



研究与开发

基于区块链的结果可追溯的可搜索加密方案

翁昕耀, 游林, 蓝婷婷

(杭州电子科技大学, 浙江 杭州 310018)

摘要: 在可搜索加密方案中, 无论是云端服务器还是用户, 都可能存在欺骗行为。为了解决这种安全问题, 给出公平性安全的定义, 提出基于区块链的可搜索加密方案。通过第三方可信机构 (trusted authority, TA) 验证数据传输过程中数据的一致性, 区块链记录完整验证结果以防止篡改, 使所涉及的实体达成一致的安全共识, 从而实现公平性安全。安全性与复杂性分析表明, 该方案是可行的。当该方案与搜索结果可验证的可搜索加密方案相结合时, 可优化成基于区块链的搜索结果可验证的可搜索加密方案, 其安全性将得到进一步的提高。

关键词: 可搜索加密; 公平性安全; 区块链; 可信机构; 散列函数

中图分类号: TP309.7

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019183

Blockchain-based result-traceable searchable encryption scheme

WENG Xinyao, YOU Lin, LAN Tingting

Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: In a searchable encryption scheme, whether it is a cloud server or a user, it may be deceptive. In order to solve the security problem, the definition of fairness security was given and a new searchable encryption scheme based on blockchain was proposed. Through the third-party trusted authority to verify the consistency of data during data transmission, the complete verification results were recorded on the blockchain to prevent tampering, and it could make all the related entities to reach the consistent security consensus and achieve the fairness security. The security and complexity analysis show that the proposed searchable encryption scheme was feasible. If it is combined with a searchable encryption scheme with its searched results verifiable, the proposed scheme can be effectively improved to be a blockchain-based searchable encryption scheme with its searched results verifiable and its security will be more strengthened.

Key words: searchable encryption, fairness security, blockchain, trusted authority, hash function

收稿日期: 2019-03-22; 修回日期: 2019-06-28

通信作者: youlin@hdu.edu.cn

基金项目: 浙江省自然科学基金重点项目 (No. LZ17F020002); 国家自然科学基金资助项目 (No. 61772166)

Foundation Items: The Key Project of Natural Science Foundation of Zhejiang Province (No. LZ17F020002), The National Natural Science Foundation of China (No. 61772166)

1 引言

随着人工智能、物联网等技术的快速发展,大数据的重要性越来越凸显。在云存储技术已经成熟的当下,用户和企业都倾向于将数据存储在云端服务器。但云端服务器是不可信赖的,虽然提供了诸多便利,但仍存在各种安全问题。用户的敏感数据(比如电子邮件、个人医疗信息或金融数据等)以明文形式上传,就会有被第三方恶意窃取的风险,导致用户隐私泄露。仅2018年,全球就发生多起重大信息安全泄露事件。2018年1月,印度一个包含姓名、联系方式、指纹等敏感信息的10亿公民身份数据库 Aadhaar 遭泄露。2018年3月,Facebook 被暴露超过1.2亿用户的个人信息经由第三方合作公司泄露,被用于非法政治用途。

为解决上述安全问题,可搜索加密技术应运而生,在加密数据上实现关键词检索,只获取感兴趣的目标数据。在保证数据安全的前提下,提高数据的使用效率。最早的可搜索加密方案是由 Song 等人^[1]提出的 SWP 方案。Curtmola 等人^[2]参考了 Goh 等人^[3]的安全索引概念,提出了 SSE-1 和 SSE-2 方案。随着大数据时代的到来,可搜索加密方案的安全性受到越来越广泛的关注。Kamara 等人^[4]提出的可搜索加密云端存储系统 CS2 以及 Kurosawa 和 Ohtaki^[5]提出的 UC-security 可验证可搜索加密方案,是基于恶意服务器的假设构建的。此外,用户设备有受到攻击和密钥被

攻击者窃取的可能性^[6]。在多用户场景下,多个恶意用户也可能为了非法获取数据,联手欺骗云端服务器^[7-9]。因此,可搜索加密系统中的任何实体都可能是恶意的,会存在欺骗行为。

区块链^[10-11]是一种基于现代密码算法的分布式数据库,具有去中心化、难以篡改、可追溯、公开透明等优势。区块链的主要作用是存储信息,任何需要保存的信息,都可以写入区块链,也可以从中获取。区块链的结构如图1所示,由一个个串联的区块组成。每个区块包含两个部分:区块头(head)和区块数据(data)。在区块头上记录当前区块的特征值,如生成时间、区块数据的散列值、上一个区块的散列值等;在区块数据上记录实际的数据。

散列函数,是密码体制中能够将任意长度的消息映射成某一固定长度消息的公开函数。在密码学中,散列函数满足3个安全性方面的特性:单向性(one-way),如果对任意给定的 y ,寻找使 $h(x)=y$ 成立的 x 在计算上是困难的;弱抗碰撞性(weakly collision resistance),已知 x ,寻找 $x' \neq x$,使得 $h(x)=h(x')$ 在计算上是困难的;强抗碰撞性(strongly collision resistance),寻找两个不同的 x 和 x' ,使得 $h(x)=h(x')$ 在计算上是困难的。根据上述安全特性可以推断,散列值在区块链中起着决定性作用。因为每个区块都拥有一个独一无二的散列值,如果区块的内容发生变化,则它的散列值也一定会发生改变。

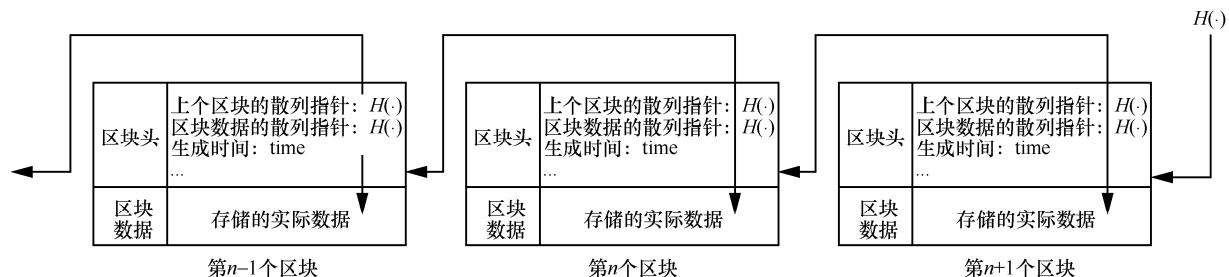


图1 区块链的结构



为了进一步提高可搜索加密技术的安全性, 本文提出基于区块链的可搜索加密方案, 主要贡献如下。

- 给出可搜索加密研究领域公平性安全的定义。完善可搜索加密的安全体系, 能同时保障用户和云端服务器的权益。
- 提出基于区块链的可搜索加密方案。结合区块链与第三方可信机构 (trusted authority, TA), 实现公平性安全。其中, 区块链起决定性作用。
- 本方案具有良好的拓展性, 能与现有可搜索加密方案相结合, 提高方案的安全性。
- 安全性分析验证了新方案的可行性与安全性, 复杂性分析验证了新方案的有效性。

2 系统模型

可搜索加密技术的一般模型主要由 3 个实体组成: 数据拥有者、数据使用者和云端服务器, 如图 2 所示。

系统的设计目的在于, 确保用户数据的安全性, 同时提高数据的使用效率。如果下列性质保持不变, 则可搜索加密方案是安全的: 当云端服务器只获得密文时, 它无法了解任何关于明文文档的信息; 当云端服务器执行搜索算法时, 它同样无法了解任何关于明文文档和关键词的任何信

息, 除了加密的检索结果。

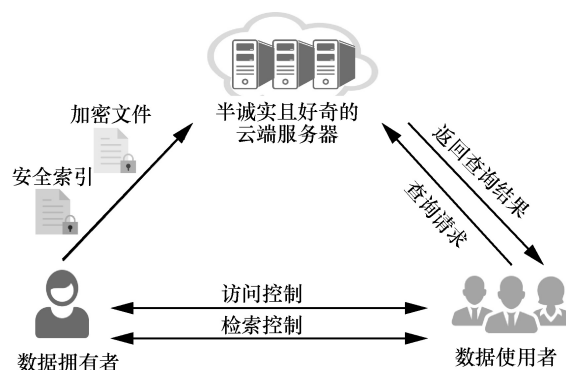


图2 可搜索加密一般模型

可搜索加密的主要过程可分为 7 个, 如图 3 所示。

(1) 索引生成 $I \leftarrow \text{Index}(F, W, K)$: 数据拥有者在本地执行索引构建算法, 扫描文档集合 $F = (f_1, f_2, \dots, f_n)$, 获得关键词集合 $W = (w_1, w_2, \dots, w_n)$ 。然后将密钥 K 、文档集合 F 和关键词集合 W 作为输入, 得到作为输出的索引 I 。

(2) 数据加密 $C, S \leftarrow \text{Enc}(F, I, K)$: 数据拥有者在本地执行加密算法, 将密钥 K 、文档集合 F 和索引 I 作为输入, 得到作为输出的加密文档集合 C 和加密索引 S , 然后将其上传给云端服务器。

(3) 权限授予 $\text{Auz}(K, \delta)$: 数据拥有者将密钥 K 和相关参数 δ 发送给数据使用者, 授予合法

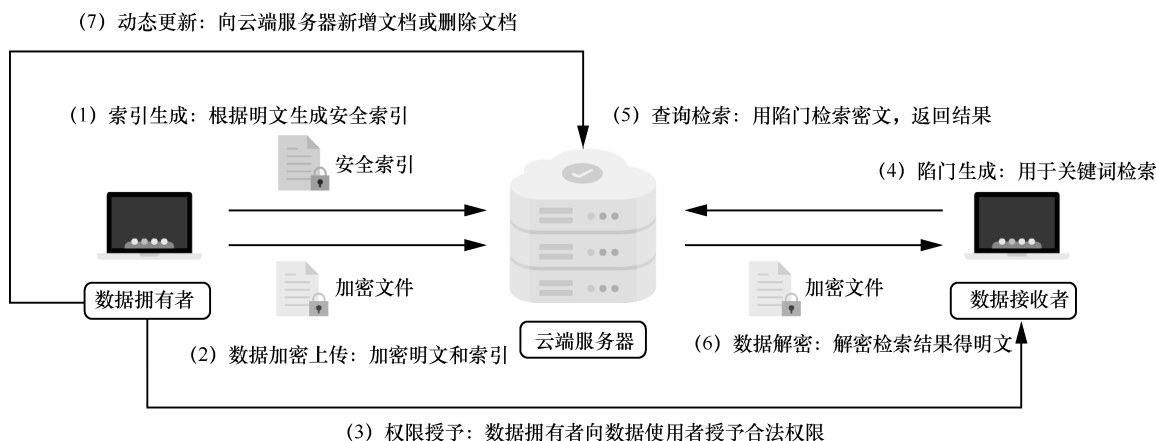


图3 可搜索加密主要过程

权限。

(4) 陷门生成 $TR_w \leftarrow \text{SrchToken}(w, K)$: 数据使用者在本地执行关键词检索陷门生成算法, 将密钥 K 和待检索关键词 w 作为输入, 得到作为输出的检索陷门 TR_w , 然后上传给云端服务器。

(5) 查询检索 $c \leftarrow \text{Search}(TR_w, C, S)$: 云端服务器执行检索算法, 输入检索陷门 TR_w 、加密文档集合 C 和加密索引 S , 得到检索结果 c 返回给数据使用者。

(6) 数据解密 $f \leftarrow \text{Enc}(c, K)$: 数据使用者在本地执行解密算法, 将密钥 K 和加密文档集合 c 作为输入, 得到作为输出的明文文档集合 f , 即关键词检索结果。

(7) 动态更新 $C', I' \leftarrow \text{Update}(F', C, K, I, \tau)$: 数据所有者向云端服务器发起文档更新请求, 新增文档或者删除文档。其中, 待更新的文档为 F' , 更新凭证为 τ ; 更新后的加密文档集合为 C' , 加密索引为 I' 。

3 公平性安全

伴随着可搜索加密方案的不断完善, 越来越多新的安全问题也随之产生。从最早的密文安全, 到后来的索引安全、关键词检索安全、动态更新安全等。经过研究后, 注意到可搜索加密仍旧存在被忽略的安全问题。

定义 1 (数据传输) 数据所有者成功发送数据, 数据接收者成功接收数据, 才能称作一次数据传输。

定义 2 (公平性安全) 在一次数据传输中, 数据发送者与数据接收者对外声明一致, 即数据发送者发送的数据与数据接收者接收的数据是相同的, 双方都不存在任何欺骗行为, 则认为方案实现了公平性安全。

在可搜索加密方案中, 数据所有者、数据使用者和云端服务器都有可能成为数据发送者或数

据接收者。对数据发送者来说, 应该如实告知发送了什么数据。对于数据接收者来说, 应该如实告知接收了什么数据。

无论是数据发送者, 还是数据接收者, 只要有一方发生欺骗行为, 就损害了方案的公平性安全。比如, 用户上传完整密文给云端服务器, 云端服务器收到完整密文, 但却谎称收到的密文有缺失, 恶意欺骗用户, 降低用户的信用评级。已知目前存在的可搜索加密方案, 都无法帮助用户验证云端服务器欺骗行为, 用户的公平性受到损害。同样, 用户也可能有欺骗行为。比如, 用户收到来自云端服务器返回的正确关键词检索结果, 但却谎称检索结果不正确, 陷害云端服务器。所以, 为了同时保障可搜索加密方案中不同实体的利益关系, 实现公平性安全是有必要的。

公平性安全模型 在任意的过程 $\text{Process}()$ 中, 实体 U 向其他实体 Q 发送数据, 实体 Q 接收来自实体 U 的数据。实体 Q 返回的接收结果可能是正确的 (true), 也可能是错误的 (false)。当且仅当实体 U 和 Q 都能判断双方发送或接收的数据是否一致以及识别存在欺骗行为的实体时, 可搜索加密方案是公平性安全的。

4 基于区块链的可搜索加密方案

4.1 基于区块链的可搜索加密方案系统模型

如果只依靠用户和服务器的自觉性, 是无法完全保证方案公平性安全的, 仍旧容易遭受不可信敌手的假冒攻击。而解决这个问题, 就需要第三方可信机构验证公平性以及区块链记录结果以防篡改, 从而实现方案的公平性安全。

在可搜索加密一般模型的基础上, 增加了第三方可信机构, 用于验证实体间传输数据的公平性。在每个实体处增加了区块链节点, 形成整体的区块链, 用于记录并同步验证结果, 如图 4 所示。

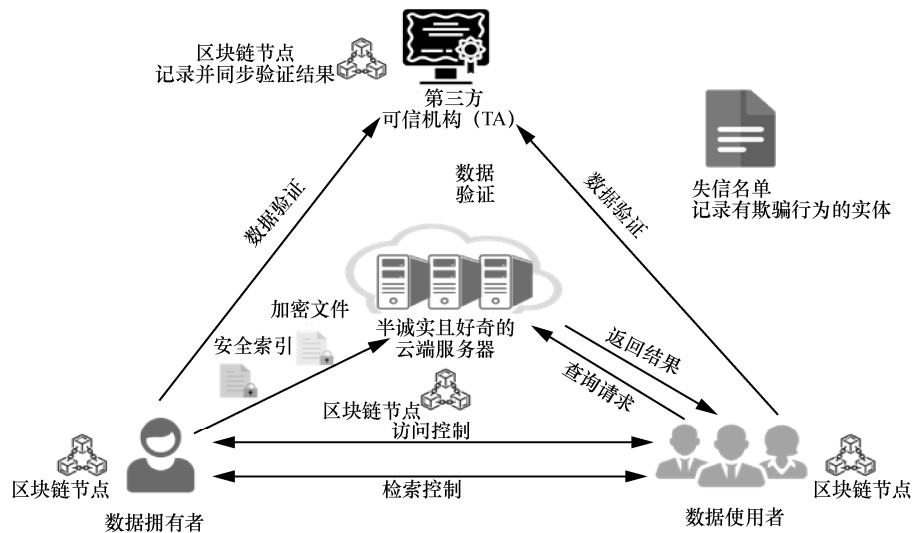


图4 基于区块链的可搜索加密方案的模型

此外，失信名单用于记录有欺骗行为的实体，为可搜索加密领域构建一个完整的信任体系。

4.2 第三方可信机构（TA）的验证流程

在可搜索加密方案中，任何实体都可能成为数据发送者和数据接收者。需要保证方案的公平性安全，也意味着需要保证方案中每次数据传输的公平性安全。第三方可信机构的验证流程，如图5所示。

步骤 1 数据所有者向数据接收者发送加密数据 $data$ ；

步骤 2 数据所有者计算加密数据 $data$ 的散列值 $Hash(data)$ ，上传给 TA；

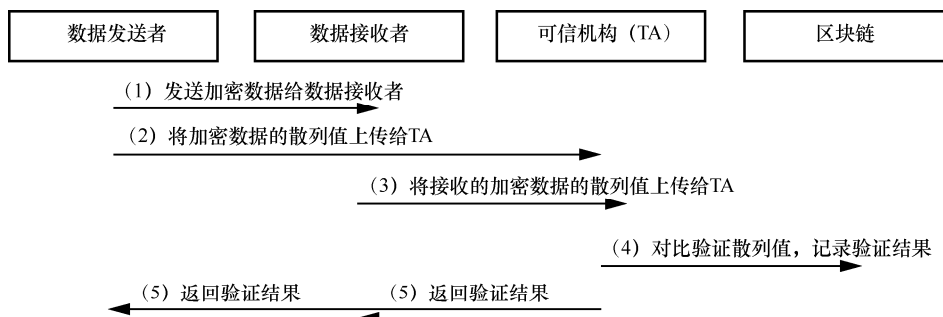
步骤 3 数据接收者计算接收到的加密数据

$data'$ 的散列值 $Hash(data')$ ，上传给 TA；

步骤 4 TA 对比验证散列值是否一致，并将验证结果记录到区块链上；

步骤 5 TA 将验证结果返回给数据发送者与数据接收者；如果第一次验证不成功，则重复步骤 2~5 继续验证，此时要求实体上传加密的数据，由 TA 来计算散列值 $Hashagain(data)$ 和 $Hashagain(data')$ 并对比验证。 $Hashagain(.)$ 是与 $Hash(.)$ 相同的散列函数，只是标记不一样。

(1) 如果 $Hashagain(data)$ 等于 $Hash(data)$ ，而 $Hashagain(data')$ 不等于 $Hash(data')$ ，则证明云端服务器是欺骗行为；如果 $Hashagain(data)$ 不等于 $Hash(data)$ ，而 $Hashagain(data')$ 等于 $Hash(data')$ ，



如果第一次验证不成功，则重复步骤 (2) ~ 步骤 (5) 继续验证，此时要求实体上传加密的数据

图5 第三方可信机构的验证流程

则证明数据拥有者是欺骗行为;

(2) 如果 $\text{Hashgain}(\text{data})$ 等于 $\text{Hash}(\text{data})$, 而 $\text{Hashgain}(\text{data}')$ 等于 $\text{Hashgain}(\text{data}')$, 则证明本次对比没有验证结果, 需进行第三次验证。再次要求数据拥有者和云端服务器上传各自拥有的加密数据, 由 TA 来计算对比散列值。在本次验证中, 虽不能证明哪一方是欺骗行为, 但能肯定两者中一定存在欺骗行为。

如果 TA 验证次数超过 5 次, 仍旧验证不出结果, 则将两者都添加到失信名单上, 留下信用可疑的记录。此外, TA 需要将每次的验证结果完整记录到区块链上。

4.3 区块链记录验证结果

为了保证可搜索加密方案的公平性安全, 也需要考虑第三方可信机构的安全性。因为其存在被恶意敌手攻击或串通其他实体来进行欺骗的可能性。

本文采用在区块链系统上记录验证结果, 使方案中所有涉及的实体达成一致的安全共识。区块链中每个参与节点都可以记录数据且保存着整个区块链的数据。区块链上的数据是永久记录且难以篡改的, 如果实体的欺骗行为被证实, 则实体会被列入失信名单, 造成巨大损失。此外, 区块链可追溯、公开透明等特点, 也使方案整体的安全性得到了一定完善。

TA 需要将每次对比认证的结果记录到区块链上。以一个区块为例, 如图 6 所示, 在区块头上记录上一个区块的散列值、区块数据的散列值、生成时间等其他相关参数。在区块数据中记录参与验证的实体、验证的次数、每次验证的内容、内容的散列值以及验证结果。每当区块链有更新时, 处于区块链中的所有节点需要同步更新, 例如数据拥有者和数据使用者的节点。

4.4 公平性安全验证过程

涉及公平性安全的, 在可搜索加密方案中主

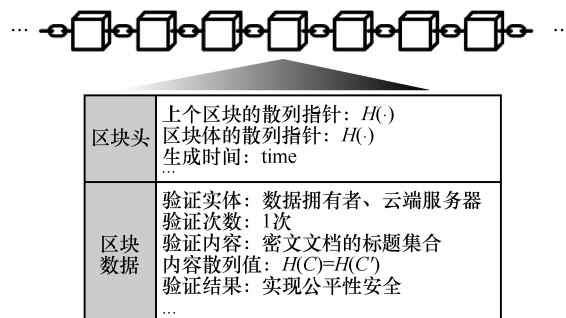


图 6 区块链中区块的结构

要有 4 个过程: 数据加密上传、权限授予、查询检索、动态更新。在不同过程中, 都需要第三方可信机构验证结果、区块链记录结果, 只是参与数据传输的实体不同, 传输的数据不同, 具体的区别如下。

(1) 数据上传过程

Upload(): 数据发送者为数据拥有者, 数据接收者为云端服务器, 传输的数据为加密后的密文 C 和索引 I 。

(2) 权限授予过程

Authorization(): 数据发送者为数据拥有者, 数据接收者为数据使用者, 传输的数据为密钥 K 和相关参数 δ 。

(3) 查询检索过程

Search(): 此过程涉及两次数据传输, 发起查询请求以及返回查询结果。其一, 数据发送者为数据使用者, 数据接收者为云端服务器, 传输的数据为加密的搜索陷门 TR_w ; 其二, 数据发送者为云端服务器, 数据接收者为数据使用者, 传输的数据为密文检索结果 R 。

(4) 动态更新过程

在动态更新过程中, 具体可以分为两种情况: 新增与删除。

- **Add()**: 数据发送者为数据拥有者, 数据接收者为云端服务器, 传输的数据为加密后的新增文档 F 与新增凭证 τ_a 。
- **Delete()**: 数据发送者为数据拥有者, 数据接收者为云端服务器, 传输的数据为加密后的删除凭证 τ_d 。



云端服务器根据删除凭证 τ_d 检索文档 D 并将其删除。此外,云端服务器删除文档 D 后,需要将删除结果返回给数据拥有者。此时,数据发送者变为云端服务器,数据接收者变为数据拥有者,传输的数据为删除的文档 D 。

5 安全性与复杂性

5.1 安全性

由于本文方案引入了区块链技术和第三方可信机构,可能带来新的安全风险。故下面将从两个方面证明本文方案的安全性。

定理 1 本文提出的方案是可行的。

证明如下。

(1) 第三方可信机构 TA 安全性。作为可信的第三方,TA 是与通信双方实体互相独立的,并且保持绝对的公平公正。在初次验证中,只需要将待验证的加密数据散列值发送给 TA。散列函数具有单向性,因此只获得了散列值的 TA 无法计算得到加密数据。在二次验证及后续验证中,需要将待验证的加密数据发送给 TA。数据都是加密的,没有解密密钥的 TA 无法从中获得关于明文数据的任何信息。在数据传输过程中,如果加密数据被第三方恶意敌手截获,只要解密密钥不泄露给敌手,同样也可以保证明文数据的安全。在结果验证完后,TA 还需要将结果完整记录到区块链上。由于区块链上记录的数据是难以篡改且可追溯的,这在一定程度上也约束了 TA 的行为。综上所述,第三方可信机构 TA 是安全的。

(2) 区块链安全性。在本方案中,记录到区块链上的相关数据有验证实体、验证次数、验证内容、内容散列值、验证结果,而这些内容都是可以公开的。其中,验证内容只是标题集合,用来指代本次验证是对什么数据进行验证的,不包括加密后的数据。根据散列函数的安全特性,仅通过内容散列值无法计算获得有关加密后数据的任何信息,从而确保区块链上记录数据的安全。

区块链本身是去中心化的,每个平等的节点都存储着完整的数据库,这意味着区块链是不能随意篡改的,并且公开透明,接受监督。综上所述,区块链是安全的。

因此,本文提出的方案是可行的。

定理 2 本文提出的方案是公平性安全的。

证明: 根据定义 1 可知,在一次数据传输中,数据发送者必定发送了数据,数据接收者必定收到了数据。根据定义 2 可知,数据发送者和数据接收者两者的数据完全一致,那么方案是属于公平性安全的。

在基于区块链的可搜索加密方案中,涉及数据传输的过程主要有:数据加密上传、权限授予、查询检索、动态更新。在这些过程中,通过第三方可信机构验证数据传输的公平性,当第一次验证成功时,则直接证明方案实现了公平性安全。若双方始终坚持自己的声明,第一次验证失败,则进行后续的验证,直到结果验证成功。若验证一直失败,在次数超过 5 次后,则证明数据传输的双方中必定存在欺骗者,是不公平的。每次验证的结果,将完整记录在区块链上,难以篡改但可追溯。综上所述,在可搜索加密方案的每次数据传输过程中,本方案可以保证各方实体享有公平对等的权益。

因此,本文提出的方案是公平性安全的。

5.2 复杂性分析

本文所提方案是基于一般可搜索加密模型,增加了区块链、第三方可信机构和失信名单,这 3 个实体是互相独立的,不影响可搜索加密方案主要操作。因此该方案能与现有的可搜索加密方案相结合,提高方案整体的安全性。比如,将本文方案与 Kurosawa 等人^[5]所提方案结合,与其他方案的比较结果,见表 1。其中, n 表示文档的数量, r 表示检索某个关键词时检索到的文档数量, p 表示并行服务器的数量, m 表示交易事务的大小。Kamara 等人^[12]所提方案采用多服

表 1 本文方案与其他方案的比较结果

方案	搜索复杂度	通信复杂度	可验证	敌手	公平性安全
Kurosawa 等人 ^[5]	$O(D(w))$	$O(r)$	是	恶意服务器	否
Kamara 等人 ^[12]	$O(r \log(n) / p)$	$O(1)$	是	诚实但好奇的服务器	否
Li 等人 ^[13]	$O(D(w))$	$O(6m)$	是	恶意服务器和恶意用户	是
本方案	$O(D(w))$	$O(r)$	是	恶意服务器和恶意用户	是

务器并行计算, 故其关键词检索效率最高, 但无法实现公平性安全。Li 等人^[13]所提方案虽然能实现公平性安全, 但其需要“数字货币”系统的支持, 且每次关键词检索至少需要 3 次数据通信和 6 次交易事务, 降低了获取检索结果的效率。本文所提方案是基于恶意服务器和恶意用户提的, 在关键词检索时, 只需要验证并记录传输数据散列值的一致性, 就能确保方案的公平性安全。

6 结束语

为解决可搜索加密中涉及的实体都可能存在的欺骗行为的问题, 本文给出了可搜索加密研究领域公平性安全的定义, 并提出基于区块链的可搜索加密方案。该方案可实现公平性安全, 使实体各方达成一致的安全共识, 具有占用资源少、可拓展性好等优势。实现公平性安全的意义在于: 同时保障实体各方的权益, 及时发现不可信赖的实体, 避免后续的损失。本文方案也存在不足, 如果双方实体传输数据本身就是错误的(物理存储故障等), 那么公平性安全检测会失效。这些问题将在今后的工作中研究与改进。

参考文献:

- [1] SONG D X, WAGNER D, PERRIG A. Practical techniques for searches on encrypted data[C]//the 2000 IEEE Symposium on Security and Privacy (S&P 2000), May 14-17, 2000, Washington, USA. Piscataway: IEEE Press, 2000: 44-55.
- [2] CURTMOLA R, GARAY J, KAMARA S, et al. Searchable symmetric encryption: improved definitions and efficient constructions[C]//the ACM Conference on Computer and Communications Security, Oct 30-Nov 3, 2006, Alexandria, USA. New York: ACM Press, 2006: 79-88.
- [3] GOH E J. Secure indexes[J]. IACR Cryptology ePrint Archive, 2003: 216.
- [4] KAMARA S, PAPAMANTHOU C, ROEDER T. CS2: A searchable cryptographic cloud storage system[R]. 2011.
- [5] KUROSAWA K, OHTAKI Y. UC-secure searchable symmetric encryption[C]//the International Conference on Financial Cryptography and Data Security, Feb 27- Mar 3, 2012, Kralendijk, Bonaire. Heidelberg: Springer, 2012: 285-298.
- [6] DAI S, LI H, ZHANG F. Memory leakage-resilient searchable symmetric encryption[J]. Future Generation Computer Systems, 2016(62): 76-84.
- [7] JARECKI S, JUTLA C, KRAWCZYK H, et al. Outsourced symmetric private information retrieval[C]//The 2013 ACM SIGSAC Conference on Computer & Communications Security, Nov 4 - 8, 2013, Berlin, Germany. New York: ACM Press, 2013: 875-888.
- [8] FISCH B A, VO B, KRELL F, et al. Malicious-client security in blind seer: a scalable private DBMS[C]//the 2015 IEEE Symposium on Security and Privacy, May 18-21, 2015, California, USA. Piscataway: IEEE Press, 2015: 395-410.
- [9] ISHAI Y, KUSHILEVITZ E, LU S, et al. Private large-scale databases with distributed searchable symmetric encryption[C]//The Cryptographers' Track at the RSA Conference, Feb 29 - Mar 4, 2016, San Francisco, USA. Cham: Springer, 2016: 90-107.
- [10] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system[Z]. 2008.
- [11] 李董, 魏进武. 区块链技术原理、应用领域及挑战[J]. 电信科学, 2016, 32(12): 20-25.
LI D, WEI J W. Theory, application fields and challenge of the blockchain technology[J]. Telecommunications Science, 2016, 32(12): 20-25.
- [12] KAMARA S, PAPAMANTHOU C. Parallel and dynamic



searchable symmetric encryption[C]//the International Conference on Financial Cryptography and Data Security, Apr 1-5, 2013, Okinawa, Japan. Heidelberg: Springer, 2013: 258-274.

[13] LI H, TIAN H, ZHANG F. Block chain based searchable symmetric encryption[J]. IACR Cryptology ePrint Archive, 2017: 447.

[作者简介]



翁昕耀（1994- ），男，杭州电子科技大学硕士生，主要研究方向为信息安全与可搜索加密。



游林（1966- ），男，博士，杭州电子科技大学网络空间安全学院教授、博士生导师，主要研究方向为密码学、区块链技术、生物特征识别及其应用等。



蓝婷婷（1994- ），女，杭州电子科技大学硕士生，主要研究方向为信息安全与生物特征加密技术。