• 医学信息技术 •

互联网自助医疗诊断中数据隐私保护方案研究*

冉玖宏[1] 李 冬[2]

([1] 重庆大学医院 重庆 400044 [2] 重庆大学计算机学院 重庆 400044)

〔摘要〕 **目的/意义** 针对互联网自助医疗诊断中易泄漏患者隐私的痛点，设计一套安全的数据隐私保护方案，既保护患者医疗健康数据，又提供准确的诊断和治疗建议。**方法/过程** 医疗机构存储的疾病数据和患者医疗健康数据，经同态加密和隐私保护控制技术处理后上传至云服务器，云服务器计算患者医疗健康数据与医疗机构疾病数据之间的相似度并匹配患者疾病情况，将治疗方法发送给患者。**结果/结论** 基于同态加密和隐私保护访问控制设计的医疗数据隐私保护方案，在互联网自助医疗诊断中能够有效保护患者隐私，同时提供准确的医疗诊断和治疗建议。

〔关键词〕 同态加密；访问控制；医疗诊断；隐私保护

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2024.05.011

Study on Data Privacy-preserving Scheme in Internet Self-service Medical Diagnosis

RAN Jiuhong[1], *LI Dong*[2]

[1]*Hospital of Chongqing University, Chongqing 400044, China*; [2]*School of Computer Science, Chongqing University, Chongqing 400044, China*

〔**Abstract**〕 **Purpose/Significance** Aiming at the pain points that are easy to leak patient privacy in internet self-service medical diagnosis, a set of secure data privacy protection scheme is designed to protect patients' medical and health data and provide accurate diagnosis and treatment suggestions. **Method/Process** The disease data stored by the medical institutions and the patients' medical and health data are processed by homomorphic encryption and privacy protection control technology and uploaded to the cloud server. The cloud server calculates the similarity between the patients' medical and health data and the disease data of the medical institutions and matches the patients' disease condition to send the therapies to the patients. **Result/Conclusion** The medical data privacy protection scheme based on homomorphic encryption and privacy protection access control can effectively protect patients' privacy in internet self-service medical diagnosis, and provide accurate medical diagnosis and treatment recommendations.

〔**Keywords**〕 homomorphic encryption; access control; medical diagnosis; privacy-preserving

1 引言

随着互联网和智能可穿戴设备等技术的逐步成熟，“互联网+医疗”作为一种自助医疗诊断框

〔修回日期〕 2023-12-11

〔作者简介〕 冉玖宏，副主任医师，发表论文 6 篇；通信作者：李冬。

〔基金项目〕 国家自然科学基金联合重点项目（项目编号：U20A20176）。

架[1-3]，可以很好地解决患者“看病难”的问题。自助医疗诊断框架包含 3 个参与方：医疗服务机构、患者和云服务器（第三方服务器），见图 1。其中医疗服务机构根据其存储的疾病数据库，采用机器学习训练医疗诊断模型并上传至云服务器；患者利用智能设备向云服务器发起查询请求；云服务器根据医疗诊断模型判断患者的患病情况，将治疗方法发送给患者。

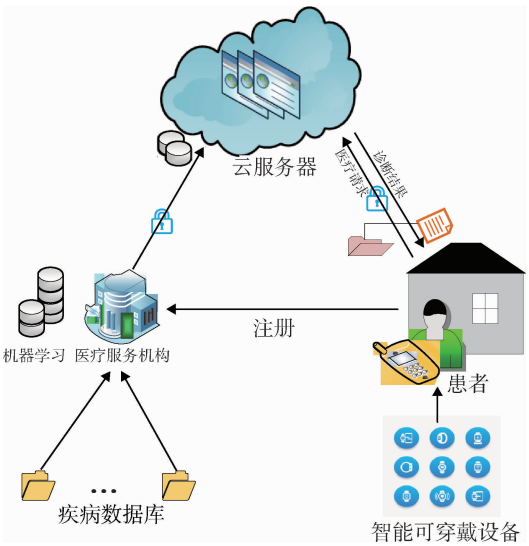

图 1 自助医疗诊断框架

尽管自助医疗诊断给患者带来诸多便利，但依然存在令人担忧的数据外泄问题[4]。为解决数据泄漏问题，许多国家制定相应法律法规[5-6]，设计基于云平台的加密软件[7-8]等。然而，这些方法很难实现理想的安全目标。传统的匿名化技术[9-10]通过脱敏规则屏蔽了敏感信息，但会使数据产生扰动，导致误诊。此外，有学者利用公钥加密[11-15]和差分隐私[16]保护数据，提高医疗诊断的准确性，但诊断效率较低。可搜索加密[17-20]是一种新的加密技术，数据所有者将数据加密后存储在云服务器上，用户向云服务器提出查询请求后，云服务器在加密文本域中匹配相关数据，然后将密文数据发送给用户。用户根据数据所有者提供的密钥对密文解密读取。这已经成为自助医疗诊断隐私保护的新范式。

基于此思想，本文提出一种新的隐私保护自助医疗诊断方案（privacy - preserving self - seviced medical diagnosis scheme，PSMDS）。假设患者可以通过智能医疗可穿戴设备获得医疗诊断所需的医疗数据信息。基于这一假设，将同态加密（homomorphic encryption，HE）与自主访问控制（discretionary access control，DAC）技术相结合，以实现理想的安全目标。此外，PSMDS 用欧氏距离定义的相似度来匹配疾病治疗方法。

2 安全互联网医疗自助诊断方案构建

2.1 方案模型

本文提出的 PSMDS 架构由用户（患者）、医疗机构和云服务器组成，见图 2。其中医疗机构承担两项核心职能：一是将本地存储的疾病数据进行加密处理，上传至云服务器；二是为患者提供注册服务，确保只有授权的患者才能访问。患者加密自己的医疗健康数据后上传至云服务器获取疾病诊治方法。云服务器的职能是对患者身份进行认证和为患者提供精准的疾病诊治方法。

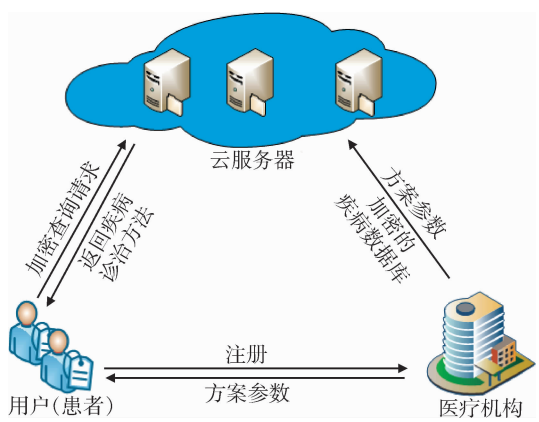

图 2 方案模型

2.2 安全性需求

2.2.1 隐私性 患者的查询数据应得到保护，即使获得了患者的全部加密查询数据，攻击者也无法恢复原始的数据信息。此外，匹配结果应受到保护，不受攻击。

2.2.2 机密性 医疗机构的疾病数据库应得到保护，即使云服务器或攻击者存储了从医疗机构传来的数据，攻击者也无法获得原始的数据信息。

2.2.3 认证 为了确保加密的查询/响应认证，

必须验证其确实来自合法的云服务器/患者，且在传输过程中未被篡改。这样可以有效防止恶意用户伪造查询或响应的行为，确保只有合法的查询和响应才可被接受。

2.3 设计目标

2.3.1 保证安全 如果方案不考虑安全问题，患者医疗健康数据和医疗机构疾病数据信息可能会遭到外泄。因此，应确保数据在其生命周期内的机密性。

2.3.2 保证医疗诊断的高精度 高精度是自助医疗诊断方案中最关键的方面，在保护患者隐私的同时不能降低其精度。因此，应提供高度精确和可靠的医疗诊断服务。

2.3.3 保证低通信开销和低计算复杂度 鉴于自助医疗诊断服务对实时性的要求，应在通信和计算方面保持低开销性能。

3 算法构造

3.1 方案初始化

首先，医疗机构通过执行 $Gen(\lambda)$ 操作获得公共参数 $(G, G_T, q_1, q_2, e, g, h, n = q_1 \cdot q_2)$。然后利用 HE 计算密钥对 (p, d)，其中 p 为公钥，d 为私钥。云服务器选取 $SK_{CS} = s_{CS} \in Z_n^*$ 作为私钥并计算 $PK_{CS} = g^{SK_{CS}}$ 作为公钥。同理，患者选取 $SK_{U_i} = s_i \in Z_n^*$ 作为私钥并计算 $PK_{U_i} = g^{SK_{U_i}}$ 作为公钥。最后，医疗机构选取 $k \in Z_n^*$ 作为疾病数据库的加密密钥，利用 hash 函数为加密后的疾病数据创建 hash 摘要 $< ID_i, H_1(E_k(D_i)) >$ 并发送给患者，其中 i 表示第 i 个疾病，$E_k(D_i)$ 表示第 i 个疾病的加密诊断方法，再向云服务器发送一个成功注册的患者列表。

3.2 患者查询

首先，患者利用便携式设备（程序）进行计算，得到一个查询向量 $\overrightarrow{q} = (q_1, q_2, \cdots, q_n)$，利用 d 对 $\overrightarrow{q}$ 进行加密 $M = Enc(Q) = (Enc(q_1), Enc(q_2), \cdots, Enc(q_n))$。其次，患者利用便携式设备随机选择一个向量 $\overrightarrow{r} = (r_1, r_2, \cdots, r_n)$，计算 $S = Enc(\sum_{j=1}^{n} \alpha_j^2)$，其中 $\{\alpha_j = q_j + r_j\}_{j=1,2\cdots n}$，通过公式（1）计算密文 M_2。此外，根据公式（2）计算签名 Sig_{U_i}，其中，TS 为时间戳。最后，患者利用便携式设备将加密的查询连同签名 $[ID_{U_i} \,||\, M_1 \,||\, M_2 \,||\, TS \,||\, Sig_{U_i}]$ 一起打包发给云服务器。

$$M_2 = S \cdot \prod_{j=1}^{n} (Enc(q_j)^{-2h_j} \cdot Enc(-r_j^2)) = Enc(\sum_{j=1}^{n} q_j^2) \quad (1)$$

$$Sig_{U_i} = (H(ID_{U_i} \,||\, M_1 \,||\, M_2 \,||\, TS))^{SK_{U_i}} \quad (2)$$

3.3 云服务器数据创建

医疗机构以 $(\{DB = (T_1, T_2, \cdots, T_m)\}, D_i)$ 的形式储存疾病数据，其中 DB 表示疾病数据库，$\{\overrightarrow{T_m} = (t_{m1}, t_{m2}, \cdots, t_{mn})\}$ 表示第 m 个疾病特征向量，D_i 表示第 i 个疾病的诊断方法。

对于疾病数据库中第 i 种疾病，医疗机构首先构建疾病特征向量 $\{\overrightarrow{T_i} = (t_{i1}, t_{i2}, \cdots, t_{in})\}$，计算疾病特征向量的密文 $M_3 = (Enc(t_{i1}), Enc(t_{i2}), \cdots Enc(t_{in}))$，然后利用 $\overrightarrow{T_i}$ 计算密文 $M_4 = Enc(\sum_{j=1}^{n} (t_{ij})^2)_{i=1,2,\cdots,m}$。此外，医疗机构利用 k 加密疾病诊治方法 $E_i = E_k(D_i)$。最后，医疗机构将 $[ID_i \,||\, M_3 \,||\, M_4 \,||\, E_i]$ 上传至云服务器。

3.4 结果匹配

云服务器提供具有快速双向认证功能的医疗诊断服务，见图 3。云服务器首先检查 TS 是否在有效期内，如果不在，则拒绝提供服务；然后判断公式（3）是否成立，如果成立，则用户身份验证通过，并进行下一步操作。

$$e(g, Sig_{U_i}) = e(PK_{U_i}, H(ID_{U_i} \,||\, M_1 \,||\, M_2 \,||\, TS)) \quad (3)$$

云服务器利用 M_1、M_3 和公式（4）计算中间量 M_5。

$$M_5 = Enc(\sum_{j=1}^{n} -2q_j t_{ij}) = \prod_{j=1}^{n} Enc(q_j)^{t_{ij}} \quad (4)$$

云服务器根据公式（5）计算欧氏距离 $Enc(D)$，根据公式（6）量化疾病信息。

$$Enc(D) = M_2 \cdot M_4 \cdot M_5 = Enc(\sum_{j=1}^{n} q_j^2 + \sum_{j=1}^{n} (t_{ij})^2 + \sum_{j=1}^{n} (-2q_j t_{ij})) \quad (5)$$

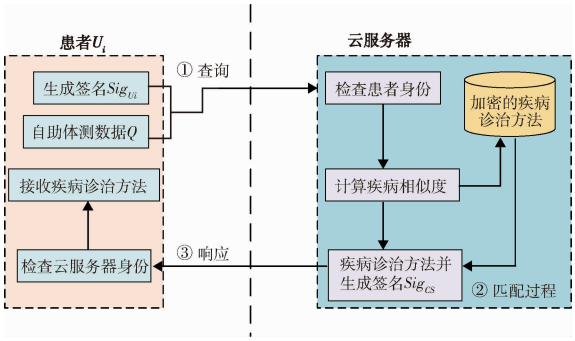

图 3 结果认证匹配

$$Sim(Q,T_i) = \frac{1}{1+Enc(D)}, i = 1,2,\cdots,m \quad (6)$$

云服务器根据 $Sim(Q,T_i)$ 的大小匹配与患者最相近的疾病，$Sim(Q,T_i)$ 越大，则患者患有第 i 种疾病的可能性越高。最后，云服务器根据公式（7）计算签名 Sig_{CS}，利用 $hash$ 函数为匹配到的疾病诊断方法创建 $hash$ 摘要，并将 $(ID_i \| H'_1(E_k(D_i)) \| E_k(D_i) \| ID_{CS} \| TS \| Sig_{CS})$ 传输至患者。

$$Sig_{CS} = H(E_k(D_i) \| ID_{CS} \| TS)^{SK_{CS}} \quad (7)$$

3.5 查询结果读取

患者通过系统且根据公式（8）和公式（9）验证云服务器的身份、检测数据是否被篡改，如果数据未篡改，则患者用 k 对传输过来的 $E_k(D_i)$ 进行解密读取对应的疾病诊治方法。反之，患者拒绝接收云服务器传输过来的数据内容。

$$e(g,Sig_{CS}) = e(PK_{CS},H(E_k(D_i) \| ID_{CS} \| TS)) \quad (8)$$

$$H_1(E_k(D_i)) = H_1'(E_k(D_i)) \quad (9)$$

4 实验分析

4.1 实验结果

提取医疗机构疾病数据库中的 5 种疾病数据[21]（慢性乙型病毒性肝炎、肺炎、心力衰竭、丙型病毒性肝炎以及急性乙型肝炎）匹配患者的医疗健康数据。

为方便实验仿真，采用主成分分析将疾病特征向量维度降至 10，其中元素分别为输血史、低热、中高热、皮肤黄染、皮肤瘙痒、咳白痰、咳嗽、咽喉痛、单侧上肢水肿和下肢水肿等。使用肺炎患者的医疗健康数据进行此次实验，结果见表 1。患者疑似患有该 5 种疾病的相似度分别为 0.434 942、0.818 843、0.569 257、0.383 371 和 0.355 343，可以立即诊断其患有肺炎。此外，通过设置阈值为 0.5，如果 $Sim \geqslant 0.5$，则可以立即诊断该患者也患有另一种疾病。因此，该患者不仅患肺炎，还患有心力衰竭，这与医学结论（肺炎能够导致心力衰竭）相同。因此，PSMDS 的准确性能够得到保证，且设置的阈值也符合标准要求。

表 1 患者医疗健康数据与医院疾病数据库疾病特征向量的相似度

患者查询向量	医疗机构疾病数据库疾病特征向量				
	慢性乙型病毒性肝炎	肺炎	心力衰竭	丙型病毒性肝炎	急性乙型病毒性肝炎
0.142 105	0.242 516	−0.064 207	0.474 277	0.027 243	−0.649 860
−0.125 631	−0.305 166	0.217 344	−0.087 409	−0.001 921	0.274 690
0.154 862	−0.165 048	0.172 303	0.065 973	0.010 267	0.045 240
0.251 785	−0.303 013	0.191 791	−0.009 206	0.117 404	−0.052 266
−0.277 293	−0.039 089	−0.367 887	−0.155 522	0.403 235	0.153 008
−0.183 205	−0.280 374	−0.204 606	0.319 365	0.504 914	−0.151 852
−0.255 395	−0.039 089	−0.367 887	−0.155 522	0.403 235	0.153 008
0.185 736	0.123 286	0.264 834	0.018 575	0.120 467	−0.076 508
−0.192 354	0.636 734	−0.061 524	0.316 654	0.090 605	0.502 150
0.062 572	−0.142 572	0.175 119	0.130 274	0.357 232	−0.071 474
相似度	0.434 942	0.818 843	0.569 257	0.383 371	0.355 343

4.2 计算复杂度对比

PSMDS 与 OTMDS[11]、CEMDS[12]的计算复杂度对比结果，见表 2。假设医疗机构疾病数据库中有 N 种疾病，将其加密后传送至服务器需要 N 次指数操作；患者进行加密查询请求和用于身份验证的签名时，分别需要 2 次乘法操作、2 次指数操作和 1 次 hash 操作；云服务器匹配患者的疾病诊治方法，并计算签名让患者验证其身份，需要 3 N 次乘法操作、1 次指数操作和 1 次 hash 操作。将乘法操作、指数操作、hash 操作分别表示为 C_m、C_e、C_h，则本文提出的 PSMDS 中患者、医疗机构和云服务器的计算复杂度分别为 $2 \times C_m + C_e + C_h$、$N \times C_e$、$3N \times C_m + C_e + C_h$。而 OTMDS 与 CEMDS 均有一个前提假设条件，即信息传输均是经过一次认证的安全通道进行的，除前文提到的基本操作外，这两种方案还包括查询次数的操作 l、矩阵运算操作 C_t 和匹配操作 C_p。因此，这两种方案的计算复杂度略高于 PSMDS。

表 2 计算复杂度比较

方案	患者	医疗机构	云服务器
PSMDS	$2 \times C_m + 2 \times C_e + C_h$	$N \times C_e$	$3N \times C_m + C_e + C_h$
OTMDS[11]	$C_e + 2 \times C_m + 2 \times C_t$	$3N \times C_e + 4N \times C_m + N \times C_t$	无
CEMDS[12]	$(2l + 1) \times C_m + (l + 2) \times C_p$	$2(l + 3) \times C_e + 2^l \times C_p + (l + 2) \times C_m$	$(l + 2) \times (C_p + C_m) + k \times C_h$

4.3 准确性对比

PSMDS 与 OTMDS、CEMDS 的准确性对比结果，见表 3。CEMDS 将患者所患疾病的概率控制在 0 和 1，其中 0 表示患者健康，1 表示患者患有疾病。然而，CEMDS 未能量化患者所患疾病的情况，不具备预测功能。显然 PSMDS 与 OTMDS、CEMDS 相比有较高的准确性。

表 3 准确性比较

预测内容	方案	慢性乙型病毒性肝炎	肺炎	心力衰竭	丙型病毒性肝炎	急性乙型病毒性肝炎
患病风险	PSMDS	0.434 942	0.818 843	0.569 257	0.383 371	0.355 343
	OTMDS	0.228 341	0.038 883	0.132 989	0.282 691	0.318 854
	CEMDS	0	1	0	0	0

4.4 安全性对比

PSMDS 与 OTMDS、CEMDS 的安全性对比结果，见表 4。

表 4 安全性比较

安全威胁	PSMDS	OTMDS	CEMDS
选择明文攻击	√	×	√
合谋攻击	√	√	√
外部窃取攻击	√	√	√
重放攻击	√	×	×
数据篡改	√	×	×

PSMDS 能够完全抵御常见安全威胁。OTMDS 的整个过程均在明文的基础上进行操作，不能抵御选择明文攻击。此外，该方案未能对患者身份进行验证，也不能抵御重放攻击；对于 CEMDS，由于其未能对患者身份和云服务器身份进行验证，不能抵御重放攻击。因此，PSMDS 与 OTMDS、CEMDS 相比有较高的安全性。

5 结语

本文提出的 PSMDS 充分利用 HE、DAC 和 hash 等技术保证了数据的机密性、完整性和诊断模式的

机密性。此外，PSMDS 利用欧氏距离定义的相似度对患者的疾病情况进行量化，利用“互联网 + 智慧医疗”整合所有疾病诊断资源，实现对患者更加精准的诊断。实验分析表明 PSMDS 不仅能抵御各种已知的安全威胁，而且诊断效率也优于与之比较的其他方案。希望 PSMDS 能对互联网自助医疗诊断中的数据发挥安全和隐私保护作用。

利益声明：所有作者均声明不存在利益冲突。

参考文献

1 KAY M, SANTOS J. MHealth: new horizons for health through mobile technologies [J]. World health organization, 2011, 64 (7): 66 - 71.

2 TKACHENKO N, CHOTVIJIT S, GUPTA N, et al. Google trends can improve surveillance of type 2 diabetes [J]. Scientific reports, 2017, 7 (1): 4993.

3 LI D, LIAO X, XIANG T, et al. Privacy - preserving self - serviced medical diagnosis scheme based on secure multi - party computation [EB/OL]. [2023 - 01 - 20]. https: //doi. org/10. 1016/j. cose. 2019. 101701.

4 李冬. 隐私保护数据安全共享交换的应用研究 [D]. 重庆: 西南大学, 2021.

5 NISSENBAUM H F. Privacy in context: technology, policy, and integrity of social life [M]. California: Stanford university press, 2011.

6 SCHWARTZ P, REIDENBERG J R. Data privacy law: a study of United States data protection [M]. New York: LEXIS law, 1996.

7 ROBERT F, SECOMBA G. Boxcryptor [EB/OL]. [2023 - 01 - 20]. https: //www. boxcryptor. com/en.

8 CHARLES B. Spideroak [EB/OL]. [2023 - 01 - 20]. https: //spideroak. com.

9 SWEENEY L. K - anonymity: a model for protecting privacy [J]. International journal of uncertainty, fuzziness and knowledge - based systems, 2002, 10 (5): 557 - 570.

10 MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. L - diversity: privacy beyond k - anonymity [J]. ACM transactions on knowledge discovery from data, 2007, 1 (1): 24 - 27.

11 SUN Y, WEN Q, ZHANG Y, et al. Privacy - preserving self - helped medical diagnosis scheme based on secure two - party computation in wireless sensor networks [EB/OL]. [2023 - 01 - 20]. https: //doi. org/10. 1155/2014/214841.

12 GUO W, SHAO J, LU R, et al. A privacy - preserving online medical prediagnosis scheme for cloud environment [EB/OL]. [2023 - 01 - 20]. https: //ieeexplore. ieee. org/document/8444618.

13 FENG C, YU K, ALOQAILY M, et al. Attribute - based encryption with parallel outsourced decryption for edge intelligent IoV [J]. IEEE transactions on vehicular technology, 2020, 69 (11): 13784 - 13795.

14 RAJAN D P, ALEXIS S J, GUNASEKARAN S. Dynamic multi - keyword based search algorithm using modified based fully homomorphic encryption and Prim's algorithm [J]. Cluster computing, 2019, 22 (5): 11411 - 11424.

15 OGATA W, KUROSAWA K. Oblivious keyword search [J]. Journal complexity, 2004, 20 (2/3): 356 - 371.

16 DWORK C. Differential privacy: a survey of results [C]. Berlin: International Conference on Theory and Applications of Models of Computation, 2008.

17 KARNIN E, GREENE J, HELLMAN M, et al. On secret sharing systems [J]. IEEE transactions on information theory, 1983, 29 (1): 35 - 41.

18 BELLARE M, BOLDYREVA A. Deterministic and efficiently searchable encryption [C]. Santa Barbara: Advances in Cryptology - CRYPTO 2007: 27th Annual International Cryptology Conference, 2007.

19 KO J G, LIM J H, CHEN Y, et al. MEDiSN: medical emergency detection in sensor networks [J]. ACM transactions on embedded computing systems, 2010, 10 (1): 1 - 29.

20 WANG H, GONG J, ZHUANG Y, et al. Healthedge: task scheduling for edge computing with health emergency and human behavior consideration in smart homes [C]. Boston: 2017 IEEE International Conference on Big Data, 2017.

21 刘桃花. 主成分分析在疾病检测中的应用 [J]. 邵阳学院学报: 自然科学版, 2014, 11 (1): 11.