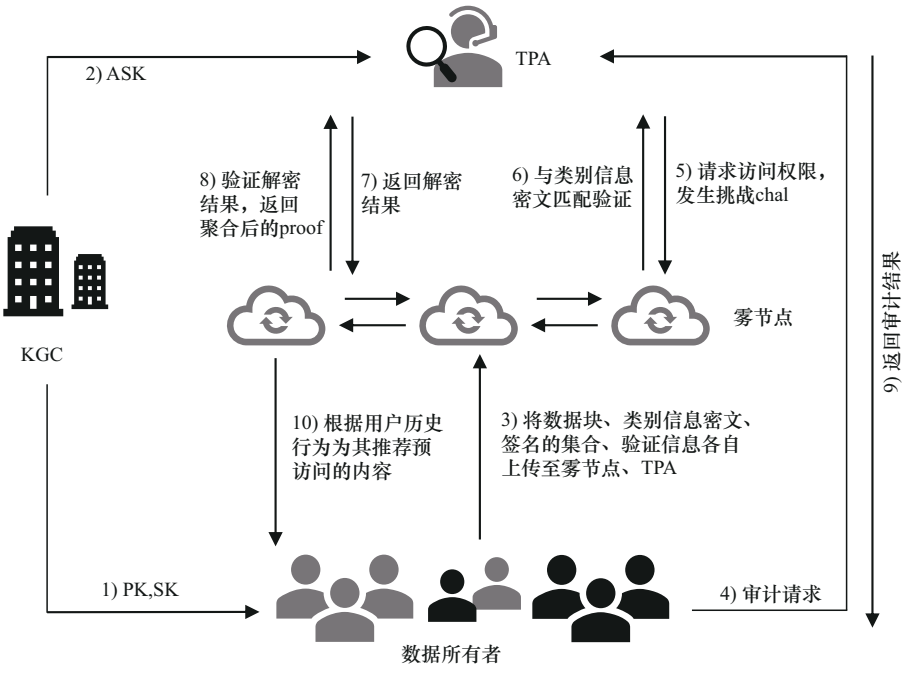

图2 系统架构

3.3　威胁模型

本文假设雾节点和TPA都是好奇的，但是并不会共谋。

1) 雾节点一旦获取到类别信息，可能将其恶意篡改。因此，确保雾节点无法获取类别信息。

2) 应确保没有审计权限的TPA无法检测雾节点上的EHR数据块完整性。

3) 假设雾节点可能试图伪造电子病历和相应的签名来欺骗审计员TPA。如果伪造成功，那么TPA的审计结果将不再真实。因此，确保雾节点不能伪造电子病历的签名是非常重要的。

3.4　文献[8]的安全缺陷

在保护类别信息隐私方面，文献[8]设计了一个令牌来精确定位EHR的类别信息。然而，其在为TPA设计令牌$t_1 = T_1^c T_2^{cH}$（其中，T_1和T_2是公共系统参数，H为类别信息关键字）后，就将此令牌发送给云，恶意云可能通过双线性对的性质$e\left(g_0^{cH},g_1\right) = e\left(g_0^c,g_1^H\right)$从该令牌猜测出类别信息关键字$H$。

解析：$e\left(T_1^c T_2^{cH},g_1\right) = e\left(T_1 T_2^H,g_1^c\right) = T_1 e\left(T_2^H,g_1^c\right)$。由于已知$T_2$，可猜测出$H$。

4　具体方案

如图3所示，假设所有病例信息隶属于oj个科室种类。每个种类kk包括k条EHR，EHR的身份标识记为$\mathrm{ID}^{(j)}$，每条EHR又包含n个数据块，即每个种类包含$k \times n$个数据块。

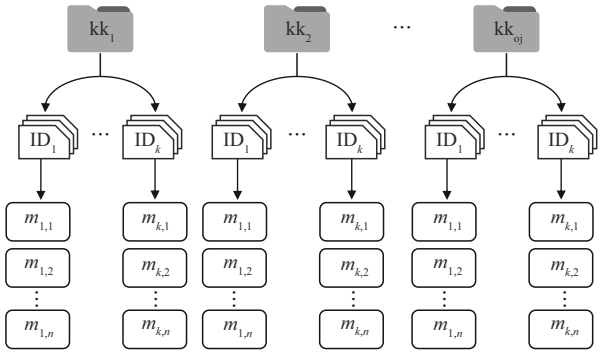

图3　EHR结构

4.1　方案构造

系统建立。设定一个系统安全参数，系统建立算法输出2个双线性映射，分别为$e{:}G_0 \times G_1 \to G_T$，$e_1{:}G_2 \times G_2 \to G_3$；建立2个防碰撞的哈希函数，分别为$H_1{:}\{0,1\}^* \to G_0$，$H_2{:}\{0,1\}^* \to Z_p$。其中，$G_0$、$G_1$、$G_2$、$G_3$和$G_T$都是素数$p$阶乘法循环群，$g_0$和$g_1$分别是群$G_0$和$G_1$的生成元，$g$是$G_2$的生成元，随机选择群元素$h_1,h_2,\cdots,h_U$。组成系统公共参数$P_{\mathrm{pub}} = \left\{p,e,e_1,G_0,G_1,G_2,G_3,G_T,H_1,H_2,h_1,h_2,\cdots,h_U\right\}$。

密钥生成。PKG选择随机数$\alpha,b \in Z_p$，计算$\mathrm{UU} = g_0^{\alpha}$，$\mathrm{VV} = g_1^{\alpha}$，$\mathrm{ZZ} = g^a$，使得$\mathrm{PK} = \{\mathrm{UU},\mathrm{VV},\mathrm{ZZ}\}$，$\mathrm{MSK} = \left\{\alpha,a,g^b\right\}$。随机选择$r \in Z_p$，计算$I = g^b g^{ar}\mathrm{Hdd}_j$（其中$\mathrm{Hdd}_j$为TPA发送的chal中包含的类别信息），$L = g^r$。对于属性集$S = S_1,S_2,\cdots,S_k$中的每个属性$j \in [1,n]$，计算$I_j = t_j^r$。最后，输出属性密钥$\mathrm{ASK} = I,L,\left\{I_j\right\}_{j \in [1,N]}$，并将其发送给TPA。

准备阶段。此阶段假定雾节点上传所有种类kk，为了简化过程，假设有种类$\mathrm{kk}_j\left(1 \leqslant j \leqslant \mathrm{oj}\right)$，则全部种类标记为$\left\{\mathrm{kk}_j\right\}_{j=1}^{t}\left(t \leqslant \mathrm{oj}\right)$。每个种类$\mathrm{kk}_j$中包括$k$条EHR，即$E_l\left(l \in k\right)$，每条EHR又包含$n$个数据块，假设有数据块$m_{l,w}^{(j)}$，$w \in n$，EHR的身份标识记为$\mathrm{ID}^{(j)}$，其中$E_l = \left\{m_{l,w}^{(j)}\right\}_{w=1}^{n}$。其他种类重复此过程。同时，在此阶段中为每条EHR中的数据块挑选一个盲化因子$v \in Z_p$，由于一共有k条EHR，分别对应盲化因子$v_1,v_2,\cdots,v_k$。为了简化过程，假设盲化第l条EHR中的数据块，其他EHR中的数据块依次类推。盲化数据块$m'_{l,w} = \dfrac{m_{l,w}}{v}$。

随机选择$n-1$个系数$A_1,A_2,\cdots A_{n-1} \in Z_p$，构造出一个$c-1$阶多项式$\chi(x) = A_0 + A_1 x + \cdots + A_{n-1}x^{c-1} \in Z_p$，其中$v = A_0$。一共有$e$个雾节点，即$\left\{\mathrm{fog}_1,\mathrm{fog}_2,\cdots,\mathrm{fog}_e\right\}$，其中$\mathrm{fog}_i \in Z_p$。接着，为每个雾节点计算$\mathrm{ss}_i = \chi\left(\mathrm{fog}_i\right)$，得到$\left\{\mathrm{ss}_1,\mathrm{ss}_2,\cdots,\mathrm{ss}_e\right\}$。定义$(c,e)$，其中恢复出秘密值的雾节点数量阈值为$c$，雾节点总数为$e$。

DO根据访问策略加密所有种类的关键字，随机指定一个一阶多项式$q(0) = s$，并定义访问策略$(\boldsymbol{M},\Re)$。随机选择一个向量$\boldsymbol{c} = \left(s,y_2,\cdots,y_k\right) \in Z_p$，用来加密秘密值$s$。雾节点根据$\boldsymbol{M}$中的每一行$i = 1$到$i = n$计算$\boldsymbol{\zeta}_i = \boldsymbol{c}(\boldsymbol{M})_i$，最后输出密文

$$\mathrm{CT}=\left\langle (\boldsymbol{M},\Re),C_1=e(g,g)^{bs}e\left(g,\mathrm{Hdd}_j\right)^{s},C_1'=g^{s},C_2'=g^{a-s},\left\{C_i=g^{aq_i}h_{\Re(i)}^{-s}\right\}_{1\leqslant i\leqslant n}\right\rangle \tag{1}$$

生成标签阶段。首先，DO为种类kk_j中的每个数据块$m_{l,w}^{(j)}$计算标记$\mathrm{sy}_l^{(j)}=H_2\left(\mathrm{kk}_j\middle\|\mathrm{ID}_l^{(j)}\middle\|E_l^{(j)}\right)$和签名$\sigma_{l,w}^{(j)}=\left(H_1\left(\mathrm{ID}_{l,w}^{(j)}\middle\|m_{l,w}^{(j)}\right)g_0^{m_{l,w}^{'(j)}}\right)^{\alpha}$。然后为种类$\mathrm{kk}_j$计算$\mathrm{Hdd}_j=H_2\left(\mathrm{kk}_j\right)$，其余种类中的数据块计算重复该操作。并且令$N_{l,w}^{(j)}=H_1\left(\mathrm{ID}_{l,w}^{(j)}\middle\|m_{l,w}^{(j)v}\right)$，$P_{l,w}^{(j)}=\left(H_1\left(\mathrm{ID}_{l,w}^{(j)}\middle\|m_{l,w}^{(j)}\right)\right)^{\alpha}$，将$\mathrm{Cf}=\mathrm{CT}\left\|\left(\mathrm{sy}_l^{(j)}\middle\|\left(m_{l,w}^{'(j)},\sigma_{l,w}^{(j)}\right)\right)_{l=1}^{k}\right.$发送给最适配的雾节点。同时，DO将$\mathrm{CTT}=\mathrm{Hdd}_j\left\|\left(\mathrm{sy}_l^{(j)}\middle\|\left(N_{l,w}^{(j)}\middle\|P_{l,w}^{(j)}\right)\right)_{l=1}^{k}\right.$发送给TPA。最后，DO将$\mathrm{TA}=\left\{\mathrm{Hdd}_j\left\|\left(\mathrm{sy}_l^{(j)}\middle\|\left(m_{l,w}^{(j)},\sigma_{l,w}^{(j)}\right)\right)\right.\right\}$发送至雾节点，其中$1\leqslant l\leqslant k,1\leqslant w\leqslant n$。利用$\mathrm{Hdd}_j$来更好地隐藏有关种类的信息，从而保护EHR隐私。

首先，DO向TPA发送请求。TPA向邻近雾节点发送挑战，然而将面临双重访问控制。通过策略属性基加密构建一个安全的可搜索关键字索引，只有属性满足雾节点中密文访问结构的TPA才拥有查询其类别信息关键字是否与雾节点上的关键字匹配的权利。只有种类信息关键字匹配成功的TPA才可以继续与雾节点交互审计过程。然后，TPA向雾节点发送一个访问请求，DO根据访问策略加密关键字配对信息，过程如下。

TPA输入公共参数P_{pub}，生成挑战chal。

生成chal。TPA发送t个类别$\left\{\mathrm{kk}_d\right\}_{d=1}^{t}\left(t\leqslant \mathrm{oj}\right)$数据块的挑战，随机选择$t_j$个EHR，$t_j\in\left[1,k_d\right]$，再随机选择其中的$w_c$个数据块，$w_c\in\left[1,n_d\right]$。TPA发送挑战请求$\mathrm{chal}:\Theta=\left\{\left\{P_j\right\}_{j=1}^{t},W,\left\{v_w\right\}_{w\in W}\right\}$，其中，$P_j=\left\{\mathrm{sy}_l^{(j)}\right\}_{l=1}^{t_j}$表示在$t$个类别中请求审计的EHR，$W=\left\{w_1,w_2,\cdots,w_c\right\}\in\{1,2,\cdots,n\}$表示在EHR中请求审计的数据块。

匹配验证。雾节点根据chal中请求审计的数据块类别信息验证TPA，TPA利用属性密钥ASK解密。解密过程如下。

假设与密文CT相关的属性集为S，TPA的属性集为$I=\left\{i:\Re(i)\in S\right\}\subseteq\{1,2,\cdots,l\}$。如果$\varrho_i$是秘密值$s$的秘密共享，TPA选择常数集合$\left\{w_i\in Z_p\right\}_{i\in I}$，则有$\sum_{i\in I}w_i\varrho_i=s$。接着TPA使用ASK执行解密算法

$$T=\frac{e\left(C_1',I\right)}{\prod_{i\in I}\left(e\left(C_i,L\right)e\left(C_1',I_{\Re(i)}\right)\right)^{w_i}}=\frac{e\left(g^{s},g^{b}g^{ar}\mathrm{Hdd}_j\right)}{\prod_{i\in I}\left(e\left(g^{aq_i}h_{\Re(i)}^{-s},g^{r}\right)e\left(g^{s},h_{\Re(i)}^{r}\right)\right)^{w_i}}=\frac{\mathrm{Hdd}_j e(g,g)^{bs}e(g,g)^{ars}}{\prod_{i\in I}e(g,g)^{arq_iw_i}}=\mathrm{Hdd}_j e(g,g)^{bs} \tag{2}$$

雾节点根据式(3)判断TPA请求审计的数据块种类关键字与雾节点中存储的关键字是否匹配。

$$e(ZZ,g)T=C_1e\left(C_2',g\right) \tag{3}$$

通过验证的TPA继续与雾节点交互审计过程。接着，根据chal请求中包含的数据块所对应类别kk_j关键字及相应的数据块信息，雾节点生成相应数据块的proof。

生成proof。每个雾节点根据TPA请求中的数据块计算聚合签名$\sigma_i=\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(\sigma_{l,w}^{(j)}\right)^{v_w}$，$\varphi=\mathrm{UU}^{r}$，$\theta=\mathrm{VV}^{\sum_{l\in P_j(1\leqslant j\leqslant t)}\mathrm{sy}_l^{(j)}}$，以及$\gamma_i=\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_wm_{l,w}^{'(j)}$。不少于$C$个雾节点通过聚合子秘密得到

$$\gamma=\sum_{i\in[1,c]}\left(\gamma_i\chi(0)\right)=\sum_{i\in[1,c]}\left(\gamma_i\sum_{\mathrm{fs}_i\in\Phi}\chi\left(\mathrm{fs}_i\right)\triangle_{i,\Phi}(0)\right)=\sum_{i\in[1,c]}\left(\gamma_iv\right)=\sum_{i\in[1,c]}\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_wm_{l,w}^{'(j)}v=\sum_{i\in[1,c]}\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_w\frac{m_{l,w}^{(j)}}{v}v=\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_wm_{l,w}^{(j)} \tag{4}$$

$$\sigma=\prod_{i\in[1,c]}\left(\sigma_i\right)^{\chi(0)}=\left(\sigma_i^{\sum_{\mathrm{fs}_i\in\Phi}\chi\left(\mathrm{fs}_i\right)\triangle_{i,\Phi}(0)}\right)=$$

$$\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(\sigma_{l,w}^{(j)}\right)^{v_w\sum_{\mathrm{fs}_i\in\Phi}\chi(\mathrm{fs}_i)\triangle_{\mathrm{i},\Phi}(0)}=$$

$$\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(\left(H_1\left(\mathrm{ID}_{l,w}^{(j)}\middle\|m_{l,w}^{(j)}\right)g_0^{m_{l,w}^{'(j)}}\right)^{\alpha}\right)^{vv_w}=$$

$$\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(H_1(\mathrm{ID}_{l,w}^{(j)}\|m_{l,w}^{(j)})^{v}g_0^{m_{l,w}^{(j)}}\right)^{\alpha v_w}\tag{5}$$

由一起恢复秘密值的最后一个雾节点得到 $\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_w m_{l,w}^{(j)}$，最后一个雾节点接着计算 $\gamma_1=\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_w m_{l,w}^{(j)}+rH_2(\varphi)$，其中 $r\in Z_p^*$。雾节点将生成的 $\mathrm{proof}_R=\{\sigma,\varphi,\theta,\gamma_1\}$ 发送给通过验证的TPA。

检查 proof。TPA 收到 proof 后，计算 $o=\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(N_{l,w}^{(j)}\right)^{v_w}$ 并通过式(6)是否成立来判断雾节点发送的proof是否正确。

$$e\left(\sigma\varphi^{H_2(\varphi)},g_1^{\sum_{l\in P_j(1\leqslant j\leqslant t)}\mathrm{sy}_l^{(j)}}\right)=e\left(og_0^{\gamma_1},\theta\right)\tag{6}$$

用户推荐。假设有一个与用户相关的矩阵 $\boldsymbol{J}_{m\times k}$，另一个是与物品（EHR）相关的矩阵 $\boldsymbol{K}_{k\times n}$，根据上述2个矩阵来对初始评分矩阵 $\mathbf{MM}_{m\times n}$（评分可以被看作用户和EHR之间的联系）进行重构，即

$$\mathbf{MM}_{m\times n}=\boldsymbol{J}_{m\times k}\boldsymbol{K}_{k\times n}=\begin{bmatrix}z_{11}&z_{12}&\cdots&z_{1n}\\ \vdots&\vdots&\ddots&\vdots\\ z_{m1}&z_{m2}&\cdots&z_{mn}\end{bmatrix}\tag{7}$$

其中，k 是隐向量的个数，$\mathbf{MM}$ 是初始评分矩阵，z_{ui} 是真实评分，m 是用户的数量，n 是物品的数量，$\boldsymbol{J}_u$ 是用户的隐向量，$\boldsymbol{K}_i$ 是物品的隐向量。本文方案利用SVD，引入用户的历史行为对物品的影响，交给模型去学习，进而提升推荐的准确性。将用户的历史行为作为一个隐式反馈信息，作用于物品的隐向量，表示用户之前对一些物品的打分行为对未打分物品的影响。其中，SVD的预测打分函数为

$$\hat{m}_{u,i}=\mu+b_u+b_i+\boldsymbol{K}_i^{\mathrm{T}}\left(\boldsymbol{J}_u+\left|N(u)\right|^{-\frac{1}{2}}\sum_{j\in N(u)}\boldsymbol{y}_j\right)\tag{8}$$

其中，$m_{u,i}$ 表示用户 u 对物品 i 的实际评分，$\hat{m}_{u,i}$ 表示预测评分，(u,i) 表示训练数据集 $R=\{(u,i)\in m_{u,i}\}$ 中的用户−物品评分对，μ 表示整体平均评分，b_u 表示用户 u 的偏差，b_i 表示物品 i 的偏差，$N(u)$ 表示用户 u 评过分的物品集合，即 $\left|N(u)\right|^{-\frac{1}{2}}\boldsymbol{K}_i^{\mathrm{T}}\sum_{j\in N(u)}\boldsymbol{y}_j$ 为隐式反馈矩阵，$\boldsymbol{y}_i$ 为具有隐式反馈信息的物品的隐向量。

为了防止模型过拟合，加入正则化系数 $\lambda_1,\lambda_2,\cdots$，则SVD的损失函数为

$$l=\sum_{(u,i\in R)}\left(m_{u,i}-\mu-b_u-b_i-\boldsymbol{K}_i^{\mathrm{T}}\left(\boldsymbol{J}_u+\left|N(u)\right|^{-\frac{1}{2}}\sum_{j\in N(u)}\boldsymbol{y}_j\right)\right)^2+\lambda\left(\sum_{u\in R}\left(b_u^2+\left\|\boldsymbol{J}_u\right\|^2\right)+\sum_{i\in R}\left(b_i^2+\left\|\boldsymbol{K}_i\right\|^2+\left\|\boldsymbol{y}_i\right\|^2\right)\right)\tag{9}$$

根据学习率，各参数相应进行更新。根据top-K聚合算法，在预测矩阵中依次选择 K 个预测值最大的物品并聚合，然后推荐给用户。如果该 K 个预测值相等，则计算来自所有用户的该物品的得分平均值。将具有较高得分平均值的物品推荐给用户 u。

4.2　扩展功能

本文方案还支持数据动态更新的功能，其中包括数据块的增加、修改、删除。假定需要增加的数据块是种类 kk_j 中的第 $\mathrm{ID}_k^{(j)}$ 个EHR中的数据块 $E_k=\left\{m_{k,w}^{(j)}\right\}_{w=1}^{n}$，数据所有者计算 $\mathrm{sy}_k^{(j)}=H_2\left(\mathrm{kk}_j\middle\|\mathrm{ID}_k^{(j)}\middle\|E_k^{(j)}\right)$，验证信息 $\mathrm{vm}=N_{k,w}^{(j)}\|P_{k,w}^{(j)}$，签名 $\sigma_{k,w}^{(j)}=\left(N_{k,w}^{(j)}g_0^{m_{k,w}^{'(j)}}\right)^{\alpha}$，以及为种类 kk_j 计算 $\mathrm{Hdd}_j'=H_2\left(\mathrm{kk}_j\right)$。最后，数据所有者将 $\mathrm{If}=\mathrm{CT}\left\|\left(\mathrm{sy}_k^{(j)}\left\|\left(m_{k,w}^{'(j)},\sigma_{k,w}^{(j)}\right)\right)\right._{l=1}^{k}$ 发送给最适配的雾节点。同时，DO将 $\mathrm{IT}=\mathrm{Hdd}_j\|N_{k,w}^{(j)}\|P_{k,w}^{(j)}$ 发送给TPA。TPA选择随机数 $d\in Z_p$，计算 $W=d\mathrm{UU}$。通过验证等式 $e\left(d\mathrm{Hdd}_j',\mathrm{VV}\right)=e\left(W,\mathrm{Hdd}_j\right)$ 来查询待添加的数据块种类关键字是否在CT中，如果不在则添加数据块信息 $N_{k,w}^{(j)}\|P_{k,w}^{(j)}$。已经通过验证的TPA利用关键字生成属性密钥匹配雾节点中的关键字。如果雾节点中不存在关键字，则添加相应数据块信息 $\left(\mathrm{sy}_k^{(j)}\left\|\left(m_{k,w}^{'(j)},\sigma_{k,w}^{(j)}\right)\right)\right._{l=1}^{k}$。

5 安全性分析

5.1 EHR类别信息的安全性

首先分析方案中的审计阶段是否对DO的EHR信息的隐私进行保护，接着分析类别信息的隐私是否被保护。

在生成标签阶段，EHR的信息主要存在于 $\mathrm{Cf}=\mathrm{CT}\left\|\left(\mathrm{sy}_l^{(j)}\left\|m_{l,w}^{'(j)},\sigma_{l,w}^{(j)}\right.\right)\right._{l=1}^{k}$ 和 $\mathrm{CT}=\mathrm{Hdd}_j\left\|\left(\mathrm{sy}_l^{(j)}\right\|\left(N_{l,w}^{(j)}\left\|P_{l,w}^{(j)}\right.\right)\right._{l=1}^{k}$ 中。其中，$m_{l,w}^{'(j)}$ 是盲化后的数据块，$\sigma_{l,w}^{(j)}$ 是数据块的签名，$(N_{l,w}^{(j)}\|P_{l,w}^{(j)})$ 是验证信息。然而，$\mathrm{sy}_l^{(j)}=H_2\left(\mathrm{kk}_j\left\|\mathrm{ID}_l^{(j)}\right\|E_l^{(j)}\right)$ 是EHR的标记，其中包括类别信息 kk_j、EHR的身份标识 $\mathrm{ID}_l^{(j)}$ 和 $E_l^{(j)}$。重要的类别信息只包含在 Hdd_j 与 $C_1=e(g,g)^{bs}e\left(g,\mathrm{Hdd}_j\right)^s$ 中。由于 H_2 是单向函数，且ABE中的q-BDHE假设是难以解决的。因此，类别信息不会被泄露。由于符合门限数值个雾节点才能恢复盲化因子，因此敌手无法获得盲化因子，即敌手不能得到数据块的明文形式。所以在数据外包阶段EHR的信息也是不会被泄露的。在数据审计阶段，敌手可能通过挑战信息 $\mathrm{chal}:\Theta=\left\{\left\{P_j\right\}_{j=1}^{t},W,\left\{v_w\right\}_{w\in W}\right\}$、属性密钥 $I=g^b g^{ar}\mathrm{Hdd}_j$ 和证明 $\mathrm{proof}_R=\{\sigma,\varphi,\theta,\gamma_1\}$ 中推断出重要的信息。如果恶意的TPA获取了 v_w 并试图通过对EHR发起 n 次审计来获取EHR的信息，它将在 $\left\{\gamma_i=\sum_{l\in P_j(1\leqslant j\leqslant t)}\sum_{w=w_1}^{w_c}v_w m_{l,w}^{(j)}+r_i H_2(\varphi)\right\}_{i=1}^{n}$ 上构建线性方程组。然而，由于在这个线性方程组中有 n 个未知的 $m_{l,w}^{(j)}$ 和 n 个未知的 r_i，因此该方程组是无解的。虽然属性密钥 $I=g^b g^{ar}\mathrm{Hdd}_j$ 包含类别信息 Hdd_j，但由于解决q-BDHE问题是困难的，攻击者无法获得有效的类别信息。因此，在审计阶段也不会泄露有效的EHR信息。

对于类别信息 Hdd_j，本文方案将其隐藏在 $C_1=e(g,g)^{bs}e\left(g,\mathrm{Hdd}_j\right)^s$ 密文陷门中，并设计属性密钥令牌来帮助查找类别。在搜索类别时，TPA只使用属性密钥令牌来匹配 C_1，不需要知道具体的类别信息 Hdd_j。因此，在本文方案中，类别信息 Hdd_j 的隐私性得到了更好的保护。同时，根据ABE访问控制一对多的属性，可以控制恶意TPA无法拥有审计权限。

5.2 TPA的权限安全性

本文方案利用基于属性的加密方法以验证TPA的权限，即恶意雾节点无法以不可忽略的概率打破q-BDHE困难假设来获取类别关键字信息，不合法的TPA也无法以不可忽略的概率打破q-BDHE困难假设来获取审计权限。假设决策q-BDHE问题对群 G_2 和 G_3 有效，则该方案在基于CPA安全的随机预言机模型中是安全的。

假设攻击者 $\mathcal{A}$ 在多项式时间内具有不可忽略的优势，并试图使用CPA安全模型突破该方案。假设挑战者 $\mathcal{B}$ 在解决决策性q-BDHE问题方面具有不可忽视的优势。$\mathcal{A}$ 隐藏ASK信息并试图解密其他普通用户的密文。因此，$\mathcal{A}$ 具有属性密钥ASK，但不包含任何有用的信息。

$\mathcal{B}$ 输入一个基于决策性q-BDHE问题的挑战对 $\left(g,h,y_{g,a,ZZ},Z\right)$，其中 Z 是一个随机元素，有可能等于 $e_1(\mathrm{ZZ},h)$，并且 $y_{g,a,ZZ}=\left(g_1,g_2,\cdots,\mathrm{ZZ},\mathrm{ZZ}_1,\cdots,\mathrm{ZZ}_n\right)\in G_2^{2n-1}$ 属于群 G_3。$\mathcal{B}$ 计算系统公共参数 $P_{\mathrm{pub}}=\left\{p,e_1,G_2,G_3,H_2,h_1,h_2,\cdots,h_U\right\}$，然后选择随机数 $a,b\in Z_p$ 生成公钥 $\mathrm{ZZ}=g^a$ 和私钥 $\mathrm{MSK}=a,g^b$。

$\mathcal{A}$ 将挑战2个相同长度的关键字信息 $m_0=\mathrm{Hdd}_j$ 以及 $m_1=\mathrm{Hdd}_j^*$ 发送给 $\mathcal{B}$，$\mathcal{B}$ 随机选择关键字信息 $m_b\in G_3$，其中 $b\in\{0,1\}$。

查询阶段I。$\mathcal{B}$ 初始化空表格 T_1，以及空属性集 E，$\mathcal{A}$ 可以对属性集重复以下任何查询。

1) 在 $\mathcal{B}$ 收到来自 $\mathcal{A}$ 的属性集 S' 的私钥查询后，说明私钥查询集合是 S。$\mathcal{B}$ 随机选择 $r\in Z_p$，计算 $I=g^b g^{ar}m_b$（其中，Hdd_j 为TPA发送的chal中包含的类别信息），$L=g^r$。对于属性集 $S=\left\{S_1,S_2,\cdots,S_k\right\}$ 中的每个属性 $j\in[1,n]$，$\mathcal{B}$ 都计算 $I_j=t_j^r$。最后，$\mathcal{B}$ 输出属性密钥 $\mathrm{ASK}=\left\{I,L,\left\{I_j\right\}_{j\in[1,N]}\right\}$，表格 T_1 中存放 (j,S,ASK)。

2) $\mathcal{B}$ 检验 (j,S,ASK) 是否存在于表格 T_1 中，如果存在，则返回ASK；否则，返回⊥。$\mathcal{B}$ 基于访问策略 $(\boldsymbol{M},\Re)$ 生成陷门，且 Hdd_j 无法满足该访问结构。$\mathcal{B}$ 随机选择一个向量 $\boldsymbol{c}=\left(s,y_2,\cdots,y_k\right)\in Z_p$ 用来

加密秘密值s。雾节点根据$\boldsymbol{M}$中的每一行$i=1$到$i=n$计算$\zeta_i=\boldsymbol{c}(\boldsymbol{M})_i$。最后输出密文CT

$$\mathrm{CT}=\left\{(\boldsymbol{M},\Re),C_1=e(g,g)^{bs}e(g,m_b)^{s},C_1'=g^{s},C_2'=g^{\alpha-s},\left\{C_i=g^{\alpha q_i}h_{\Re(i)}^{-s}\right\}_{1\leqslant i\leqslant n}\right\} \tag{10}$$

并将其发送给$\mathcal{A}$。

查询阶段II。该阶段与查询阶段I相似，$\mathcal{A}$继续提交一个属性列表给$\mathcal{B}$。

猜测阶段。$\mathcal{A}$输出$b'\in\{0,1\}$作为对关键字的猜想，如果$b'=b$，则$\mathcal{B}$输出0代表猜想$Z=e_1(\mathrm{ZZ},h)$；否则，$\mathcal{B}$输出1代表猜想Z是G_3内的一个随机数。当$Z=e_1(\mathrm{ZZ},h)$时，$\mathcal{B}$可以提供有效的模拟。相应地，$\Pr[\mathcal{B}(g,h,y_{g,a,l},e(g_l,h))=0]=\frac{1}{2}+\mathrm{Adv}_{\mathcal{A}}$。当$Z$是$G_3$内的一个随机数，那么$m_b$对$\mathcal{A}$来说是完全随机的。因此，$\Pr[\mathcal{B}(g,h,y_{g,a,l},Z)=0]=\frac{1}{2}$。那么$\mathcal{B}$的优势不足以打破q-BDHE假设。

5.3　签名的不可伪造性

5.3.1　验证等式正确性分析

在类别信息关键字匹配阶段中，等式的正确性为

$$e(\mathrm{ZZ},g)T=e(g^{a},g)\mathrm{Hdd}_j e(g,g)^{bs}=e(g,g)^{bs}\mathrm{Hdd}_j e(g,g)^{s}e(g^{a-s},g)=C_1 e(C_2',g)$$

在验证proofverify阶段中，等式的正确性为

$$e\left(\sigma\varphi^{H_2(\varphi)},g_1^{\sum_{l\in P_j(1\leqslant j\leqslant t)}\mathrm{sy}_l^{(j)}}\right)=$$

$$e\left(\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(H_1(\mathrm{ID}_{l,w}^{(j)}\|m_{l,w}^{(j)})^{v}g_0^{m_{l,w}^{(j)}}\right)^{\alpha v_w}\mathrm{UU}^{r H_2(\varphi)},g_1^{\sum_{l\in P_j(1\leqslant j\leqslant t)}\mathrm{sy}_l^{(j)}}\right)=$$

$$e\left(\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(H_1(\mathrm{ID}_{l,w}^{(j)}\|m_{l,w}^{(j)})^{v}g_0^{m_{l,w}^{(j)}}\right)^{v_w}g_0^{r H_2(\varphi)},g_1^{\alpha\sum_{l\in P_j(1\leqslant j\leqslant t)}\mathrm{sy}_l^{(j)}}\right)=$$

$$e\left(\prod_{l\in P_j(1\leqslant j\leqslant t)}\prod_{w=w_1}^{w_c}\left(H_1(\mathrm{ID}_{l,w}^{(j)}\|m_{l,w}^{(j)})^{v}\right)^{v_w}g_0^{\gamma_1},\mathrm{VV}^{\sum_{l\in P_j(1\leqslant j\leqslant t)}\mathrm{sy}_l^{(j)}}\right)=e(og_0^{\gamma_1},\theta) \tag{11}$$

5.3.2　签名的不可伪造性分析

本节主要分析恶意雾节点错误行为成功的概率，保证任何对手都不能以不可忽略的概率伪造有效签名。

定理1　在双线性群G_0中，如果解决CDH问题是困难的，则攻击者不能以不可忽略的概率伪造有效签名。

证明　分析设计了如下博弈。假设存在攻击者$\mathcal{A}$，$\mathcal{A}$在本文方案中最多可以进行Q_H次哈希查询和Q_S次签名查询，并以不可忽略的概率伪造有效签名以通过TPA的验证，那么如何构造挑战者$\mathcal{B}$，使其可以解决一个概率为$\varepsilon\varsigma(1-\varsigma)^{Q_S}$的CDH问题实例，这违背困难问题假设。

系统建立。$\mathcal{B}$首先执行系统建立算法和密钥生成算法得到系统公共参数$P_{\mathrm{pub}}=\{p,e,G_0,G_1,G_T,H_1,H_2\}$、$\mathrm{PK}=\{\mathrm{UU},\mathrm{VV}\}$和$\mathrm{MSK}=\alpha$。给定一个在双线性群$G_0$中的CDH实例$(g_0,g_0^{\alpha})$交将其设置为签名验证的公钥，$\mathcal{B}$不知道签名密钥$\mathrm{MSK}=\alpha$。

哈希查询。攻击者$\mathcal{A}$在此阶段最多可以进行Q_H次哈希查询。首先，$\mathcal{B}$创建一个初始值为空的哈希列表H_l来记录查询和回应。

在收到哈希查询$(\mathrm{ID}_l^{(j)},d_{l,w}'^{(j)},m_{l,w}'^{(j)})$后，$\mathcal{B}$随机选择$s_{l,w}\in Z_p^{*}$并投掷一枚均匀的硬币$c_{l,w}\in\{0,1\}$，其中，$\Pr[c_{l,w}=1]=\delta$且$\Pr[c_{l,w}=0]=1-\delta$，接着设置$H_{i,j}^{(j)}$为

$$H_{l,w}^{(j)}=H_1(\mathrm{ID}_l^{(j)}\|d_{l,w}'^{(j)})g_0^{m_{l,w}'^{(j)}}=\begin{cases}g_0^{s_{l,w}}g_0^{c}, & c_{l,w}=1\\ g_0^{s_{l,w}}, & \text{其他}\end{cases} \tag{12}$$

$\mathcal{B}$用$H_{i,j}^{(j)}$元组回应此哈希查询，并将元组$\left(H_{l,w}^{(j)},(\mathrm{ID}_l^{(j)},d_{l,w}'^{(j)},m_{l,w}'^{(j)}),s_{l,w},c_{l,w}\right)$添加到哈希列表$H_l$。如果此查询已经被记录在$H_l$中，$\mathcal{B}$将回应此查询作为哈希列表$H_l$需要记录的内容。

签名查询。$\mathcal{A}$在此阶段最多执行Q_S次签名查询。对于在$(\mathrm{ID}_l^{(j)},d_{l,w}'^{(j)},m_{l,w}'^{(j)})$上的签名查询，$\mathcal{B}$在$H_l$中找到元组$\left(H_{l,w}^{(j)},(\mathrm{ID}_l^{(j)},d_{l,w}'^{(j)},m_{l,w}'^{(j)}),s_{l,w},c_{l,w}\right)$。如果

$c_{l,w}=1$，$\mathcal{B}$丢弃；否则，$\mathcal{B}$计算$\sigma_{l,w}=\left(g_0^{\alpha}\right)^{s_{l,w}}$。根据签名的定义，可得

$$\sigma_{l,w}=(H_1(\mathrm{ID}_l^{(j)}\|d_{l,w}^{'(j)})g_0^{m_{l,w}^{'(j)}})^{\alpha}=\left(H_{l,w}^{(j)}\right)^{\alpha}=\left(g_0^{s_{l,w}}\right)^{\alpha}=\left(g_0^{\alpha}\right)^{s_{l,w}} \tag{13}$$

因此，$\sigma_{l,w}$在$\left(\mathrm{ID}_l^{(j)},d_{l,w}^{'(j)},m_{l,w}^{'(j)}\right)$是一个有效的签名。

伪造性。$\mathcal{A}$返回一个在$\left(\mathrm{ID}_l^{(j)},d_{l,w}^{'(j)},m_{l,w}^{'(j)}\right)$上伪造的签名$\sigma_{l,w}^{*}$且未被查询过。$\mathcal{B}$从哈希列表$H_l$找到元组$\left(H_{l,w}^{(j)*},\left(\mathrm{ID}_l^{(j)*},d_{l,w}^{'(j)*},m_{l,w}^{'(j)*}\right),s_{l,w}^{*},c_{l,w}^{*}\right)$。如果$c_{l,w}=1$，$\mathcal{A}$丢弃；否则，$H_{l,w}^{(j)*}=g_0^{s_{l,w}}g_0^{c}$。根据签名的构造，$\mathcal{A}$伪造的签名$\sigma_{l,w}^{*}$为

$$\sigma_{l,w}^{*}=\left(H_{l,w}^{(j)*}\right)^{\alpha}=\left(g_0^{s_{l,w}^{*}}g_0^{c}\right)^{\alpha}=\left(g_0^{s_{l,w}^{*}+c}\right)^{\alpha}=g_0^{\alpha s_{l,w}^{*}+\alpha c} \tag{14}$$

$\mathcal{B}$通过使用伪造签名计算给定CDH实例的解

$$\frac{\sigma_{l,w}^{*}}{\left(g_0^{\alpha}\right)^{s_{l,w}^{*}}}=\frac{g_0^{\alpha s_{l,w}^{*}+\alpha c}}{\left(g_0^{\alpha}\right)^{s_{l,w}^{*}}}=g_0^{\alpha c} \tag{15}$$

定义R_1表示在签名查询中$\mathcal{B}$不丢弃元组的事件，R_2表示$\mathcal{A}$能够在$\left(\mathrm{ID}_l^{(j)*},d_{l,w}^{'(j)*},m_{l,w}^{'(j)*}\right)$上伪造出有效签名的事件，$R_3$表示事件$R_2$发生的同时$c_{l,w}=0$在元组$\left(\mathrm{ID}_l^{(j)*},d_{l,w}^{'(j)*},m_{l,w}^{'(j)*}\right)$上。因此，$\mathcal{B}$成功解决CDH困难问题的概率为

$$\Pr\left[R_1\wedge R_2\wedge R_3\right]=\Pr[R_1]\Pr[R_2|R_1]\Pr[R_3|R_1R_2] \tag{16}$$

其中，$\Pr\left[R_1\right]=(1-\varsigma)^{Q_S}$，$\Pr\left[R_2|R_1\right]=\varepsilon$，$\Pr\left[R_3|R_1R_2\right]=\varsigma$。因此，$\mathcal{B}$成功破解CDH困难问题的概率为$\Pr\left[R_1\wedge R_2\wedge R_3\right]=\varepsilon\varsigma(1-\varsigma)^{Q_S}$。

6　性能分析

本节将本文方案与经典的公共审计方案CIPPPA[22]、DLIT-LA[25]、DPDP-MHT[18]、DAAS[19]进行性能比较，通过比较各个方案在审计过程中的通信开销、计算开销，来体现本文方案的性能优势。实验在带有2.4 GHz Quad Cores CPU/i5-1135G7U处理器、100 GB硬盘、6 GB内存下的64-bit Ubuntu 20.04.3 LTS平台上进行。实验的库依赖于OpenSSL和PBC。此外，实验使用了f曲线来构造配对操作，其中的开销结果是通过实验100次取平均值获得的。Shamir秘密共享阈值设置最小恢复秘密值的雾节点数目为4，总参与方数目为8。表1为实验参数，表2为不同群中元素的大小。

表1　实验参数

符号	含义
n	被挑战的数据块数目
$E_{0/1/T}$	群$G_{0/1/T}$中的模幂运算
P	配对操作
PM	点乘运算
H	哈希操作
M	乘法操作
Inv	群Z_p中的求逆运算
A	加法操作
$\left\|G_{0/1/T}\right\|$	群$G_{0/1/T}$中元素的大小
$\left\|Z_p\right\|$	群Z_p中的元素的大小
$\|H\|$	字符串$\{0,1\}^{*}$的大小
Cmp	群G_T中的配对比较操作
k	Shamir秘密共享参与方的数量
$\|\mathrm{Index}\|$	被挑战的数据块的索引大小

表2　不同群中元素的大小

元素	大小/B
$\left\|G_0\right\|$	40
$\left\|G_1\right\|$	80
$\left\|G_T\right\|$	240
$\left\|Z_p\right\|$	20
$\|H\|$	32
X	4
ASK	888
Hdd_j	589
ss_i	28

6.1　通信开销

首先在表3中比较了5种方案在整个审计过程中的通信开销，具体包括生成挑战、生成proof这2个阶段的总通信开销。总通信开销最低的是DLIT-LA[25]，其只在生成挑战阶段产生通信开销，通信复杂度为$O(n)$，然而其不支持分布式审计；总通信开销最高的是DPDP-MHT[18]，因为它利用默克尔哈希树作为数据存储结构，数据结构复杂，在TPA与雾节点通信的过程中需要更多的元数据来检测EHR的完整性；DAAS[19]支持分布式审计，通信开销相较于DPDP-MHT[18]略低；本文方案和CIPPPA[22]相差不多，然而CIPPPA[22]也不支持分布

式审计机制。图4展示了在审计过程中，当被挑战的数据块数量变化时，各个方案的通信开销对比。随着挑战数据块数量的增加，本文方案在支持分布式审计的同时，通信开销相较于DPDP-MHT[18]降低了将近93%，相较于DAAS[19]低了约74%，与CIPPPA[22]、DLIT-LA[25]相差不大。

表3　　审计过程中的通信开销对比

方案	通信开销
本文方案	$\|Z_p\| + 2\|G_1\| + \|G_T\| + n\|\text{Index}\|$
CIPPPA	$\|Z_p\| + 3\|G_1\| + \|H\| + n\|\text{Index}\|$
DLIT-LA	$\|Z_p\| + \|G_1\| + n\|\text{Index}\|$
DPDP-MHT	$\|Z_p\| + (n+1)\|G_1\| + n\|H\| + n\|\text{Index}\|$
DAAS	$(n+1)\|Z_p\| + \|G_1\| + n\|\text{Index}\|$

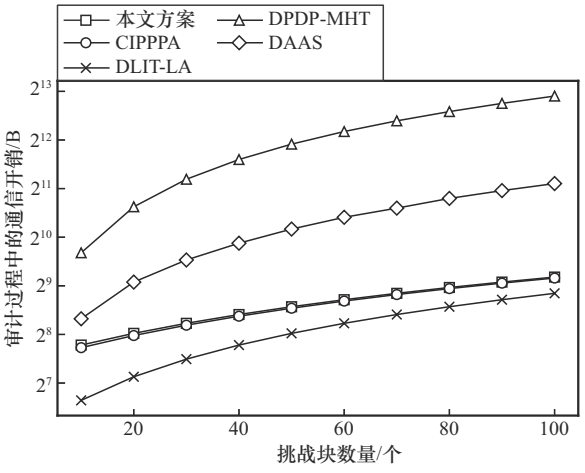

图4　各个方案的通信开销对比

6.2　审计过程的计算开销

本文方案与CIPPPA[22]、DLIT-LA[25]、DPDP-MHT[18]、DAAS[19]方案在生成proof和验证proof阶段的计算开销分别如图5和图6所示，对比如表4所示。本文方案在这2个阶段的总计算开销相较于DPDP-MHT[18]降低了28.5%，相较于CIPPPA[22]降低了49.4%，相较于DLIT-LA[25]降低了95%，相较于DAAS[19]降低了37.1%。在生成proof阶段，本文方案计算开销与EHR成线性关系，而CIPPA、DLIT-LA和DPDP-MHT这3个方案的计算开销与每条EHR中的数据块数量成线性关系。在验证proof阶段，DLIT-LA[25]包含大量双线性对的聚合操作，因此计算开销最高；DAAS[19]由于需要计算多个标签且在验证proof时用到多次双线性配对运算，导致计算开销较高；CIPPPA[22]计算开销高于DPDP-MHT[18]和本文方案，这是因为其TPA需要计算额外的双线性对以及哈希的聚合；DPDP-MHT[18]和本文方案相差不多，由于其TPA还需额外的哈希计算等，因此计算开销稍高。

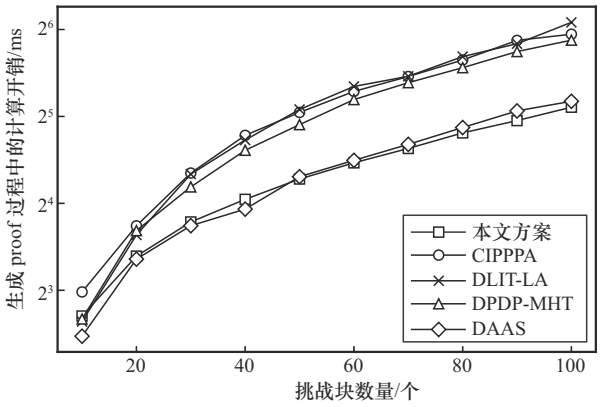

图5　生成proof阶段的计算开销

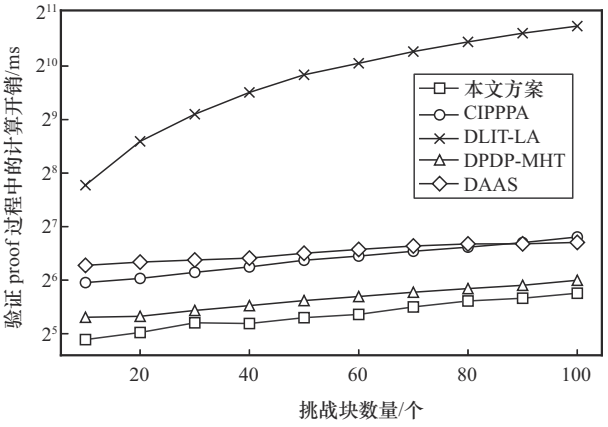

图6　验证proof阶段的计算开销

表4　　生成proof和验证proof阶段计算开销对比

方案	生成proof	验证proof
本文方案	$(n+1)E_1 + (n-1)PM_1 + nM_p + (n+k+1)A_p + (k+1)E_p + H_p$	$2P + \text{Cmp} + H_p + 2E_1 + 2PM_1 + E_2$
CIPPPA	$(n+1)M_1 + H_p + (n+1)M_p + nA_p + (n-1)A_1 + \text{Inv}$	$A_p + H_p + (n+1)M_1 + nH_1 + (n-1)A_1 + 3P + PM_T + \text{Cmp}$
DLIT-LA	$nE_1 + (n-1)PM_1 + nM_p + (n-1)A_p$	$nH_1 + (n+2)P + nPM_T + E_1$
DPDP-MTH	$nM_p + (n-1)A_p + nE_1 + (n-1)PM_1$	$2P + \text{Cmp} + nE_1 + nE_p + nM_1 + (n-1)PM_1$
DAAS	$n\left(M_p + A_p + E_1 + H_p\right)$	$3P + n\left[kM_p + kA_p + M_1\right] + (k+2n)E_1 + \text{Cmp} + PM_T$

本文方案各阶段的通信开销和计算开销如表5所示。通信开销：设置被挑战的数据块数量为10，Shamir秘密共享参与方为8，恢复秘密值的最小阈值为4，并且不考虑不合法的TPA解密属性基加密密文失败的情况。本文方案全部流程通信开销为3969 B。计算开销：准备阶段包括实际方案步骤中的系统建立和准备阶段，总耗时为39.922 ms；标签生成阶段和秘密共享阶段总耗时为24.946 ms；生成挑战阶段耗时9.012 ms；生成proof阶段涵盖属性基可搜索加密的匹配验证的时间，耗时6.518 ms；在验证proof阶段中，雾节点共同恢复秘密值，耗时29.633 ms。本文方案的全部过程计算开销为110.031 ms。

表5　本文方案各阶段的通信开销和计算开销

阶段	通信开销/B	计算开销/ms
准备阶段	1 452	39.922
标签生成	2 029	24.817
秘密共享	224	0.129
生成挑战	80	9.012
生成proof	180	6.518
验证proof	4	29.633
累计	3 969	110.031

6.3　SVD的训练结果

本节选用1.88 MB的公开数据集，其中一共包括100 000条评分记录，训练集占70%，测试集占30%，每条数据包括3个元素（用户、物品、评分）。不同训练次数下的均方根误差（RMSE, root mean squared error）如图7所示，训练一共迭代了30次，随着训练次数的增加，RMSE逐渐降低，即预测准确率逐渐增高。每轮训练平均耗时36.051 1 s，在测试集上计算得到的RMSE为0.916 9。

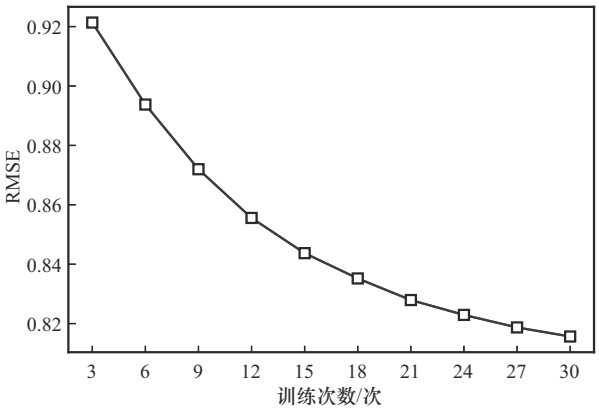

图7　不同训练次数下的RMSE

7　结束语

本文提出了一个基于电子医疗的分布式审计方案，该方案既可以保护EHR类别信息的隐私，又可以保护EHR本身的隐私。通过部署带有哈希链表的计数式布隆过滤器，本文方案实现了数据的动态更新操作，并利用Shamir秘密共享方法实现了多个雾节点协同验证审计机制。正确性及安全性分析表明，本文方案是正确的。基于CDH、q-BDHE等困难问题，本文方案是安全的。通过进一步计算各方案的通信开销、计算开销，表明本文方案在充分保护EHR隐私的同时是高效的。在电子医疗领域中，本文通过设计了一种高效的分类审计EHR的方案使得医护人员可以在分布式存储场景中更加高效且安全地审计EHR。该方案可以有效确保EHR数据的安全存储和传输，保护EHR中的隐私信息及类别信息不被篡改。这对于促进医疗信息的安全性和隐私保护具有重要意义，有助于提高患者和医疗机构对电子健康记录系统的信任度。与此同时，本文方案还可降低医疗机构的成本，提高整个医疗系统的效率和可靠性。

未来着重考虑的是如何在保证审计方案安全性的同时进一步降低计算开销，优化双线性配对带来的复杂计算成本。

参考文献：

[1] HU Y, CHENG L, YAO Q, et al. Exploiting combined locality for wide-stripe erasure coding in distributed storage[C]//19th USENIX Conference on File and Storage Technologies (FAST 21). Berkeley: USENIX Association, 2021: 233-248.

[2] LI J, NELSON J, MICHAEL E, et al. Pegasus: tolerating skewed workloads in distributed storage with in-network coherence directories[C]//14th USENIX Symposium on Operating Systems Design and Implementation (OSDI 20). Berkeley: USENIX Association, 2020: 387-406.

[3] SHICKEL B, TIGHE P J, BIHORAC A, et al. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis[J]. IEEE Journal of Biomedical and Health Informatics, 2017, 22(5): 1589-1604.

[4] DAGLIATI A, MALOVINI A, TIBOLLO V, et al. Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview[J]. Briefings in Bioinformatics, 2021, 22(2): 812-822.

[5] GAO H, HUANG H, XUE L, et al. Blockchain-enabled fine-grained searchable encryption with cloud-edge computing for electronic health records sharing[J]. IEEE Internet of Things Journal, 2023, 10(20): 18414-18425.

[6] XU S, NING J, HUANG X, et al. Untouchable once revoking: a practical and secure dynamic EHR sharing system via cloud[J]. IEEE Transac-

tions on Dependable and Secure Computing, 2021, 19(6): 3759-3773.

[7] KIMOVSKI D, MEHRAN N, KERTH C E, et al. Mobility-aware IoT applications placement in the cloud edge continuum[J]. IEEE Transactions on Services Computing, 2021, 15(6): 3358-3371.

[8] SU Y, LI Y, ZHANG K, et al. A privacy-preserving public integrity check scheme for outsourced EHR[J]. Information Sciences, 2021, 542: 112-130.

[9] 张佳乐, 赵彦超, 陈兵, 等. 边缘计算数据安全与隐私保护研究综述[J]. 通信学报, 2018, 39(3): 1-21.
ZHANG J L, ZHAO Y C, CHEN B, et al. Overview of edge computing Data Security and Privacy Protection [J]. Journal on Communications, 2018, 39(3): 1-21.

[10] BERA B, CHATTARAJ D, DAS A K. Designing secure blockchain-based access control scheme in IoT-enabled Internet of Drones deployment[J]. Computer Communications, 2020, 153: 229-249.

[11] YU S, WANG C, REN K, et al. Achieving secure, scalable, and fine-grained data access control in cloud computing[C]//Proceedings of IEEE INFOCOM. Piscataway: IEEE Press, 2010: 1-9.

[12] HE X, GAO W. Research on blockchain-based data sharing and access control model[C]//Proceedings of 2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA). Piscataway: IEEE Press, 2023: 614-618.

[13] ATENIESE G, BURNS R, CURTMOLA R, et al. Provable data possession at untrusted stores[C]//Proceedings of the 14th ACM Conference on Computer and Communications Security. New York: ACM Press, 2007: 598-609.

[14] JUELS A, KALISKI J. PORs: Proofs of retrievability for large files[C]//Proceedings of the 14th ACM Conference on Computer and Communications Security. New York: ACM Press, 2007: 584-597.

[15] 王子园, 杜瑞忠. 边缘环境下基于无证书公钥密码的数据完整性审计方案[J]. 通信学报, 2022, 43(7): 62-72.
WANG Z Y, DU R Z, A data integrity audit scheme based on certificateless public key cryptography in edge environments [J]. Journal on Communications, 2022, 43(7): 62-72.

[16] 熊金波, 宋良均, 孙罡, 等. 多接入边缘计算网络的资源共享与激励机制[J]. 通信学报, 2023, 44(11): 67-78.
XIONG J B, SONG L J, SUN G, et al. Resource sharing and incentive mechanism of multi access edge computing network [J]. Journal on Communications, 2023, 44(11): 67-78.

[17] 沈剑, 周天祺, 曹珍富. 云数据安全保护方法综述[J]. 计算机研究与发展, 2021, 58(10): 2079-2098.
SHEN J, ZHOU T Q, CAO Z F. Overview of Cloud Data Security Protection Methods [J], Journal of Computer Research and Development, 2021, 58(10): 2079-2098.

[18] WANG Q, WANG C, REN K, et al. Enabling public auditability and data dynamics for storage security in cloud computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 22(5): 847-859.

[19] KE G, ZHANG W B, WANG X Q, et al. Dual attribute-based auditing scheme for fog computing-based data dynamic storage with distributed collaborative verification[J]. IEEE Transactions on Network and Service Management, 2023, 20(4): 4982-4999.

[20] ZHU Y, WANG H, HU Z, et al. Dynamic audit services for integrity verification of outsourced storages in clouds[C]//Proceedings of the 2011 ACM Symposium on Applied Computing. New York: ACM Press, 2011: 1550-1557.

[21] YANG K, JIA X. An efficient and secure dynamic auditing protocol for data storage in cloud computing[J]. IEEE transactions on parallel and distributed systems, 2012, 24(9): 1717-1726.

[22] ZHANG X, ZHAO J, XU C, et al. CIPPPA: conditional identity privacy-preserving public auditing for cloud-based WBANs against malicious auditors[J]. IEEE transactions on cloud Computing, 2019, 9(4): 1362-1375.

[23] WU H, LIU Y, CHENG G, et al. Real-time identification of VPN traffic based on counting Bloom filter and chained hash table from sampled data in high-speed networks[C]//Proceedings of IEEE International Conference on Communications. Piscataway: IEEE Press, 2022: 5070-5075.

[24] XUE K, GAI N, HONG J, et al. Efficient and secure attribute-based access control with identical sub-policies frequently used in cloud storage[J]. IEEE Transactions on Dependable and Secure Computing, 2020, 19(1): 635-646.

[25] SHEN J, SHEN J, CHEN X, et al. An efficient public auditing protocol with novel dynamic structure for cloud data[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(10): 2402-2415.

[作者简介]

张晓旭（1998– ），女，天津人，南开大学博士生，主要研究方向为云数据完整性检测、云数据隐私保护、密码学应用。

陈宇辰（2000– ），男，福建永春人，南开大学硕士生，主要研究方向为密码学应用、加密数据去重。

哈冠雄（1995– ），男，回族，天津人，南开大学博士生，主要研究方向为云数据安全、密码学应用、加密数据去重。

贾春福（1967– ），男，河北文安人，博士，南开大学教授、博士生导师，主要研究方向为网络与信息安全、可信计算、恶意代码分析、密码学及应用等。