区块链上基于B+树索引结构的密文排序搜索方案

牛淑芬*[①] 王金凤[①] 王伯彬[①] 贾向东[①] 杜小妮[②]

[①](西北师范大学计算机科学与工程学院 兰州 730070)

[②](西北师范大学数学与统计学院 兰州 730070)

摘 要：为了克服云存储不可信及云存储中密文检索效率低的问题，该文提出区块链上基于B+树的密文排序可搜索加密方案。该方案结合区块链技术解决了在互不了解的多方建立可靠信任的问题；使用向量空间模型降低了文本的复杂性实现了高效的文本检索系统；采用B+树的索引结构提高了区块链上密文交易的检索速度；利用加权统计(TF-IDF)算法实现了多关键词查询结果的排序。在随机预言机模型下，证明该方案是适应性不可区分安全的，通过效率对比分析，表明该方案在区块链上实现了高效的密文检索。

关键词：云存储；区块链；B+树；排序搜索

中图分类号：TP309.7 **文献标识码**：A **文章编号**：1009-5896(2019)10-2409-07

DOI: 10.11999/JEIT190038

Ciphertext Sorting Search Scheme Based on B+ Tree Index Structure on Blockchain

NIU Shufen[①] WANG Jinfeng[①] WANG Bobin[①] JIA Xiangdong[①] DU Xiaoni[②]

[①](*School of Computer Science and Engineering*, *Northwest Normal University*, *Lanzhou* 730070, *China*)

[②](*College of Mathematics and Statistics*, *Northwest Normal University*, *Lanzhou* 730070, *china*)

Abstract: In order to overcome the problem that cloud storage is not trusted and the low efficiency of ciphertext retrieval in cloud storage, a searchable ciphertext sorting encryption scheme based on B+ tree on the block chain is proposed. Combined with the blockchain technology, the problem of establishing reliable trust in multiple parties that do not understand each other is solved. A vector space model is used to reduce the complexity of the text and an efficient text retrieval system is implemented. The index structure of the B+ tree is used to improve the retrieval of ciphertext transactions on the blockchain. The ranking of multi-keyword query results is realized by the Term Frequency–Inverse Document Frequency (TF-IDF) algorithm. Under the random oracle model, it is proved that the scheme is adaptive and indistinguishable. Through the comparative analysis of efficiency, it is shown that the scheme achieves efficient ciphertext retrieval on the blockchain.

Key words: Cloud storage; Blockchain; B+ tree; Sorting search

1 引言

随着信息技术的快速发展，越来越多的数据需要存储在云服务器上，但现有的云存储服务器不是完全可信的。为了实现数据的保密，数据拥有者将数据做加密处理，并将数据以密文的形式上传到云存储服务器，使得云服务器无法获得有关明文数据的内容，但如何对云服务器上的密文进行检索是一困难问题。2000年Song等人[1]首次提出了对称可搜索加密算法，解决了密文上的搜索问题，但该方案是基于线性扫描搜索效率极低。Curtmola等人[2]提出了安全索引的定义，并采用布隆过滤器构建了安全索引，但搜索结果却存在部分错误。Golle等人[3]提出了链接关键词的可搜索方案，可以使用户检索多个关键词，但方案中需要指定关键词的位置。Ma等人[4]提出了移动医疗系统下的公钥可搜索加密方案，该方案可以抵御随机预言机模型中的选择关键词攻击，但不能抵抗内部服务器的关键词猜测攻

收稿日期：2019-01-15；改回日期：2019-06-05；网络出版：2019-06-12

*通信作者： 牛淑芬 sfniu76@nwnu.edu.cn

基金项目：国家自然科学基金(61562077, 61462077, 61662071, 61662069)，西北师范大学青年教师科研提升计划(NWNU-LKQN-14-7)，甘肃省杰出青年项目(1308RJDA007)

Foundation Items: The National Natural Science Foundation of China (61562077, 61462077, 61662071, 61662069), The Young Teacher's Scientific Research Ability Promotion Program of Northwest Normal University (NWNU-LKQN-14-7), The Natural Science Foundation of Gansu Province for Distinguished Young Scholars (1308RJDA007)

击。为了防止内部关键字猜测攻击，Huang等人[5]提出了一种安全有效的可搜索加密方案。为了提高检索结果的准确性，Xia等人[6]提出了支持动态更新的多关键词排序搜索方案，该方案采用平衡二叉树建立索引并利用贪婪深度优先搜索算法实现检索结果的排序。杨旸等人[7]提出了能实现数据隐私保护的多关键词语义排序搜索方案，该方案不仅提高了数据搜索效率，而且返回了更加满足用户需求的搜索结果。

上述所有方案中都假设云存储服务器是诚实的，并返回的搜索结果都是正确的。事实上，目前存在的云服务器是半诚实且好奇的。通过使用现有的技术，来解决云服务器的不可信问题，并实现数据的共享。目前区块链的应用是一个研究热点，因为区块链技术[8]能自由地实现数据的访问和共享。瑞士的Healthbank公司采用区块链公开透明技术保证了数据存储的绝对安全[9]，Lvan[10]提出了一种基于区块链上的电子病历安全存储方法。Andrychowicz等人[11]将比特币引入安全多方计算中来解决公平性问题。Dagher和Xia等人[12,13]利用区块链不可篡改、可追溯的特点分别提出了可以应用区块链的方案，提供了区块链上的数据存储结构，但是没有提供有效的搜索方法。Li等人[14]提出了区块链上的单关键字可搜索加密方案,实现了区块链上数据的检索，但搜索结果不准确且搜索效率低。Zhang等人[15]首次实现了服务器端数据的可验证性，在数据存储阶段保护了诚实的云服务器免受恶意数据所有者的陷害，使用区块链技术和散列函数，可以在不引入第三方的情况下，实现搜索费用的支付公平性，但并没有实现区块链上密文交易的精确搜索。

现有的云存储服务器是半诚实的，区块链技术具有去中心化、集体维护、高度透明和去信任的特点，因此，如何在区块链上实现加密文件和索引的存储以及密文的高效检索还有待解决。在检索过程中，使得搜索节点尽可能泄露少量的信息，同时能使区块链系统中的其他节点无需解密加密文档，就可以替用户查询满足搜索条件的密文文档，并对满足条件的文档进行相关度排序。

本文针对上述问题，在Xia等人[6]方案的基础上提出了区块链上支持B+树索引结构的密文排序搜索方法，并对本方案进行了安全性和效率分析。本文主要贡献为：

(1) 本方案利用区块链技术不仅可以为数据存储及检索提供安全、透明的通信保障，还能通过去中心化的共识机制，解决了传统的云存储服务器安全性低、可靠性差以及不诚实搜索的问题，实现了数据真正的共享。

(2) 利用空间向量模型将索引和查询关键词向量化，降低了文本信息的复杂性，并使用B+树构造了索引树，降低了索引结构的空间复杂度，同时也提高了区块链上密文交易的检索速度。

(3) 使用加权统计算法(Term Frequency–Inverse Document Frequency, TF-IDF)衡量加密文件与搜索陷门之间的相似度，降低了传输过程的通信开销，并提高了密文检索的精确度和用户的查找效率，实现了区块链上的密文排序检索。

2 预备知识

2.1 相关定义

(1) TF-IDF加权统计方法

TF-IDF是一种用于信息检索的加权技术，用来度量一个关键词对于文件集中的其中一份文件的重要程度。TF是词频表示一个关键词在某个文件中出现的频率，IDF是逆文档频率用来衡量一个关键词的普遍重要性。在向量模型中，每个文件可以表示成一个向量，该向量由关键字的TF值组成，查询向量由检索关键词的IDF值组成。在本方案中，当节点u是B+树的内部节点时，通过u的子节点的词频向量计算节点u的词频向量$\boldsymbol{D}$，当节点u为叶子节点时，具体计算如式(1)所示

$$\mathrm{TF}_{w_j}=\frac{\mathrm{TF}'_{f,w_j}}{\sqrt{\sum\limits_{w_j\in W}\left(\mathrm{TF}_{f,w_j}\right)^2}} \tag{1}$$

其中，$\mathrm{TF}'_{f,w_j}=1+\ln N_{f,w_j}$是文件$f$包含的关键词$w_j\,(1\le j\le m)$的词频值，$N_{f,w_j}$是文件$f$中关键词$w_j$的数量。查询关键词的向量计算如式(2)所示

$$\mathrm{IDF}_{w'_j}=\frac{\mathrm{IDF}_{w'_j}{}'}{\sqrt{\sum\limits_{w'_j\in W'}\left(\mathrm{IDF}_{w'_j}{}'\right)^2}} \tag{2}$$

其中，$\mathrm{IDF}_{w'_j}{}'=\ln\left(1+N/N_{w'_j}\right)$是查询关键词$w'_j$的逆文档频率，$N_{w'_j}$是包含查询关键词$w'_j$的文档数量，$N$是总文件的数量。相关性分数计算函数如式(3)所示

$$\mathrm{CS}\left(\boldsymbol{D}_u,\boldsymbol{Q}\right)=\boldsymbol{D}_u\cdot\boldsymbol{Q}=\sum_{w'_j\in W'}\mathrm{TF}_{w'_j}\cdot\mathrm{IDF}_{w'_j} \tag{3}$$

其中，$\boldsymbol{D}_u$为节点u处的词频向量，$\boldsymbol{Q}$为查询关键词的向量。

(2) B+树索引结构

在数据检索中，使用B+树构造索引的应用比较广泛，其检索时间与树的高度成正比，因此本方案选择使用B+树作为索引结构。其中B+树的每一

个叶子节点存储着与文档相关的信息，内部节点是基于叶子节点生成的。为了使搜索者能够在区块上快速查找到与陷门交易相关的密文文档的交易标识符，数据拥有者将与密文交易相关的交易标识符存储在B+树的叶子节点中，其中节点u的结构如式(4)所示

$$u=(N,\boldsymbol{D},P_{\mathrm{l}},P_{\mathrm{r}},\mathrm{TXid}) \tag{4}$$

其中，N为树形索引结构中的节点编号，$\boldsymbol{D}$为每个文档的m个关键词组成的m维向量，P_{l}和P_{r}分别是节点u的左指针和右指针，如果u为索引树的叶子节点，则将密文文档的交易标识符TXid存储在叶子节点中，如果u为内部节点，则节点向量$\boldsymbol{D}$的计算如式(5)所示

$$\boldsymbol{D}[j]=\max\{u.P_{\mathrm{l}}\to\boldsymbol{D}[j],u.P_{\mathrm{r}}\to\boldsymbol{D}[j]\},\quad j=1,2,\cdots,m \tag{5}$$

节点插入B+树的过程是建立密文交易索引的过程，算法的具体构造如表1所示。

表 1　B+树算法

算法1 BuildIndexTree(I)
if(u是叶子节点)then
将密文交易标识符TXid放到叶子节点，并计算叶子节点的$\boldsymbol{D}$的向量
return
else
根据新节点的位置，向下查找该节点插入的子节点ul
if(节点ul的数值为最大)then
对节点进行分割，重新确定向下插入的子节点ul
end if

2.2　系统模型

本系统由$\{F,W,O,U,P,M\}$组成，其中F为数据拥有者O要存储在区块链上的文档集合，W为关键字词典，U为数据的使用者，P为在区块链系统中用户指定的搜索者或其中一个矿工，M为区块链系统中收集交易单的矿工，系统模型如图1所示

区块链上的多关键字可搜索加密主要有以下4个参与者，分别是数据拥有者O、用户U、搜索者P和维护区块链的矿工M，具体算法的形式化定义如下：

(1) KeyGen$\left(1^{\lambda}\right)$：是一个密钥生成算法，由数据拥有者O执行,输入安全参数λ，输出m维的密钥向量$\boldsymbol{k}$。

(2) Enc$(\boldsymbol{k},F)$：是一个加密算法，由数据拥有者执行，输入密钥向量$\boldsymbol{k}$和文档集F，输出密文集合，将密文$C_i(1\le i\le n)$嵌入交易$\mathrm{TX}_i(1\le i\le n)$

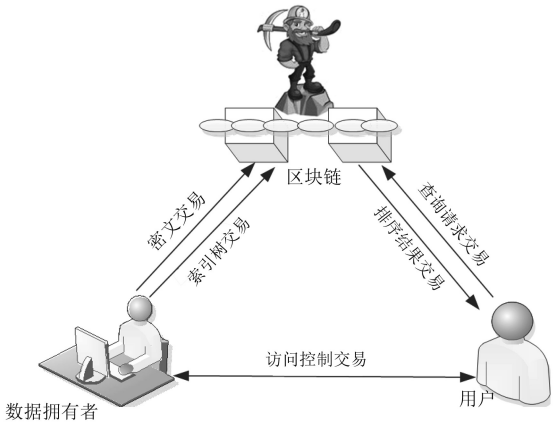

图 1 区块链系统检索图

中并向比特币网络广播该笔交易，然后由矿工写入区块链。

(3) BuildIndex$(\boldsymbol{k},W)$：是由数据拥有者来执行的索引构成算法，输入密钥向量$\boldsymbol{k}$和关键词词典W，输出关键词的索引向量，并构成索引树结构$\boldsymbol{I}$，将索引结构以交易$\mathrm{TX}_{\boldsymbol{I}}$的形式向全网广播。

(4) Trapdoor$(\boldsymbol{k},W')$：由节点用户U生成一个检索请求交易，输入密钥向量$\boldsymbol{k}$和查询关键字集W'，输出陷门$\boldsymbol{T}_{W'}$，将该陷门以交易$\mathrm{TX}_{\boldsymbol{T}_{W'}}$的形式向区块链网络广播。

(5) CalcScore$\left(\boldsymbol{I}_u,\boldsymbol{T}_{W'}\right)$：是一个确定性算法，由用户在陷门交易$\mathrm{TX}_{\boldsymbol{T}_{W'}}$中指定区块链系统中某一搜索者$P$来执行搜索包含关键字集$W'$的密文文档，输入索引树$\boldsymbol{I}_u$和陷门$\boldsymbol{T}_{W'}$，输出包含关键字$W'$的密文排序交易标识符TXid。

(6) Dec$(\boldsymbol{k},C)$：由用户U执行的解密算法，输入密钥向量$\boldsymbol{k}$和密文集合C，输出明文F。

2.3　安全性定义

在任何安全高效的模型中，大多数方案都会泄露一部分消息，为了弱化安全性定义，并允许泄露一部分消息给敌手，采用文献[14]中的方法，用一些泄露函数来弱化安全性定义，这些泄露函数指出了密文和陷门中泄露的信息，这些函数的定义如下：

定义1　搜索历史(query history)。文档集合F上的q次搜索历史为$H_q=(F,\varepsilon)$，其中$\varepsilon=(w_1,w_2,\cdots,w_q)$表示包含$q$个关键词的向量。

定义2　迹(trace)。由q次搜索历史$H=(F,\varepsilon)$产生$\tau(H)=(|\mathrm{TX}_1|,|\mathrm{TX}_2|,\cdots,|\mathrm{TX}_n|,\alpha(H),\delta(H))$，其中$|\mathrm{TX}_i|(1\le i\le n)$表示每笔交易的长度，$|\mathrm{TX}_{\boldsymbol{I}}|$表示索引的长度，$\alpha(H)$表示访问模式，$\delta(H)$表示搜索模式，迹是愿意泄露的有关搜索历史的信息泄露函数。

定义3 方案Π=(KeyGen, Enc, BuildIndex, Trapdoor, CalaScore, Dec)是适应性选择关键词攻击的不可区分性(INDistinguishably adaptive Chosen Keyword Attack, IND-CKA2)安全：假设A是敌手，S是一个挑战者模拟器，进行如下的概率实验：

$\mathrm{Real}_{\mathrm{A}}^{\Pi}(\lambda)$：挑战者随机选取一个安全参数$\lambda$，运行密钥生成算法生成密钥向量$\boldsymbol{k}$，敌手A随机选择两个文档，并将$(f_1, f_2)$发送给挑战者，挑战者将加密后的数据$(\boldsymbol{I}_b, C_b) \leftarrow \mathrm{Enc}(\boldsymbol{k}, F)$返回给敌手。敌手A在多项式时间内选择不同的关键字$w_i$向挑战者询问，挑战者返回$\boldsymbol{T}_{w_i} \leftarrow \mathrm{Trapdoor}(\boldsymbol{k}, w_i)$给敌手A，最后敌手输出1bit的$b'$，如果$b'=b$，则实验输出1，否则输出位0。

$\mathrm{Ideal}_{\mathrm{A}}{}^{\Pi}{}_{\mathrm{S}}(\lambda)$：敌手A随机选取文档，模拟器S通过泄露函数$\tau(H)$，模拟生成密文$(\boldsymbol{I}, C) \leftarrow S(\tau(H))$并发送给敌手A。敌手在多项式时间内选取不同的关键字向模拟器做一系列询问，模拟器根据$\alpha(H)$，$\delta(H)$将近似结果发送给敌手A，最后敌手输出1 bit的b'，如果$b'=b$，则实验输出1，否则输出位0。如果对于所有多项式时间内的敌手，存在一个多项式的模拟器S，则本文认为能够抵御适应性选择关键字攻击，即本方案Π是安全的。

3 区块链上基于B+树的密文排序搜索方案

区块链上的可搜索加密方案Π=(KeyGen, Enc, BuildIndex, Trapdoor, CalaScore, Dec)由6个多项式时内的算法构成，具体算法如下：

(1) $\mathrm{KeyGen}(\lambda) \rightarrow \boldsymbol{k}$：由数据拥有着执行该算法，输入安全参数$\lambda$，随机生成$m$维的秘密向量$\boldsymbol{\mu}$，以及两个$m \times m$维的可逆矩阵$\boldsymbol{M}_1$和$\boldsymbol{M}_2$，输出密钥$\boldsymbol{k}=(m, \boldsymbol{M}_1, \boldsymbol{M}_2)$。

(2) $\mathrm{Enc}(\boldsymbol{k}, F) \rightarrow C$：由数据拥有者执行该算法，输入密钥$\boldsymbol{k}$和文档集$F=(f_1, f_2, \cdots, f_n)$，数据拥有者$O$按如下步骤计算：

(a) 数据拥有者使用密钥$\boldsymbol{k}$将$F=(f_1, f_2, \cdots, f_n)$加密成密文$C_i = E(\boldsymbol{k}, f_i)\ (1 \le i \le n)$；

(b) 为了将加密后的密文C_i存储在区块链上，O需要找n个未花费的交易输出$\mathrm{UTX}_i\ (1 \le i \le n)$构造交易$\mathrm{TX}_{f_i}$；

(c) O将密文$C_i\ (1 \le i \le n)$嵌入交易TX_{f_i}，并对其签名，再以交易TX_{f_i}的形式向全区块链系统广播，由矿工将验证通过的交易记录到区块链上并给O返回一个嵌入密文的交易标识符$\mathrm{TXid}_i\ (1 \le i \le n)$。

(3) $\mathrm{BuildIndex}(\boldsymbol{k}, W) \rightarrow \boldsymbol{I}$：该算法由数据拥有者来构建索引树，B+树的构建如下：

(a) O对于每个关键词$w_j\ (1 \le j \le m)$，计算TF值，其中$\boldsymbol{D}=(\mathrm{TF}_{w_1}, \mathrm{TF}_{w_2}, \cdots, \mathrm{TF}_{w_m})$；

(b) 索引B+树中的每个节点存储向量$\boldsymbol{D}$，O根据密钥向量$\boldsymbol{\mu}$将节点u处的$\boldsymbol{D}$向量随机分成两个向量$\{\boldsymbol{D}_u', \boldsymbol{D}_u''\}$，如果$\boldsymbol{\mu}[j]=0$，则$\boldsymbol{D}_u'[j]=\boldsymbol{D}_u''[j]=\boldsymbol{D}_u[j]$；如果$\boldsymbol{\mu}[j]=1$，则$\boldsymbol{D}_u'[j]+\boldsymbol{D}_u''[j]=\boldsymbol{D}_u[j]$，B+树中的每个节点存储的加密向量为$\boldsymbol{I}_u=\{\boldsymbol{M}_1{}^{\mathrm{T}}\boldsymbol{D}_u', \boldsymbol{M}_2{}^{\mathrm{T}}\boldsymbol{D}_u''\}$；

(c) O需要找到一个未花费的交易输出UTX_0来计算交易$\mathrm{TX}_{\boldsymbol{I}}$，再将索引树$\boldsymbol{I}$嵌入到交易$\mathrm{TX}_{\boldsymbol{I}}$的外部脚本中，然后$O$以交易$\mathrm{TX}_{\boldsymbol{I}}$的形式向全网广播，矿工将索引交易上传到区块链后$M$将交易标识符$\mathrm{TXid}_{\boldsymbol{I}}$返回给$O$, O将$\mathrm{TXid}_{\boldsymbol{I}}$广播出去。

(4) $\mathrm{Trapdoor}(\boldsymbol{k}, W') \rightarrow \boldsymbol{T}_{W'}$：由用户执行该算法，根据查询关键集$W'$生成$m$维的查询向量$\boldsymbol{Q}$，$U$创建陷门交易的具体流程如下：

(a) 如果$w_j' \in W'$，$\boldsymbol{Q}[j]$存储关键字w_j'的IDF值，即$\boldsymbol{Q}[j]=\mathrm{IDF}_{w_j'}$，否则$\boldsymbol{Q}[j]=0$。查询向量$\boldsymbol{Q}[j]$是由两个随机向量$\boldsymbol{Q}'[j]$和$\boldsymbol{Q}''[j]$组成，如果$\boldsymbol{\mu}[j]=0$则$\boldsymbol{Q}'[j]+\boldsymbol{Q}''[j]=\boldsymbol{Q}[j]$，如果$\boldsymbol{\mu}[j]=1$则，陷门为$\boldsymbol{T}_{W'}=\{\boldsymbol{M}_1{}^{-1}\boldsymbol{Q}', \boldsymbol{M}_2{}^{-1}\boldsymbol{Q}''\}$；

(b) 用户U使用未花费的交易TX_y计算搜索交易T的主体，并将陷门$\boldsymbol{T}_{W'}$嵌入交易T的外部脚本，U对该交易签名后向全区块链系统广播交易T的交易标识符$\mathrm{TX}_{\boldsymbol{T}_{W'}}$。

(5) $\mathrm{CalcScore}(\boldsymbol{I}_u, \boldsymbol{T}_{W'}) \rightarrow \mathrm{Score}$：该算法由用户$U$在交易$T$中指定的搜索者$P$或其他矿工执行，$U$构造一笔搜索包含关键词$w'$的密文文档的交易$T$。

(a) 搜索者P根据交易T计算交易g的主体，运算$\phi(\mathrm{TX}_{T_{W'}}, \mathrm{TXid}_I)$获得$\boldsymbol{I}_{\boldsymbol{u}}$和$\boldsymbol{T}_{\boldsymbol{W}'}$，$P$计算$\theta=\mathrm{Score}(\boldsymbol{Q}, \boldsymbol{D}_u)=\boldsymbol{I}_u \cdot \boldsymbol{T}_{W'}=(\boldsymbol{M}_1{}^{\mathrm{T}}\boldsymbol{D}_u') \cdot (\boldsymbol{M}_1{}^{-1}\boldsymbol{Q}') + (\boldsymbol{M}_2{}^{\mathrm{T}}\boldsymbol{D}_u'') \cdot (\boldsymbol{M}_2{}^{-1}\boldsymbol{Q}'')=\boldsymbol{D}_u' \cdot \boldsymbol{Q}' + \boldsymbol{D}_u'' \cdot \boldsymbol{Q}''$，搜索者$P$根据$\theta$的大小将满足搜索条件的密文文档进行相关性分数排序；

(b) P将排序后的k个密文$C_i\ (1 \le i \le n)$嵌入交易g中，并向全区块链广播带有自己签名的交易，获得交易T中的服务费d\$；

(c) 若交易g没出现在区块链上，U可以通过创建一笔新的交易r来追回上一笔交易T中的手续费，其具体搜索流程如图2所示：

(6) $\mathrm{Dec}(\boldsymbol{k}, C) \rightarrow F$：数据使用者获得搜索者返回的与搜索关键字相关性最高的k篇文档C，数据使用者用密钥$\boldsymbol{k}$解密获得明文文档$f_i=\mathrm{Dec}(\boldsymbol{k}, C_i)\ (1 \le i \le n)$。

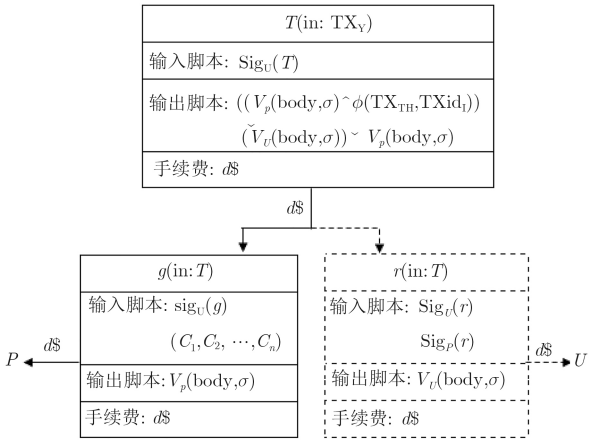

图 2 搜索相关度排序

4　安全性和效率分析

4.1　安全性分析

定理1 在随机预言机模型下，根据定义证明方案Π是IND-CKA2安全的，其中$\tau(H)$表示泄露的访问模式和搜索模式。

证明　存在概率多项式时间内的模拟器$S=\{S_0,S_1,\cdots,S_q\}$, $q=\text{ploy}(k)$，模拟器S使得对任意概率多项式时间内的攻击者$A=\{A_0,A_1,\cdots,A_q\}$, $\text{Real}_A^{\Pi}(k)$和$\text{Ideal}_A^{\Pi}{}_S(k)$的输出结果在计算上是不可区分的。然后运行$\text{Ideal}_A^{\Pi}{}_S(\lambda)$游戏，假设模拟器S被给予泄露函数$\tau(H)=(|\text{TX}_1|,|\text{TX}_2|,\cdots,|\text{TX}_n|,|\text{TX}_I|,\alpha(H),\delta(H))$，模拟器S通过泄露函数计算的到$\left(\text{TX}_1^*,\text{TX}_2^*,\cdots,\text{TX}_n^*,\text{TX}_I^*,\text{TX}_{T_{W'}}^*\right)$如下：通过证明加密、陷门和关键字的不可区分性来证明本方案是安全的。

(1) 模拟交易$\text{TX}_1^*,\text{TX}_2^*,\cdots,\text{TX}_n^*$

因为加密算法在选择明文攻击下是安全的，在游戏$\text{Ideal}_\text{A}^{\Pi}{}_\text{S}(k)$中，模拟器$S$生成的密文$C_1^*,C_2^*,\cdots,C_n^*$与在游戏$\text{Real}_A^{\Pi}(k)$中生成的密文$C_1,C_2,\cdots,C_n$在计算上是无法区分的，模拟器分别将密文以交易$\text{TX}_1^*,\text{TX}_2^*,\cdots,\text{TX}_n^*$的形式在区块链系统中向全网广播与$\text{Real}_\text{A}^{\Pi}(k)$游戏中挑战者生成的模拟器S可以从$\tau(H)$中获得文交易$\text{TX}_1,\text{TX}_2,\cdots,\text{TX}_n$是不可区分的。

(2) 模拟索引交易$\text{TX}_{\boldsymbol{I}}^*$

档嵌入每笔交易的交易长度，模拟器通过泄露函数中的信息随机选择长度为m的字符串作为$\boldsymbol{I}^*$，并将该索引树以交易$\text{TX}_{\boldsymbol{I}}^*$的形式发送给敌手，因为$\varepsilon=(\text{Enc},\text{Dec})$在选择明文攻击下是不可区分的，因此，$\text{TX}_{\boldsymbol{I}}^*$在计算上与$\text{TX}_{\boldsymbol{I}}$是不可区分的。

(3) 模拟陷门交易$\text{TX}_{\boldsymbol{T}_{W'}}^*$

敌手A向模拟器S查询关键字w_j的陷门，其中$w_j\in W(1\le j\le m)$，模拟器S收到该关键字后，查询该关键字是否之前询问过，若是，模拟器发送以前相同的陷门给敌手，否则模拟器通过泄露函数$\tau(H)$，生成一个相似的陷门$T_{w'}^*$并将该陷门以交易$\text{TX}_{T_{W'}}^*$的形式发送给敌手A，因此，在游戏$\text{Real}_\text{A}^{\Pi}(k)$中生成的陷门交易$\text{TX}_{\boldsymbol{T}_{W'}}$与在游戏$\text{Ideal}_\text{A}^{\Pi}{}_\text{S}(k)$中生成的近似陷门交易$\text{TX}_{\boldsymbol{T}_{W'}}^*$在计算结果上时无法区分的。

(4) 模拟排序结果交易g^*

当$q=0$时，如果A想得到交易T中的服务费，模拟器S根据访问模式和搜索模式，将$(C_1,C_2,\cdots,C_n)$返回给A，其中$C_i\leftarrow\{0,1\}^m$。当$q\ge1$时，模拟器S将$\left(C_{w_q1},C_{w_q}2,\cdots,C_{w_qn}\right)$返回给A，其中$C_{w_qi}(1\le i\le n)$表示关键词$w_q$的历史访问模式。因此$S$模拟一笔相似的交易结果$g^*$，使得通过游戏$\text{Ideal}_\text{A}^{\Pi}{}_S(k)$生成的交易$g^*$与游戏$\text{Real}_\text{A}^{\Pi}(k)$生成的交易$g$在计算以及在访问模上是不可区分的。

综上，$\text{Real}_\text{A}^{\Pi}(k)$的输出和$\text{Ideal}_\text{A}^{\Pi}{}_S(k)$的输出是不可区分的，根据定理1表明该方案是不可区分性安全的。本方案是通过敌手-挑战者游戏模式和敌手-挑战模拟器模式输出的结果存在不可区分性来证明此方案的安全性。现有的大多数方案选择基于敌手-挑战者游戏模式的安全性定义来证明方案的安全性，基于游戏的安全性定义弱化了敌手在攻击前所获得的一些信息，并且这种证明具有计算的隐藏性。本方案从加密、索引建立、陷门建立以及搜索结果在两种模式证明下存在不可区分性安全，该证明方法弱化了方案的安全性定义，使得方案具有更高的安全性。

4.2　效率分析

表2表示了本方案与其他方案在陷门时间复杂度、搜索时间复杂度等方面的区别，其中n为文档的个数，m为关键词词典的大小，φ为检索中包含搜索关键词$w_j(1\le j\le m)$的子集的叶子节点的数量。通过比较可以发现，本文方案具有搜索时间固定，满足交易的公平性并对满足搜索条件的结果进行排序的优势。在陷门时间复杂度方面，文献[14]和文献[16]的门陷较小，文献[7]和本方案具有相同的陷门计算量，其陷门计算量与检索关键词的数量有关系，在文献[7]和本方案中每个关键词的加密需要一个矩阵向量，从而导致需要更高的陷门计算量，但同时也实现了多关键词的检索。文献[14]和文献[16]实现的是单关键词的检索，故单关键词的陷门计算复杂度优于多关键词的陷门时间复杂度。

在搜索时间复杂度方面，本方案具有较高的检索效率。在本方案中，由于检索是从B+树的根节

表 2 效率对比分析

方案	|Trapdoor|	|SearchComplexity|	公平性	搜索结果是否排序
文献[6]	$O(m^2)$	$O(\varphi_m \log n)$	否	是
文献[14]	$O(1)$	$O(\|D(w)\|)$	是	否
文献[16]	$O(1)$	$O(\|D(w)\|)$	否	否
本文方案	$O(m^2)$	$O\left(\varphi_m \log_{\lceil \frac{m}{2} \rceil} \frac{n+1}{2}\right)$	是	是

点到叶子节点的遍历，关键词的搜索时间复杂度与B+树的高度有关系，方案中m阶索引树的高度为$\log_{\lceil m/2 \rceil}(n+1/2)+1$，因此B+树索引结构的搜索时间复杂度为$O\left(\varphi_m \log_{\lceil m/2 \rceil} n+1/2\right)$。事实上许多包含关键词的叶子结点都没有被访问，并且在被访问的叶子节点中，许多节点具有共同的父节点和相同的访问路径，每次搜索时不需要从索引树的根节点重新遍历，因此本方案的实际搜索时间要小于$O\left(\varphi_m \log_{\lceil m/2 \rceil} n+1/2\right)$。文献[14]和文献[16]的检索时间复杂度与检索关键词呈线性关系，在检索少量关键词的情况下具有较高的效率，当检索多个关键词时不再具有较高的效率，而基于B+树索引结构的最大检索时间复杂度是定值，而实际检索复杂度小于该最大值，因此本方案具有较高的搜索效率。

在公平上，文献[6]和文献[16]是基于云存储的检索，现有的云存储服务器是半可信的，不具有公平性，文献[14]及本方案是在区块链上检索的，根据区块链的特性可知区块链上的每个参与者都是公平的；对搜索结果的表示，文献[14]和文献[16]对检索结果没有排序，文献[7]及本方案使用空间向量模型对检索出的结果做了一定的相似度排序，提高了检索结果的准确度。

5 结论

本文提出了一个高效的区块链上基于B+树索引结构的密文排序方案，本方案在一定程度上解决了云存储服务器不可信和服务器恶意搜索的问题。本方案不仅可以实现区块链上加密文档的交易检索，同时还能利用向量空间模型、TF-ID加权统计方法对满足条件的检索结果进行相关性排序。并在随机预言机模型下证明该方案是安全的，通过效率对比分析表明本方案具有存储安全、共享安全、搜索效率固定且比其他方案效率高等优势。最后，搜索者需要通过6笔交易才能将包含搜索结果的交易上传到区块链上，这使的比特币系统中的每个参与节点都是公平的。

参 考 文 献

[1] SONG D X, WAGNER D, and PERRIG A. Practical techniques for searches on encrypted data[C]. 2000 IEEE Symposium on Security and Privacy, Berkeley, USA, 2000: 44–55.

[2] CURTMOLA R, GARAY J, KAMARA S, *et al*. Searchable symmetric encryption: Improved definitions and efficient constructions[C]. The 13th ACM Conference on Computer and Communications Security, Alexandria, USA, 2006: 79–88.

[3] GOLLE P, STADDON J, and WATERS B. Secure conjunctive keyword search over encrypted data[C]. The 2nd International Conference on Applied Cryptography and Network Security, Yellow Mountain, 2004: 31–45.

[4] MA Mimi, HE Debiao, KHAN M K, *et al*. Certificateless searchable public key encryption scheme for mobile healthcare system[J]. *Computers & Electrical Engineering*, 2018, 65: 413–424.

[5] HUANG Qiong and LI Hongbo. An efficient public-key searchable encryption scheme secure against inside keyword guessing attacks[J]. *Information Sciences*, 2017, 403-404: 1–14. doi: 10.1016/j.ins.2017.03.038.

[6] XIA Zhihua, WANG Xinhui, SUN Xingming, *et al*. A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2016, 27(2): 340–352. doi: 10.1109/tpds.2015.2401003.

[7] 杨旸, 刘佳, 蔡圣暐, 等. 云计算中保护数据隐私的快速多关键词语义排序搜索方案[J]. 计算机学报, 2018, 41(6): 1346–1359. doi: 10.11897/SP.J.1016.2018.01346.
YANG Yang, LIU Jia, CAI Shengwei, *et al*. Fast multi-keyword semantic ranked search in cloud computing[J]. *Chinese Journal of Computers*, 2018, 41(6): 1346–1359. doi: 10.11897/SP.J.1016.2018.01346.

[8] NAKAMOTO S. Bitcoin: A peer-to-peer electronic cash system[EB/OL]. https://bitcoin.org/en/bitcoin-paper, 2016.

[9] Healthbank[EB/OL]. http://www.healthbank.coop, 2016.

[10] LVAN D. Moving toward a blockchain-based method for the secure storage of patient records[EB/OL]. https://www.healthit.gov/sites/default/files/9-16-drew_ivan_20160804_blockchain_for_healthcare_final.pdf, 2016.

[11] ANDRYCHOWICZ M, DZIEMBOWSKI S, MALINOWSKI

D, *et al*. Fair two-party computations via Bitcoin deposits[C]. 2014 International Conference on Financial Cryptography and Data Security, Christ Church, Barbados, 2014: 105–121. doi: 10.1007/978-3-662-44774-1_8.

[12] DAGHER G G, MOHLER J, MILOJKOVIC M, *et al*. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology[J]. *Sustainable Cities and Society*, 2018, 39: 283–297. doi: 10.1016/j.scs.2018.02.014.

[13] XIA Qi, SIFAH E B, SMAHI A, *et al*. BBDS: Blockchain-based data sharing for electronic medical records in cloud environments[J]. *Information*, 2017, 8(2): 44. doi: 10.3390/info8020044.

[14] LI Huige, TIAN Haibo, ZHANG Fangguo, *et al*. Blockchain-based searchable symmetric encryption scheme[J]. *Computers & Electrical Engineering*, 2019, 73: 32–45. doi: 10.1016/j.compeleceng.2018.10.015.

[15] ZHANG Yinghui, DENG R H, SHU Jiangang, *et al*. TKSE: Trustworthy keyword search over encrypted data with two-side[J]. *IEEE Access*, 2018, 6: 31077–31087. doi: 10.1109/access.2018.2844400.

[16] 王尚平, 刘利军, 张亚玲. 可验证的基于词典的可搜索加密方案[J]. 软件学报, 2016, 27(5): 1301–1308. doi: 10.13328/j.cnki.jos.004912.

WANG Shangping, LIU Lijun, and ZHANG Yaling. Verifiable dictionary-based searchable encryption scheme[J]. *Journal of Software*, 2016, 27(5): 1301–1308. doi: 10.13328/j.cnki.jos.004912.

牛淑芬：女，1976年生，博士，副教授，研究方向为云计算、大数据和区块链上的数据安全.

王金风：女，1992年生，硕士，研究方向为区块链上的数据检索.

王伯彬：男，1992年生，硕士，研究方向为车联网隐私保护.

贾向东：男，1971年生，博士，教授，研究方向为无线传感器.

杜小妮：女，1972年生，博士，教授，研究方向为流密码和分组密码.