文章编号:1001-9081(2019)S2-0140-07

基于可搜索加密的区块链数据隐私保护机制

刘格昌,李 强*

(四川大学 计算机学院(软件学院),成都 610065)

(* 通信作者电子邮箱 liq@scu.edu.cn)

摘 要:针对存储在区块链的交易数据都是公开透明的,账户隐私和交易信息得不到保护的问题,提出一种基于可搜索加密的区块链数据隐私保护机制的方法。首先,使用区块链来存储重要数据,存储的是数据的加密形式,可以有效防止明文数据的泄露;其次,为了提供加密数据的搜索功能,利用双线性映射的数学特性来构造区块链交易单,在交易单中增加对关键词的加密信息用于关键词搜索;最后,由用户使用可搜索加密私钥构造陷门,来进行指定关键词搜索。将该机制运用于个人医疗数据区块链系统,解决了个人医疗隐私信息的泄露等问题,搜索时间为2 400个关键词用时为11.23 s。运用该机制不仅增强了区块链数据隐私保护能力,也提供了加密隐私数据的查找和实用的便利。

关键词:区块链数据库;可搜索加密;隐私保护;防篡改;个人医疗数据

中图分类号:TP311.13;TP309.2 文献标志码:A

Blockchain data privacy protection mechanism based on searchable encryption

LIU Gechang, LI Qiang*

(*College of Computer Science (College of Software Engineering), Sichuan University, Chengdu Sichuan* 610065, *China*)

Abstract: Aiming at the problem that the data stored in the blockchain are open and the account privacy cannot be protected, a method based on the searchable encryption of the blockchain data privacy protection mechanism was proposed. Firstly, when using blockchain to store important data, the encrypted form of data were stored, which can effectively prevent the disclosure of plaintext data. Secondly, in order to provide the search function of encrypted data, the mathematical feature of bilinear mapping was used to construct the block chain transaction sheet, and the encrypted information of keywords in the transaction sheet was added for keyword search. Finally, a trap door was constructed by the user using the searchable encrypted private key to perform the specified keyword search. The mechanism was applied to the blockchain system of personal medical data, which solved the leakage of personal medical privacy information and other problems. The search took 11.23 s for 2400 keywords. The use of this mechanism not only enhances the blockchain data privacy protection capabilities, but also provides the search for encrypted privacy data and practical convenience.

Key words: blockchain database; searchable encryption; privacy protection; tamper-proof; personal medical data

0 引言

区块链(blockchain)技术是数字加密货币的底层核心技术,比特币和以太坊等都是使用区块链技术构建区块链系统。随着区块链的研究和应用不断深入,不再局限于比特币(BitCoin)交易应用,各种应用层出不穷,但其数据隐私泄露的问题也越来越突出。为了达到共识节点的一致性,区块链数据库全网的交易记录必须公开给区块链中所有的节点,这必将引发交易记录泄露的风险。在比特币交易中,一些专业分析人员可以通过已产生的交易记录推测出用户的交易规律,进一步可以推测出比特币账户的用户真实身份和位置信息。区块链在能源等敏感数据的应用上也会出现同样的问题,若这些涉及敏感数据的交易被泄露,将给个人、机构和国家造成不可挽回的后果。所以,对区块链数据库的数据隐私安全问题的研究同样重要。

然而,区块链数据库[1]由于其与中心化管理明显不同,研究的方法也不一样。传统中心化关系数据库管理系统、NoSQL数据库管理系统的安全问题都是由单一机构进行管理和维护,拥有绝对的数据控制权,因此可以提高中心节点的防御能力来增强数据库的安全。区块链技术中,数据分散存储在各个共识节点上,没有中心的管理者,各个节点的性能和防御能力参差不齐,很容易对一些设施弱的节点进行攻击,所以区块链达到交易防篡改的同时,也要加强交易数据的隐私保护。

目前针对恶意节点及数据防泄露问题,有以下几种解决办法。针对区块链节点的加入设置规则,没有得到允许的节点不能加入区块链中,超级账本(hyperledger)[2]中,采用了节点认证机制。对于已经存在的恶意节点,Huang等[3]提出了一种基于恶意节点行为分析的检测方法,可以识别定位恶意节点,消除安全隐患。在数据货币中,应用广泛的是"混币(Coin Shuffle)"机制[4],该方案的主要措施是对交易单中的内容进行混淆,从而增加攻击者的分析难度。Zcash的加密方

收稿日期:2019-01-17;修回日期:2019-02-25。 基金项目:四川省科技厅重点项目(2018GZ0105,2018GZ0104)。

作者简介:刘格昌(1994—),男,河南新乡人,硕士研究生,CCF会员,主要研究方向:区块链; 李强(1963—),男,四川成都人,副教授,博士,主要研究方向:移动云计算、大数据、区块链。

案是一种利用零知识证明技术(Zero-Knowledge Proof) 来达到不需透露交易的相关信息,但该算法证明过程非常缓慢,在效率上有瓶颈[5-6]。文献[7]采用区块链技术来解决物联网的隐私数据保护问题,其主要改进是对物联网中产生的数据进行区块化,来实现一种去中心化的存储方式,每个区块只存储数据的一小部分,然后对其进行加密处理从而达到隐私保护的目的。这种方式把问题复杂化,计算资源耗费较大。文献[8]同样采用区块链数据库来存放个人隐私数据的加密形式,但是在构造区块链数据库时,还是保留了一些明文来达到对数据的标识,无法完全实现对数据的保密。文献[9]中,同样是对医疗数据进行加密处理,只有病人才能对加密数据进行解密,但是由于该区块链系统同样暴露了一些明文信息,会导致恶意用户对数据的搜索和破坏,对隐私保护的能力不够彻底。

本文针对隐私保护问题,提出了基于可搜索加密的区块链数据隐私保护机制。该机制利用可搜索加密技术实现对存储在区块链数据库中的关键数据的加密及密文搜索,提高了数据隐私保护能力,同时也提供了加密隐私数据使用的便利,节省了区块链数据库的网络和计算开销。

1　区块链技术及可搜索加密技术

1.1　区块链技术

区块链在架构上一般公认是由数据层、网络层、共识层、合约层和应用层组成[10],如图 1 所示。数据层的用途是交易数据,每个交易最后被封装成区块上传到区块链数据库中,其中每个交易单包含时间戳(timestamp) 、哈希值(hash values) 、数据文件等等。网络层最明显的特征是去中心化的分布式组网机制。共识层的作用主要是区块链节点参与共识的过程,包括一些共识算法。合约层是一些脚本、合约机制,自动执行的代码,智能合约(Smart contract) 是区块链的一个发展方向。应用层是针对具体的使用场景,设计不同的区块链应用软件。

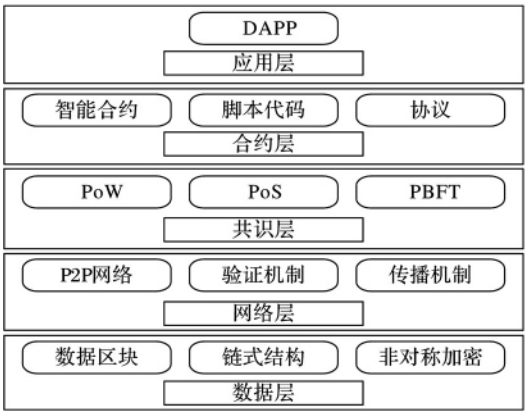

图 1　区块链通用架构模型

1.1.1　基于时间戳的链式数据库结构

区块一般由区块头(Header) 和区块体(Body) 构成,而区块头一般是由前一区块地址(Prev-block) 、哈希值、Merkle 根(Merkle-root) 以及时间戳组成,如图 2 所示。在区块头里,通过时间戳来记录区块的生成时间,是区块链中不可缺少的条件,可以作为交易单存在的依据。时间戳为区块链提供了时间维度,使得区块链可以通过时间戳来重现区块生成过程。哈希值通过哈希函数计算得出,它具有不可逆性、定时性、定长性和抗碰撞的特点,是用来识别文件是否发生篡改的有利工具。区块链中在每个交易单中和区块头中使用哈希值来对交易单和区块进行标识。SHA256 哈希运算具有 256 位比特,其散列空间可以达到 2^{256} 个,可以满足平常的应用需求。区块头部还有一个重要的数据结构,即 Merkle 树,是由交易单进行分组哈希,并将新生成的哈希值再进行分组哈希,最后得出一个哈希值记录在区块头内。

数据层的特点是链式结构,是由各个区块链接形成一条由创始区块到最新区块的数据存储结构。每笔交易单都必须对其进行数字签名,用来证明该交易单的有效性。当一个时间段内的交易单验证通过后,组成区块并加盖时间戳链接到区块链数据库中,随着区块的不断增多,这样就形成一条以时间戳为时间轴的数据库,从而可以按时间对数据进行追溯。这是区块链第一道数据防篡改机制。

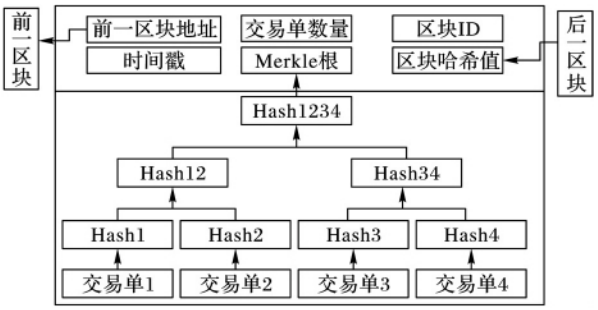

图 2　区块

1.1.2　基于点对点的数据传播机制

数据层中对区块使用非对称加密来达到安全性需求和区块的所有权验证。在区块的传播过程中,通过非对称加密对区块里的交易单进行验证。非对称加密一般由两个不同的密钥组成,一个称为私钥,另一个称为公钥。简单地理解非对称加密即是由 x 计算 $y = f(x)$ 是简单的,而反过来由 y 计算出 x 是困难的。私有密钥由用户保存,不能泄露; 公开密钥可以公开给区块链中的节点。常见的非对称加密算法有 RSA (Rivest-Shamir-Adleman) 加密算法、Elgamal 加密算法、椭圆加密算法(Elliptic Curve Cryptography, ECC) 等。其特点是信息由其中一个密钥加密,只能由另一个密钥解密。

点对点传输机制是一种在节点之间均衡负载的数据传输机制,是一种分布式网络。相比于中心化网络,点对点网络中的各个节点具有相同的地位。负载均衡是指每个节点都会接收到区块并传播给其他建模与仿真节点。当节点接收到待验证的区块时,从区块里的交易单中提取出非对称加密的公钥,对里边的数字签名进行解密并验证[11]。这是区块链第二道数据安全验证机制。

1.1.3　全网共识节点的共识机制

共识机制主要是用来对区块链数据库中数据的真实性和一致性作出验证。共识机制也就是运行在节点上约定好的对区块验证规则。由于全网的节点会出现恶意节点的情况,即拜占庭将军问题。目前共识机制分为两类: 强一致性共识算法有 Raft、拜占庭容错(Practical Byzantine Fault Tolerance, PBFT) ; 最终一致性共识算法有工作量证明(Prof of Work, PoW) 、股权证明 (Proof of Stake, PoS) 。各类共识算法各有优缺点,比如强一致共识算法安全性比最终一致性共识算法高,但算法复杂度高、效率低。

本文采用PBFT算法，该算法在1999年被提出，用来解决拜占庭问题。共识过程共包含三种角色：客户端、主节点和从节点。该共识算法可以容忍的恶意节点为f个，则从节点的个数为$n = 3f + 1$。

PBFT算法的过程[12]如下：

1）客户端负责微主节点传送交易，获取数据库服务；

2）在一个时间段内，主节点对全网交易单进行打包，并广播给从节点；

3）所有从节点执行验证操作，并将结果发回客户端；

4）客户端接收从节点返回来的执行结果，若验证正确结果数大于$f + 1$，则表明数据共识成功，存入区块链数据库中。

对应的共识过程有四步：

1）预准备阶段：主节点接收客户端的请求后，生成消息，格式为$\langle\langle PRE\text{-}PREPARE, v, n, d\rangle, m\rangle$。其中：$v$为该时间段的视图编号，$n$为该消息编号，$d$为该消息$m$的主节点生成的摘要。随后主节点广播该消息给从节点。

2）准备阶段：从节点从主节点接收消息后，对消息里的视图编号进行验证，同时广播该消息给其他建模与仿真从节点，格式为$\langle PREPARE, v, n, d, i\rangle$。

3）确认阶段：当从节点接收到消息后，验证消息，并向从节点广播确认消息，格式为$\langle COMMIT, v, n, D(m), i\rangle$。

4）返回结果阶段：从节点返回验证结果给客户端，若正确结果大于$f + 1$，则表示上链存储成功。

PBFT共识算法与工作量证明不同的是，它不需要依靠大量的算力来争夺出块的权力，在保证了资源的前提下，有效地解决了共识存在的问题。PBFT也和权益证明不同，不依靠哪个节点权益份量大而获得出块的权力。

这是区块链数据安全的第三道数据共识机制。

1.1.4 区块链隐私问题的定义

数据的隐私问题通常是指不愿意被泄露给外人的数据，只能被数据拥有者拥有，这些隐私数据若被泄露，将对用户造成无法弥补的损失。

区块链是一种去中心化的、分布式的数据存储技术，其存储信息一般对区块链节点是公开的，在区块链数字货币的应用中，交易单记录的数据通常是公开的，没有额外的数据保护方法。交易记录会显示出一些涉及用户的敏感信息，例如比特币的交易单会显示交易双方及金额。为了保护用户的隐私数据，需要对隐私数据进行加密处理，以减少隐私泄露的风险。加密过的隐私数据若被泄露，攻击者也无法从中得出明文数据，这是区块链技术的又一道隐私保护措施。

1.2 可搜索加密技术

在分布式存储或者云存储环境下，数据的存储通常不是明文存储，而是由用户加密后上传到分布式数据库中。那么就会产生一个问题，即服务提供者如何对加密的文件进行管理，比如想通过查询某个关键词而获得加密文件，而可搜索加密的实现解决了这个问题。

可搜索加密有两种，其中加解密密钥相同的为对称可搜索加密，加解密密钥不同的为非对称可搜索加密。它们分别适用于不同的场景中：对称可搜索加密为只有拥有私钥的用户才可以对关键词进行加密和对关键词进行搜索；非对称可搜索加密为拥有公钥的用户可以对关键词进行加密，只有拥有私钥的用户才能对关键词进行搜索。为了实现区块链数据库的数据共享，本文采用非对称可搜索加密[13]。

1.2.1 可搜索加密介绍

2004年，Boneh等[14]通过利用椭圆曲线非对称加密，设计出公钥可搜索加密（Public-key Encryption with Keyword Search，PEKS）。该技术使用素数阶为p的群和一个双线性映射e：$G_1 \times G_1 \rightarrow G_2$。双线性映射是把两个向量空间上的元素，映射到另一个向量空间上的一个元素，这个映射有以下特性[14]：

1）可计算性：存在多项式时间算法对任意的g，$h \in G_1$，计算$e(g, h) \in G_2$的值。

2）双线性性：对于任意x，$y \in [1, p]$，有$e(g^x, g^y) = e(g, g)^{xy}$。

3）非退化性：若g是一个群G_1的一个生成元，那么$e(g, g)$映射得出的是群G_2的一个生成元。

通过利用非对称密码，公钥可搜索加密算法可通过如下过程实现：

1）$(pk\_peks, sk\_peks) = \mathrm{KeyGen}(\lambda)$：输入参数$\lambda$，输出可搜索加密公钥$pk_peks$和可搜索加密私钥$sk_peks$。

2）$C_w = \mathrm{Encrypt}(pk\_peks, W)$：输入公钥$pk_peks$和关键词$W$，输出关键词密文$C_w$。

3）$T_w = \mathrm{Trapdoor}(sk\_peks, W)$：输入私钥$sk_peks$和关键词$W$，输出陷门$T_w$。

4）$b = \mathrm{Test}(pk\_peks, C_w, T_w)$：输入可搜索加密公钥$pk_peks$、关键词陷门$T_w$和关键词密文$C_w$，根据算法算出匹配结果，然后通过判定值$b = \{0,1\}$判断是否完全匹配。

1.2.2 可搜索加密步骤

可搜索加密技术与基于明文搜索不同，它是基于加密文件进行搜索，所以需要提供类似于明文搜索的“标签”，“标签”是对关键字用可搜索加密公钥进行加密制作而成。对关键字进行加密，是不可能通过加密后的关键字获得任何关于明文的信息，所以相对于明文标签，用可搜索加密技术制作的“标签”是安全的。

可搜索加密的步骤如图3，具体执行如下：

Step1：用户使用密钥对明文文件进行加密，同时使用可搜索加密密钥对关键词进行加密，并上传至数据库。

Step2：查询过程，用户使用可搜索加密密钥对待查关键词进行加密，生成陷门，同时陷门不会透露出任何关于关键词的信息，发送给数据库。

Step3：数据库以陷门作为输入，并执行匹配算法，返回所有和陷门匹配成功的密文文件。

Step4：用户接收密文文件，并用密钥解密。

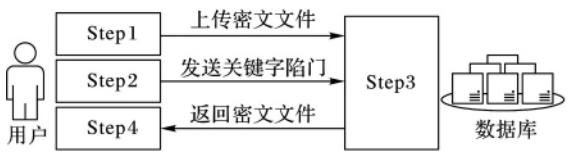

图3 可搜索加密步骤

可搜索加密机制在保证密文的安全性的同时，对使用者提供了搜索的能力，能完成传统加密方式所达不到的功能，在电子邮件系统、大量加密文件搜索等方面具有广泛的应用，目前已广泛应用于云计算领域中。本文对云计算领域的可搜索

加密运用于区块链数据库进行研究,以适用于区块链数据库[15]。

1.2.3　可搜索加密的时间效率

可搜索加密的时间效率主要由 4 个方面决定: 表达能力、通信效率、计算效率、属性特性,在实际应用中要对这 4 个方面进行优化和折中。表达能力直接影响数据搜索的精细程度,表达能力越强,则数据搜索返回的范围越小,搜索效果越好,但也会造成计算资源的耗费。通信效率则是与可搜索加密后的密文长度和网络的速率有关,密文长度越短,网速越快,数据传输所费的时间就越短。计算效率则是与可搜索加密采用的加解密算法有关,比如本文采用的基于椭圆曲线的双线性映射,也是大多数可搜索加密所采用的,其实现效率方面没有传统的指数运算高。在传统构的搜索方案中,密文中涉及很多属性时,也会造成时间上的开销。但是,可搜索加密的未来发展方向是朝着大属性集合,因此,构造高效的算法将会对可搜索加密提供有效的解决方案。

1.2.4　PEKS 的多对一模式

可搜索加密的多对一模式是从公钥密码体制发展而来,与公钥密码体制相同,都是允许多个数据拥有者将其加密文件上传到数据库中,并且通过可搜索加密对其建立安全索引,而仅单个用户可以进行基于关键字查询。多对一模式用途比较广泛,可以适用于较多的使用场景。多对一模式要求数据接收者拥有具备陷门制作与指定关键字查询的能力,同时执行查询请求的区块链数据库服务器无法从密文中获取明文信息。

2　基于可搜索加密的区块链隐私保护机制

基于可搜索加密的区块链隐私保护机制是利用可搜索加密技术对区块链数据进行加密保护并提供有效密文搜索的新的区块链隐私保护机制。该机制不仅增加了数据隐私保护能力,而且提供了加密数据的查询及使用的便利,大大节省了网络资源及计算开销。

2.1　新机制下的区块链架构

通过结合可搜索加密技术,区块链五层架构模型中数据层就增加了一个模块,为可搜索加密技术。区块链应用层则提供密文搜索功能。全新的区块链五层模型在提高隐私数据保护能力的同时,也提供了加密隐私数据搜索的解决办法。新的五层模型如图 4。

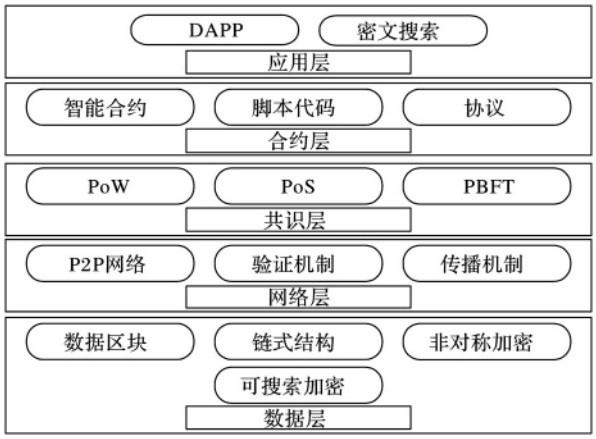

图 4　新的五层模型

2.2　区块链交易单的构造

在不改变区块链数据库的原有体系下,对区块链交易单的改进为增加一项用于密文搜索的 PEKS_关键词,是由用户使用可搜索加密公钥对关键词进行加密而成,作用是在区块链数据库中进行加密文件搜索。同时交易单对应的加密医疗数据,分为两种情况: 1) 若该文件仅由用户 A 拥有,且文件的使用者也是本人,则用户 A 使用对称加密密钥对文件进行加密,计算出哈希值后写入交易单中; 2) 若该文件是由用户 A 发送给用户 B,则用户 A 使用 B 的非对称公钥对文件加密,同时用户 A 使用 B 的可搜索加密公钥对关键词进行加密后上传到区块链数据库,用户 B 就可以进行关键词搜索,获取该加密数据。交易单中加密数据这一项是用来存储用户上传的加密文件,可以存储加密文本、图片、视频等,若上传数据较大,可以先进行压缩后再加密,形成交易单后上传。交易单构造如图 5。

图 5　交易单

针对区块链数据上链存储时数据量庞大的情况,同时为了提高区块链的使用效率,通过研究发现,用户可以分为两类: 一类为轻量级用户,他们的基础设施不完善,使用频率小,没有很大的计算能力; 一类为完备级用户,他们的设备完善,存储空间大,使用频率大,拥有很强的数据处理能力。其中轻量级用户不需要保存记录,当需要使用区块链数据库的数据时,向完备级用户提出申请; 而完备级用户存储全网所有的交易记录,参与共识,是分布式区块链数据库的组成部分。

2.3　加密数据上链

加密数据上链所需要的步骤如下:

1) 对于要提交给区块链数据库存储的文件,根据文件的用途可以采用两种加密方式: 若文件仅是由自己拥有,则可以用对称密钥加密文件; 若文件是发送给其他建模与仿真用户,可以使用非对称加密技术对文件进行加密。同时用户通过算法 KeyGen (λ) 得到可搜索加密公钥和私钥,其中 λ 是参数。接着对数据文件提取出关键词,使用非对称可搜索加密算法 Encrypt (pk_peks, W) 用可搜索加密公钥对关键词进行加密,其中 pk_peks 是可搜索加密公钥,W 是关键词。接着写下时间戳和交易 ID,生成交易单后向主节点提交验证请求。

2) 区块链数据库全网节点执行共识算法,由于本文采用 PBFT 算法,算法设置主节点对一个时间段接收到的交易单打包后,发送给从节点用来验证。

3) 从节点接收区块并对区块里的交易单进行验证,提取出交易单中的非对称加密公钥,对交易单中的数字签名进行解密后得到解密后的哈希值,与其交易单中存储的哈希值进行对比,若完全匹配则验证通过,说明数据没有问题; 若出现差错,则说明数据已被篡改,退回该交易单并不记入区块,同时发送交易单的验证结果给用户。

4) 用户接收验证节点的验证到结果: 若验证结果正确数大于 $f+1$，则表明医疗区块链数据库接收该区块并存入区块链数据库中，也即主节点、从节点都保存该区块; 若验证结果正确数小于 $f+1$，则说明上传的交易单出现问题，需要再次上传以待验证。

2.4 加密数据查找

在区块链数据库中查询指定关键词的步骤如下:

1) 用户想查找关键词为 W 的加密文件，使用可搜索加密算法生成陷门 T_w = Trapdoor(sk_peks, W) ，sk_peks 为私钥，W 为关键词，T_w 为陷门。发送查找请求给区块链数据库上的共识节点。

2) 区块链数据库上的共识节点收到查找请求后，从查找请求中提取出陷门，接着执行可搜索加密的 b = Test(pk_peks, C_w, T_w) 匹配出结果，若 $b=1$，则表示查询成功，$b=0$ 表示查询失败。其中 pk_peks 为公钥，C_w 为密文。

3) 用户接收区块链数据库返回来的交易单，从返回的交易单中得到包含关键词为 W 的加密文件，然后用密钥进行解密后得到明文数据文件。用户若想验证存储的医疗数据文件是否被篡改，则可以对加密文件进行哈希值计算，若得出的哈希值与交易单中记录的哈希值一样，则说明文件正确无误。

3 隐私保护机制在个人医疗数据中的应用

电子病历是一名病人最为重要的隐私信息，现在的医疗信息都是通过电子数据形式展现给病人，电子数据的安全性成为人们最为关注和担心的问题。医疗数据的安全问题可以通过法律和技术两种方式进行保障。现在的区块链技术是最适合用来解决隐私保护防篡改问题的有效解决方法。还可以通过可搜索加密，实现在区块链数据库中搜索加密文件的功能，使医疗数据的防泄露得到进一步得到保障。区块链技术由于其去中心化、永久记录和便于审计的特点，可以满足人们对隐私数据的完整性、可限制性以及防泄露的安全需求; 同时，为了节省网络数据传输和计算开销，充分利用计算能力强的节点来提供在密文上进行关键词查找。

3.1 交易单存储电子病历

交易单中存储的信息，通过分析可以归为以下几个方面。首先是用户信息方面，对于患者来说，包括患者的姓名、性别、年龄、身份证号等个人信息; 对于医护人员来说，增加了其所在的医院、科室、级别等。这些信息是涉及用户的隐私，需要对其进行加密处理后上传到区块链中，以保证其不会被明文泄露给非法的数据侵入者。在用到这些涉及隐私的数据时，须经过实体拥有者同意后，拥有者对加密信息进行解密后方可获取这部分信息。其次是医疗信息方面，主要是记录了患者在就诊时所产生的医疗信息。比如某一患者在某天去某个医院进行就诊，就生成一条就诊记录，包括就诊时间、就诊医师、时间、费用等，若就诊时还有拍片等图片、视频数据产生，对其进行加密处理，并存入区块链交易单中，这样，若图片或视频被不法分子获取、篡改、删除时，就可以通过区块链去中心化存储来防止数据被泄露和篡改。

3.2 加密医疗数据上链

交易单中存储的是医疗数据的加密信息，是涉及患者最为重要的个人就诊信息，若是泄露给不法分子，将会对患者、医院造成巨大损失。以往的方法是对加密的文件留出明文标签，对明文标签进行查找，但是这种做法对数据隐私保护不彻底，还是会给不法分子进行数据分析造成泄露的后果。比如留出的明文标签涉及一些用户的 ID 或就诊的一些信息，这些明文信息同样具有隐私性，是不能暴露在数据库中供搜索识别的。为了对数据隐私保护进行得更为彻底，提出不留明文标签，而是采用可搜索加密技术对明文标签进行加密处理，同时还可以对密文实现搜索功能，实现明文搜索同样的功能。

上传医疗数据到区块链数据库所需要的步骤如下:

1) 对要提交给医疗区块链数据库存储的医疗文件进行加密。同时用户通过算法 KeyGen (λ) 得到可搜索加密公钥和私钥，对医疗数据文件提取出关键词，使用非对称可搜索加密算法 Encrypt (pk_peks, W) 用可搜索加密公钥对关键词进行加密。接着写下时间戳和交易 ID，若有图片或视频等大文件，可以对其进行压缩加密后上链。生成交易单后向主节点提交验证请求。

2) 医疗区块链数据库全网节点执行共识算法，由于本文采用 PBFT 算法，算法设置主节点将一个时间段接收到的交易单打包后，发送给从节点用来验证。

3) 从节点接收区块并对区块里的交易单进行验证，提取出交易单中的非对称加密公钥，对交易单中的数字签名进行解密后得到解密后的哈希值，与其交易单中存储的哈希值进行对比: 若完全匹配则验证通过，说明数据没有问题; 若出现差错，则说明数据已被篡改，退回该交易单并不记入区块，同时发送交易单的验证结果给用户。

4) 用户接收验证节点的验证结果: 若验证结果正确数大于 $f+1$，则表明医疗区块链数据库接收该区块并存入区块链数据库中，也即主节点、从节点都保存该区块; 若验证结果正确数小于 $f+1$，则说明上传的交易单出现问题，需要再次上传以待验证。

上传医疗数据到区块链数据库加密数据上传函数 uploadDatatoDatabase 关键伪代码如下:

```
boolean uploadDatatoDatabase ( W, data, pk_peks) {
/* 输入: W 为关键词,data 为上传明文数据,pk_peks 为用户可搜索加密公钥
输出: true 为上传成功,false 为上传失败* /
/* 使用 pk_peks 私钥对关键词 W 进行加密,得到关键词密文 C_w,是交易单中用于搜索* /
element_t H1_W2;                                    //element_t 类型
sha512 ( W, lenW, hashedW2) ;                       //对关键词进行转换
element_from_hash ( H1_W2, hashedW2, strlen( hashedW2) ) ;
//对转换后的关键词转换成 element_t 类型
element_t r, hR, t;
element_random ( r) ;
element_pow_zn ( hR, pk_peks, r) ;
element_apply ( t,H1_W2, hR) ;                      // 初始化 r, hR, t 参数
peks peks;
/* peks 为最终存储可搜索加密关键词信息的加密密文* /
element_pow_zn ( peks, pk_peks, r) ;
//对 peks 运算
while ( bitswanted) {                               //对每一个比特进行运算
  for ( i = 7; i > = 0 && bitswanted > 0; i --, bitswanted --)
    * ptr ++ = '0' + ( ( byte >> i ) & 0x01) ;      //按比特运算
    if( bitswanted = = 0)                           //如果位数得到要求则退出运算
```

```
        break;
    }
    encData ( data)
    /* 对上链的明文数据进行加密,可以选择对称加密,也可以选择非对称加密,具体加密方法按用途定* /
    transaction = dataTransaction( data, C_w)
    /* 关键词密文 C_w,加密文件 data 添加到交易单中,同时写入时间戳、ID 和哈希值等生成完整交易单* /
    if( uploadtoNode ( transaction) > ( f + 1) ) { return true}
    else{ return false}
    /* 完整的交易单上传至主节点以待验证,同时返回上传结果,若收到验证成功的个数满足 PBFT 共识算法的条件,返回 true,否则返回 false* /
    }
```

3.3 加密医疗数据查找

本文提出的隐私保护方案是对医疗数据进行加密存储,所以随之而来的是要解决加密文件的搜索问题。对加密文件进行搜索相对于基于明文搜索来说,实现起来有些局限性,不能像明文搜索那样方便,只能通过可搜索加密技术搜索指定关键词的加密文件,而不能更为精确地搜索指定的文件; 不过通过可搜索加密技术,搜索包含指定关键词的加密文件也一定程度上减少了计算开销和网络传输数据的开销。

在医疗区块链数据库中查询指定关键词的步骤如下:

1) 用户想查找关键词为 W 的加密医疗数据,使用可搜索加密算法生成陷门 T_w = Trapdoor(sk_peks, W) ,sk_peks 为私钥,W 为关键词,T_w 为陷门。生成查询交易单后发送查找请求给区块链数据库上的共识节点。

2) 区块链数据库上的共识节点收到查找请求后,从查找请求中提取出陷门,接着执行可搜索加密匹配算法 b = Test(pk_peks, C_w, T_w) 匹配出结果,若 b = 1,则表示查询成功,b = 0 表示查询失败。其中 pk_peks 为公钥,C_w 为关键词密文集。

3) 用户接收区块链数据库返回来的查找结果,从结果中得到包含关键词为 W 的加密医疗数据,然后用密钥对交易单中的加密数据进行解密后得到明文医疗数据。用户若想验证存储的医疗数据文件是否被篡改,则可以对加密文件进行哈希值计算,若得出的哈希值与交易单中记录的哈希值一样,则说明文件正确无误。

在医疗区块链数据库中查询指定关键词的陷门制作函数 Trapdoor 和匹配函数 Test 伪代码如下:

```
void Trapdoor( element_t T_w, element_t alpha, element_t H1_w) {
    /* 输入待计算的陷门 T_w,可搜索加密私钥 alpha,包含关键词信息 element_t 类型的 H1_w* /
    element_init_G1 ( T_w) ;              //对 T_w 所在的群进行初始化
    element_pow_zn ( T_w, H1_w, alpha) ;
    //对 T_w 进行运算,使其包含私钥和关键词信息
}
int Test ( pk_peks, C_w, T_w) {
    /* 输入关键词陷门 T_w,关键词密文集 C_w,可搜索加密公钥 pk_peks,输出查找密文 C_w * /
    for ( id = 0; id < max; id ++) {
        /* max 为区块链数据库中交易单的总数* /
        temp = pairing_apply( T_w, C_w)
        char * char_t =
            malloc( sizeof( char) * element_length( temp) ) ;
        element_snprint ( char_temp, temp) ;
        /* 对区块链数据库存储的交易单中的加密关键词进行双线性映射,得到 temp* /
        match = memcmp ( temp, C_w)
        /* 对 temp 和区块链数据库的加密关键词中的结果进行匹配,得到的匹配结果保存在 match 变量中 * /
        f( match = = 1) {
        /* 说明匹配成功,获得密文并保存下来,否则匹配失败* /}
        //接着对区块链数据库中所有的关键词密文进行匹配
    }
    //匹配完成后返回获得的密文集合,并传回给用户
}
```

4 模型实现及评估

可搜索加密用到的双线性映射,需要用到 PBC(Pairing-Based Cryptography) 函数库,它是已经开发好可以完全实现双线性映射的函数库; 同时还需要用到开源数学运算库(GNU MP Bignum Library,GMP) ,它是进行任意精度运算的函数库。

4.1 节点配置

本文实验共 5 个节点,每个节点的配置如表 1。

表 1　本模型所用节点配置如下

节点	CPU 核心数	CPU 频率/GHz	内存/GB	系统	备注
1	1	1.6 ~ 3.4	1.4	Ubuntu	主节点/用户
2	2	2.6	2.0	Ubuntu	从节点/用户
3	1	2..4 ~ 2.9	2.2	Ubuntu	从节点/用户
4	1	1.8 ~ 2.6	2.0	Ubuntu	从节点/用户
5	2	1.8 ~ 2.6	2.0	Ubuntu	用户

搭建区块链数据库: 节点 1、2、3、4 为重量级用户,是区块链数据库中的数据存储结节,同时也是区块链数据库的使用者; 节点 5 是轻量级用户,是区块链数据库的使用者,没有数据存储功能,它的数据存储在区块链数据库中。

针对区块链数据库中存储的内容,本次测试是通过 5 个用户产生文本文档作为医疗数据,进行加密后上传至区块链数据库中存储,每个文档的大小不一,同时每个文档的关键词由用户产生。

4.2 加密医疗数据上链时间

通过 5 个节点生成不同的医疗数据,每个用户上传的医疗数据都为 500 个,每个加密医疗数据文件大小不一,但是文件都不大,仅由文本文件和小图片组成,同时 5 个用户各自指定每个医疗文件的关键字,实验测得 500、1 000、1 500、2 000、2 500个加密医疗数据文件上链所需要的时间如表 2。

表 2　加密文件上链时间

加密医疗文件个数	上链时间/s
500	16.93
1 000	20.54
1 500	26.79
2 000	34.02
2 500	37.23

4.3 加密医疗文件查找时间

本次实验分别对区块链数据库中存储的医疗数据关键词个数为 300、600、900、1 200、1 500、1 800、2 100、2 400 进行了指定关键词查找,测得加密医疗数据搜索指定关键词时间如表

3。可以看出，当区块链数据库里的加密文件共有2 400个关键词时，主节点的搜索指定关键词时间是11.23 s，说明用户在加密的区块链数据库中进行关键词搜索是可行的。

表3 加密文件查找时间

关键词个数	查找时间/s
300	1.25
600	3.31
900	4.09
1 200	7.11
1 500	8.03
1 800	8.92
2 100	10.72
2 400	11.23

5 结语

本文提出并设计了基于可搜索加密的区块链数据隐私保护机制，以有效解决个人医疗信息泄露问题，并通过实验证明该机制有效可行。下一步我们将优化区块链数据库与可搜索技术，考虑实现更优的可搜索加密算法，比如寻求表达能力丰富的可搜索加密方案、更加高效的加解密方案等，充分发挥两者优势。由于本模型还处于实验阶段，可搜索加密的时间效率不是很高，因此未来将尝试推广可搜索加密来保护区块链数据的隐私，用来应用到实际生活应用中对数据隐私要求比较高的场景中。

参考文献：

[1] 邵奇峰，金澈清，张召，等. 区块链技术：架构及进展[J]. 计算机学报，2018，425(05)：3－22.

[2] Hyperledger architecture working group paper[EB/OL]. (2017－06－10) [2018－09－10]. https://www.hyperledger.org/.

[3] HUANG B，LIU Z，CHEN J，et al. Behavior pattern clustering in blockchain networks [J]. Multimedia Tools & Applications，2017，76(19)：20099－20110.

[4] RUFFING T，NORENO-SANCHEZ P，KATE A. CoinShuffle: practical decentralized coin mixing for bitcoin[C]// Proceedings of the 2014 European Symposium on Research in Computer Security. Cham：Springer，2014：345－364.

[5] SASSON E B，CHIESA A，GERMAN C，et al. Zerocash：Decentralized anonymous payments from bitcoin[C] // Proceedings of the 2014 IEEE Symposium on Security and Privacy. Piscataway，NJ：IEEE，2014：459－474.

[6] 祝烈煌，高峰，沈蒙，等. 区块链隐私保护研究综述[J]. 计算机研究与发展，2017，54(10)：2170－2186.

[7] 陈捷，高英. 区块链在物联网隐私保护中的应用[J]. 物流技术，2018，37(7)：33－38，133.

[8] 章宁，钟珊. 基于区块链的个人隐私保护机制[J]. 计算机应用，2017，37(10)：2787－2793.

[9] 梅颖. 安全存储医疗记录的区块链方法研究[J]. 江西师范大学学报(自然科学版)，2017，41(5)：484－490

[10] 袁勇，王飞跃. 区块链技术发展现状与展望[J]. 自动化学报，2016，42(4)：481－494.

[11] 刘敖迪，杜学绘，王娜，等. 区块链技术及其在信息安全领域的研究进展[J]. 软件学报，2018，29(7)：2092－2115.

[12] 刘肖飞. 基于动态授权的拜占庭容错共识算法的区块链性能改进研究[D]. 杭州：浙江大学，2017.

[13] 董晓蕾，周俊，曹珍富. 可搜索加密研究进展[J]. 计算机研究与发展，2017，54(10)：2107－2120.

[14] BONEH D，CRESCENZO G D，OSTROVSKY R，et al. Public key encryption with keyword search[C]// EUROCRYPT 2004：Advances in Cryptology. Heidelberg：Springer-Verlag Berlin，2004：506－522.

[15] CAI C，YUAN X，WANG C. Towards trustworthy and private keyword search in encrypted decentralized storage[C]// Proceedings of the 2017 IEEE International Conference on Communications. Piscataway，NJ：IEEE，2017：1－7.