医院信息系统数据中的安全与隐私保护技术研究

陈　权　王永艳
（武警上海总队医院　上海　201103）

摘　要　当前，医院信息系统中对数据的安全与隐私保护并不到位，使医院信息数据极易泄露。因此，文中提出了医院信息系统中的数据安全与隐私保护技术。在提取医院信息系统中的基础数据后，需对其进行同态加密安全存储，再构建基于Spark框架的数据隐私保护模型，对数据进行隐私保护。实验结果显示，随着数据量的增加，与基于安全多方计算的医院信息系统中的数据安全与隐私保护技术和基于门限Paillier密码的医院信息系统中的数据安全与隐私保护技术相比，文中探讨的医院信息系统中的数据安全与隐私保护技术运行时间更短，在提高数据安全与隐私保护效率、降低时间成本方面具有极大的优越性。

关键词：医院信息系统；数据安全；隐私保护
中图分类号　TP393

Research on Security and Privacy Protection Technologies in Hospital Information System Data

CHEN Quan and WANG Yongyan
(Armed Police Shanghai General Hospital,Shanghai 201103,China)

Abstract　Currently, the security and privacy protection of data in hospital information systems are not sufficient, making hospital information data highly susceptible to leakage. Therefore, this paper proposes data security and privacy protection technologies in hospital information systems. After extracting the basic data from the hospital information system, it is necessary to perform homomorphic encryption and secure storage on the extracted data, and then construct a data privacy protection model based on the Spark framework to protect the data privacy. The experimental results show that as the amount of data increases, compared with the data security and privacy protection technologies in hospital information systems based on secure multi-party computing and the data security and privacy protection technologies in hospital information systems based on threshold Paillier passwords, the data security and privacy protection technologies discussed in this paper have a shorter running time, and have great advantages in improving the efficiency of data security and privacy protection and reducing time costs.

Key words　Hospital information system,Data security,Privacy protection

0　引言

随着信息技术的飞速发展，医院信息系统已经成为当代医疗行业中的关键元素。它提升了医疗服务的质量和运行效率，但也催生了一系列难题，尤其是数据安全和患者隐私保护方面的挑战[1-2]。由于医院信息系统包含了大量病患的个人敏感信息，这些信息的保护状况直接影响着病患的权益及医疗机构的信誉。目前，该系统在数据安全和隐私保护方面表现出明显的短板[3]。首要问题在于数据在传输和存储环节中可能遭遇风险，易受到非法侵入或损坏。传统的传输手段往往对黑客攻击和信息截取缺乏足够的防御，从而威胁患者信息的机密性和完整性。其次，医院在管理患者信息时，隐私保护意识不足，部分机构在收集、利用和处理病患数据时并未严格遵守法规，增加了信息滥用或泄露的可能性。再者，面对突发状况时，医院信息系统缺乏有效的应急响应策略，若系统遭受攻击或发生故障，则不能迅速应对，从而引发严重后果，给医院带来重大损失。此外，常见的本地加密存储和数据处理方法效率低，安全防护效能不足，无法有效防止上述问题的发生。鉴于此，医疗机构有必要提升对医疗信息的重视，探索并实施更先进的医院信息系统数据安全与隐私保护技术，以确保海量医疗数据的安全。

针对医院信息系统内的数据安全与隐私挑战，众多研究者提出了一系列保护策略。初期，侧重于运用安全多方计算技术来处理数据隐私。然而，这种方法要求客户端进行多轮安全交互，在群智感知系统中显得不切实际，容易因

作者简介：陈权(1997—)，本科，助理工程师，研究方向为医院信息化建设；王永艳(1977—)，本科，主管技师，从事医院信息化建设工作。

用户移动频繁和终端资源有限而受到影响。另一种方案是借助门限Paillier密码技术，通过将部分原始信息外包给云端以减少直接通信，但这依然面临门限Paillier算法计算和通信成本高的局限性。鉴于此，本文将深入研究医院信息系统的数据安全与隐私保护技术，旨在提升系统的安全性与隐私保护能力，以确保患者的信息权利和医院的信誉不受侵犯。同时，这也将为医疗行业及其他相关领域的理论研究和实践操作提供有价值的参考，促进医疗信息化和现代化的持续发展。

1 提取医院信息系统基础数据

医疗机构的信息技术环境复杂，承载着海量的个体医疗数据，这些数据对病人和医疗机构至关重要。首先，需精准识别出提取的关键信息，即个人隐私数据。在医院信息系统中，无论是门诊咨询还是住院治疗，患者通常会提交详尽的个人信息，如遗传病档案、联系方式（如电话号码）、身份标识（如年龄和性别）以及身份证号等，这是构成个人隐私的重要组成部分。为了有效管理和利用这些信息，本文将依托Hadoop技术架构来提升数据处理效率。Hadoop平台由多个组件构成，其中包括数据采集模块、数据存储模块和数据处理模块。数据采集模块主要依赖第三方工具，负责从各个渠道收集并整合患者提供的个人信息。在数据处理阶段，Hadoop框架发挥核心作用，它支持数据的分布式存储和处理，使大规模数据的管理变得更便捷。此外，系统会对这些敏感信息进行严格地筛选、分类和整理，提炼出所需的基础数据，从而顺利完成医院信息系统基础数据的提取过程。

2 利用同态加密安全存储数据

在提取医院信息系统中的基础数据后，还需对数据进行加密处理，使患者的数据被安全存储[4]。由于移动终端或计算机计算能力较弱，因此本文提出了一种利用同态加密安全存储医院信息系统数据的方法。同态加密指对原始的明文数据进行加密，在这里指加密上文提取的医院信息系统基础数据，并对密文进行各种算术运算，具体如式（1）所示：

$$\begin{aligned} x_1 x_2,\cdots,x_n &\xrightarrow{\text{encrypt}} [x_1],[x_2],\cdots,[x_n] \\ f([x_1],[x_2],\cdots,[x_n]) &\rightarrow f(x_1,x_2,\cdots,x_n) \end{aligned} \tag{1}$$

其中，$x_1,x_2,\cdots,x_n$是上文提取的医院信息系统基础数据，即原始的明文数据；$[x_1],[x_2],\cdots,[x_n]$是对密文进行加密处理后的数据；f表示同态加密的运算函数。根据上述原理，利用同态加密安全存储医院信息系统基础数据的步骤如下。

（1）密钥生成。

1）选取2个不同的素数p和q，两者进行乘积运算，即$n = p \times q$。

2）当p和q为不同的素数时，由欧拉函数可知，存在$\Psi=(n)=(p-1)(q-1)$，记为λ。

3）选择随机整数g，$2 < g < n^2$，并且g和n互质。

4）取$\mu = [L(g^{\lambda} \bmod n^2)]^{-1} \bmod n$，其中$L(x) = (x-1)/n$。

5）令(n,g)为解密的公钥，(λ,μ)为加密的私钥。

（2）m表示待加密的医院信息系统基础数据，r为随机数。利用加密函数Encrypt时，r需满足$0 < r < n$这一条件，且r要和n互质。而paillier加密函数如式（2）所示：

$$c = (g^m x r^n) \bmod n^2 \tag{2}$$

在此切割该数据，再对切割后的数据分别进行同态加密，得到加密后的数据$c_1,\cdots,c_k$。

拥有密钥的医疗机构对提取的医院信息系统基础数据进行加密后，得到切割后的分段医院信息系统密文数据，其是本文需要的安全存储后数据。因此，利用上述步骤可以完成对医院信息系统基础数据的安全存储。

3 构建基于Spark框架的数据隐私保护模型

在利用同态加密安全存储医院信息系统数据后，还需对数据进行隐私保护。Spark框架可在满足海量数据计算效率需求的同时，提供切实可行的隐私保护。因此，需在大数据背景下构建基于Spark框架的数据隐私保护模型，如图1所示。

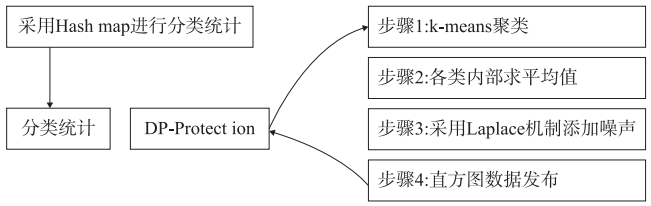

图1 基于Spark的数据隐私保护模型

由图1可知，在导入安全存储的医院信息系统中的数据集后管理数据，即将得到的医院信息系统中的数据从HDFS读取到Spark框架形成RDD医院信息系统基础数据集。再基于Spark框架开始map操作，输出处理医院信息系统的基础数据后的结果，通过分类统计、聚类分组等操作，对医院信息系统中的基础数据进行最优划分，完成分组的目的，再对每个分组中的数据求取平均值。此外，还需在最后生成的Spark隐私保护后数据直方图中加上Laplace噪声。数据直方图即为对医院信息系统数据进行隐私保护后的结果展示。通过上述步骤，可以构建基于Spark框架的数据隐私保护模型，从而对数据进行隐私保护[5]。

4 实验

为验证本文医院信息系统中的数据安全与隐私保护技术的处理效率，以某市中心医院信息管理系统为依托，使用医院信息系统数据库中的医疗数据开展了相关实验。

4.1 实验准备

所有实验都在64位操作系统的电脑上进行，硬件配置

为Intel x86-cpu，cache size（缓存大小）为25 600 KB。实验代码均是在Pycharm软件平台上用python编写的，实验在Microsoft Windows 10环境下完成，所有参数均保持默认值。在上述的软硬件环境下，采用相同的数据集格式，对医院信息系统不同的数据集进行实验。根据医院信息系统数据库中存储的数据，将数据加密存入数据库，观察消耗的时间。数据信息如表2所列。

表2　医院信息系统数据集

数据集	患者人数/人	记录信息
1	1 000	身份证号码，处方药，遗传病史，性别年龄
2	2 000	身份证号码，处方药，遗传病史，性别年龄
3	3 000	身份证号码，处方药，遗传病史，性别年龄
4	4 000	身份证号码，处方药，遗传病史，性别年龄
5	5 000	身份证号码，处方药，遗传病史，性别年龄

由表2可知，实验分为五组数据集。在整体实验中，将本文提出的医院信息系统中的数据安全与隐私保护技术和基于安全多方计算的医院信息系统数据安全与隐私保护技术以及基于门限Paillier密码的医院信息系统中的数据安全与隐私保护技术进行了对比，以运行时间作为实验效果的评估指标。基于上述准备开始相关实验。

4.2　实验结果与分析

为达到良好观感以及能准确进行分析之间的平衡，将本文提出的医院信息系统数据安全与隐私保护技术列为方法一，基于安全多方计算的医院信息系统数据安全与隐私保护技术列为方法二，基于门限Paillier密码的医院信息系统数据安全与隐私保护技术列为方法三，结果如表3所列。

表3　三种方法不同数据集的运行时间

数据集	1	2	3	4	5
方法一运行时间/s	0.97	1.94	2.91	3.88	4.85
方法二运行时间/s	6.60	13.84	21.27	29.37	37.81
方法三运行时间/s	25.93	52.06	77.87	104.02	130.41

从表3可以看出，本文提出的医院信息系统数据安全与隐私保护技术在这5个数据集级别上，运行时间基本维持在个位数，而另外两种方法的运行时间达到了两位数甚至三位数。本文提出的医院信息系统中的数据安全与隐私保护技术的运行时间与基于安全多方计算的医院信息系统中的数据安全与隐私保护技术运行时间最接近的结果是在第一个数据集级别上。本文提出的技术的运行时间是0.97 s，而基于安全多方计算的医院信息系统中的数据安全与隐私保护技术的运行时间是6.60 s，运行时间差是5.63 s，此时基于安全多方计算的医院信息系统中的数据安全与隐私保护技术的运行时间是本文提出的技术的运行时间的6.8倍左右。而基于门限Paillier密码的医院信息系统中的数据安全与隐私保护技术的运行时间更是与前两种技术的运行时间相差巨大。根据实验结果，本文提出的医院信息系统中的数据安全与隐私保护技术与其他两种技术相比，花费了更短的运行时间。由此可以发现，本文提出的医院信息系统中的数据安全与隐私保护技术具备高效性和优越性。

5　结语

近年来，大数据、人工智能等新兴科技迅猛发展，让各行各业的数据量呈现断崖式增长。因此，在数字经济背景下，数据的安全和隐私保护问题受到人们的重视，医院信息系统中的数据安全与隐私保护更是成为重点关注对象。综上所述，本文从医院信息系统的数据安全和隐私保护两个角度出发，对医院信息系统中的数据安全和隐私保护技术进行了研究，以期更好地解决医院信息系统中出现的数据安全和隐私保护问题。但从整体上来看，目前国内外对医院信息系统数据安全性和隐私保护性方面的研究尚处于起步阶段，还需从技术和政策上进行完善，以更好地解决医院信息系统处理数据安全以及隐私保护的性能问题。

参考文献

[1] 袁青霞，赵洪宇.大数据发展背景下网络安全与隐私保护探讨[J].网络安全技术与应用，2022(8)：60-62.

[2] 邓桦，宋甫元，付玲，等.云计算环境下数据安全与隐私保护研究综述[J].湖南大学学报（自然科学版），2022，49(4)：1-10.

[3] 余健，胡孔法，丁有伟.一种面向中医药临床数据的区块链安全与隐私保护方案[J].世界科学技术-中医药现代化，2021，23(10)：3688-3695.

[4] 梁秀波，吴俊涵，尹可挺，等.区块链数据安全管理和隐私保护技术研究综述[J].浙江大学学报(工学版)，2022，56(1)：1-15.

[5] 何赖江，黄伟杰，彭欢.特种设备大数据互联互通与共享平台中数据安全与隐私保护[J].中国特种设备安全，2021，37(11)：50-55.