计算机科学
COMPUTER SCIENCE

云环境下可验证关键词密文检索研究综述

周倩，戴华，盛文杰，胡正，杨庚

引用本文

周倩，戴华，盛文杰，胡正，杨庚. 云环境下可验证关键词密文检索研究综述[J]. 计算机科学, 2022, 49(10): 272-278.

ZHOU Qian, DAI Hua, SHENG Wen-jie, HU Zheng, YANG Geng. Research on Verifiable Keyword Search over Encrypted Cloud Data:A Survey[J]. Computer Science, 2022, 49(10): 272-278.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220500285

云环境下可验证关键词密文检索研究综述

周 倩[1] 戴 华[2,3] 盛文杰[2] 胡 正[2] 杨 庚[2,3]
1 南京邮电大学现代邮政学院 南京 210023
2 南京邮电大学计算机学院 南京 210023
3 江苏省大数据安全与智能处理重点实验室 南京 210023

摘 要 云计算便捷高效的特点使其拥有巨大的发展潜力,越来越多的企业与个人通过使用云计算提供的各类外包服务而获得实际收益。为了保护云端外包数据的私密性和一致性,具有隐私保护能力的可验证密文检索技术正逐渐成为当前云计算领域的一个研究热点。针对关键词密文检索的一致性验证问题,阐述现有研究工作主要采用的系统模型、威胁模型和通用框架;从可验证单关键词密文检索和可验证多关键词密文检索两个角度,综述现有研究工作的技术方案,并分析这些技术方案的优缺点;最后,通过综合分析和对比现有研究工作的研究重点及其所使用的关键技术,对现有工作进行总结,并展望未来可能的研究方向和趋势。

关键词:云计算;隐私保护;一致性验证;关键词检索
中图法分类号 TP391

Research on Verifiable Keyword Search over Encrypted Cloud Data:A Survey

ZHOU Qian[1],DAI Hua[2,3],SHENG Wen-jie[2],HU Zheng[2] and YANG Geng[2,3]
1 School of Modern Posts,Nanjing University of Posts and Telecommunications,Nanjing 210023,China
2 School of Computer Science,Nanjing University of Posts and Telecommunications,Nanjing 210023,China
3 Jiangsu Key Laboratory of Big Data Security and Intelligent Processing,Nanjing 210023,China

Abstract The convenience and efficiency of cloud computing have brought great potential for its development. More and more enterprises and individuals obtain real benefits by using various outsourcing services provided by cloud computing. In order to protect the confidentiality and integrity of outsourced data in the cloud,the keyword search over encrypted cloud data with privacy protection and integrity verification is becoming a research hotspot in the field of cloud computing. In this paper,we focus on the issue of the verifiable keyword search over encrypted data. The system models,threat models and frameworks adopted in the existing works are firstly introduced. Related works are overviewed from the aspects of verifiable single keyword search and verifiable multi-keyword search over encrypted data,and the ideas of these works are briefly described together with the advantages and disadvantages. At last,the conclusion is presented through a comprehensive analysis and comparison of the related works,and the possible research directions and trends in the future are prospected.

Keywords Cloud computing,Privacy protection,Integrity verification,Keyword search

1 引言

云计算自其诞生之初就因其按需付费、随时随地访问等优势而备受关注,云计算提供平台、存储及计算资源,为许多企业与个人解决了软硬件问题。然而,在提供服务与便捷的同时,也存在着外包数据安全的问题和隐患[1-6]。在云环境中,用户将数据上传至云服务器(Cloud Server,CS)存储,同时也失去了对数据的直接控制权,CS 中的数据所面临的各项风险将是用户无法预料和避免的。一方面,CS 中存储着大量数据,外部攻击者为获取最大化收益,通常将 CS 作为首选攻击目标以获取数据信息;另一方面,云服务提供商(Cloud Server Provider,CSP)也不一定是完全可信的,它可能会为用户提供不完整或不准确的搜索结果以节约计算开销。内部漏洞与外部威胁都是云计算必须面对的安全问题[7-13]。

为了保护云端数据的私密性,在外包数据上传至 CS 前进行加密是最直观有效的方法。而为了提高数据的可用性,

到稿日期:2022-05-31 返修日期:2022-08-06
基金项目:国家自然科学基金面上项目(61872197,61972209,61902199);中国博士后自然科学基金项目(2019M651919);南京邮电大学自然科学基金(NY217119,NY219142)
This work was supported by the National Natural Science Foundation of China(61872197,61972209,61902199),Postdoctoral Science Foundation of China(2019M651919) and Natural Science Foundation of NJUPT(NY217119,NY219142).
通信作者:戴华(daihua@njupt.edu.cn)

可搜索加密技术应运而生，该技术允许用户在密文中执行相关数据检索。然而，在保护数据私密性的同时，检索结果是否准确和完整也是云计算必须考虑的一个方面。在医疗、金融等对数据准确性要求极高的行业，数据的错误或者不完整可能会导致非常严重的后果。为了检测 CS 返回的检索结果是否被篡改或伪造，确保检索结果的正确性和完整性，有必要研究支持检索结果一致性验证的密文检索方法。基于此，可验证对称密文检索技术（Verifiable Symmetric Searchable Encryption，VSSE）越来越受到研究者们的关注和重视，目前也已经有一些研究方案陆续被提出。VSSE 方案的核心问题是如何设计和构建适用于云环境的可搜索加密索引结构、密文检索算法和检索结果一致性验证算法。

针对云环境下密文检索结果的一致性可验证问题，本文首先阐述现有研究工作主要采用的系统模型、威胁模型和通用框架；然后，从可验证单关键词密文检索和可验证多关键词密文检索两个不同的角度出发，对国内外已有的相关研究工作进行综述研究，分析其核心技术原理以及各自的优缺点；最后，通过综合分析和对比现有研究工作的研究重点及其所使用的关键技术，对现有工作进行总结，并对未来可能的研究方向和研究趋势进行展望。基于本文的研究和分析，在构建安全高效、便于动态更新的索引的基础上实现高效的检索结果一致性验证，是目前可验证密文检索技术研究中亟需解决的关键问题。

本文第 2 节介绍了相关模型、问题描述和通用框架；第 3 节和第 4 节分别从单关键词可验证密文检索和多关键词可验证密文检索这两个角度，对现有研究工作进行综述分析；最后总结全文并展望未来。

2　模型与问题描述

2.1　系统模型

云环境下可验证密文检索的系统模型如图 1 所示。

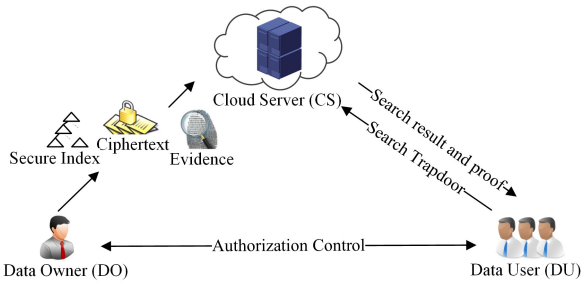

图 1　可验证关键词密文检索系统模型

Fig. 1　System model of verifiable keyword search over encrypted data

该系统模型主要包括 3 个不同的实体，分别为：数据拥有者（Data Owner，DO）、云服务器（Cloud Server，CS）和数据使用者（Data User，DU）。它们的主要工作方式如下：

（1）DO 对数据进行预处理，构造安全索引以及用于验证检索结果一致性的证据信息，同时对数据进行加密，最后将密文数据、安全索引和证据信息发送至 CS。

（2）CS 存储 DO 发送的外包数据；同时，CS 在接收到 DU 发起的检索请求后，根据收到的检索陷门执行密文检索，并将密文检索结果以及相应的证据信息发送给 DU。

（3）DU 在执行检索时，将检索关键词转换为检索陷门，并将其作为检索指令发送给 CS；当 DU 接收到 CS 返回的密文检索结果和证据信息时，利用证据信息验证收到的检索结果的一致性，并对检索结果进行解密，获得最终的明文检索结果。

2.2　威胁模型

在面向云计算环境的可验证关键词密文检索技术的研究中，DO 和 DU 均假设为可信参与方。根据 CS 的可信度的不同，一般采用如下两种威胁模型。

（1）诚实而好奇（Honest-but-curious）模型

在诚实而好奇模型下，CS 能够严格执行既定协议，对 DO 上传的数据进行诚实存储，同时执行搜索和计算等服务；但 CS 将对外包数据保持好奇，即 CS 存在企图通过窥探、分析等方法获取外包数据的私密信息。该模型又被称作“半可信”（Semi-Honest）威胁模型。

（2）恶意攻击（Malicious Attack）模型

在恶意攻击模型下，CS 完全不可信。CS 可能为了节省计算成本不再执行原有服务或只执行部分计算；同时，外包数据还可能会遭到 CS 的主动攻击，这些攻击包括篡改数据、伪造数据以及恶意删除等，且 CS 也将无法保证能抵抗外部攻击。在该威胁模型下，数据发生隐私泄露的可能性大大增加，同时 CS 返回的检索结果的一致性也无法得到保证。

在可搜索加密技术研究中，大部分研究工作都采用诚实而好奇威胁模型。然而，在现实情况下，由于各种内外部因素的干扰，也存在着 CS 不可信的应用场景，这就需要 DU 具备针对检索结果一致性的验证能力。现有的可验证关键词密文检索技术方案主要采用恶意攻击威胁模型，在保证外包数据机密性的同时，也支持针对检索结果的一致性验证。

2.3　问题描述

云环境下可验证关键词密文检索技术主要解决云计算技术目前面临的隐私保护和可验证性问题。其中，隐私保护主要实现对外包数据私密性的保护，包括文档的机密性、检索关键词的机密性等；可验证性主要实现对检索结果的一致性验证，具体内容包括准确性验证、完整性验证和新鲜度验证。

（1）准确性（Correctness）验证：检索结果满足检索要求，且未被篡改或伪造。

（2）完整性（Completeness）验证：检索结果包含满足检索要求的全部数据，不存在局部检索结果的缺失。

（3）新鲜度（Freshness）验证：检索结果是当前最“新鲜”的数据，检索结果中不存在旧版本的数据。

2.4　通用框架

现有研究工作实现可验证关键词密文检索主要采用包含如下 4 个算法的通用框架结构。

（1）Setup：初始化算法，主要负责完成数据外包前的预处理工作，包括生成密钥、构造索引、加密文档或索引、构造证据等步骤。该算法由 DO 负责执行，执行完成后 DO 将预处理后的数据外包存储至 CS 端。

（2）Trapdoor：陷门构造算法，主要负责为检索关键词构造检索陷门。该算法由 DU 负责执行，DU 将生成的检索陷门作为检索指令发送给 CS，检索陷门应满足不泄露

检索关键词任何信息的要求。

(3)Search:关键词检索执行算法,主要负责执行密文检索。该算法由 CS 负责执行,CS 在接收到 DU 发送的检索陷门后,利用加密索引执行密文检索,并将检索结果以及相应的证据信息返回给 DU。

(4)Verify:检索结果一致性验证算法。该算法由 DU 负责执行,DU 在接收到 CS 返回的检索结果和证据信息后,利用证据信息验证检索结果的一致性。

上述 4 个算法是大部分可验证关键词密文检索技术方案普遍采用的通用框架。此外,在一些支持动态更新或个性化搜索的解决方案中,还存在其他实现特定功能的算法,本节不做详细阐述。

3 可验证单关键词密文检索技术

在可验证密文检索研究中,针对单关键词的可验证密文检索方案最先被提出,此类方案重点关注针对单一关键词的检索,其实现检索结果一致性验证的结构和策略是此类方案的关键,决定了验证所需的时空开销以及抗攻击能力。在现有方案中,使用较多的验证技术有 Merkle 哈希树、RSA 累加器、签名链技术、位图技术等。

3.1 基于位图的可验证单关键词检索方法

文献[14]利用带密钥散列函数结合位图(Bitmap)技术构建了一种前缀树,首次提出了可验证单关键词密文检索方案。在该方案中,将前缀签名及位图数据保存在节点中,其中前缀签名用于检索,而位图数据保存了中间节点的孩子节点信息或终端节点的文档集信息。构建陷门时,用户将关键词的每一个明文字母依次通过带密钥散列函数生成专属签名,CS 通过陷门中每个字母的签名与节点中的前缀签名进行比较以实现检索。利用检索路径中每个节点的位图数据可重构关键词与文档集信息,以此实现对搜索结果一致性的验证。文献[15]结合文献[16]中实现模糊检索的技术提出了首个可验证的单关键词模糊检索方案,该方案利用编辑距离实现了模糊检索,同时将包含同一关键词的文档 ID 串联,通过对比 ID 串和检索路径的签名将验证所需的时间减少至 O(1)。但是,该方法只能实现在静态数据上的检索与验证,同时前缀树需要占用大量的空间。文献[17]结合位图与同态 MAC 提出了首个动态的可验证单关键词密文检索方案,但是该方案在构造索引、更新操作和验证检索结果的算法中需要更多的时空消耗,同时不支持对重放攻击的防范。

3.2 基于 Merkle 哈希树的可验证单关键词检索方法

Merkle 哈希树(Merkle Hash Tree,MHT)在文献[18]中首次被提出,其特点是通过根节点哈希值验证集合中元素的归属关系。文献[19]基于 MHT 结构提出了一种支持检索结果验证的单关键词密文检索方法,该方案分别构造了文档 MHT 和倒排 MHT,其中文档 MHT 用于文档的完整性验证,而倒排 MHT 则用于验证文档对关键词相关度分数的排名。这种对 MHT 的简单应用需要遍历整个倒排列表才能重构倒排 MHT,其空间利用率极低。针对这些问题,该文献又提出了一种针对倒排列表的 MHT 链,避免了遍历整个倒排列表的情况,但其空间消耗仍然偏高,降低了方案的实用性。使用 Merkle 哈希树作为验证方式的方案还包括文献[20-21]等,但在空间效率方面并没有有明显的改善。基于 MHT 的上述问题,文献[22]将 MHT 与 Patricia 树相结合,提出了 MPT(Merkle Patricia Trie)树结构,该方案通过将搜索路径上多个节点压缩至一个节点中来降低索引树的高度,提高搜索效率;同时将关键词中的字母转化为十六进制形式,从而控制索引树的宽度,并降低空间消耗。文献[23]首次将 MPT 树应用于云环境中的密文检索的一致性验证中,文献[24]在此基础上又引入增量哈希,该技术方案减小了集合中元素增减带来的计算量,提高了数据动态更新的效率。

3.3 基于链式结构的可验证单关键词检索方法

排序检索要求按用户需求将与被检索关键词相关度最高的 k 个文档作为检索结果返回给用户。文献[25]提出了基于链式结构的可验证单关键词排序检索,该方法将数据集中的元素按相关分数排序,将排序相邻的元素链接后计算一个电子签名。用户在收到检索结果后重构检索结果中每个链接元素的签名,由于签名中包含排序信息,因此可同时验证检索结果的一致性和排序正确性。然而,该方案需要给数据集中每一个链接元素计算签名,且计算过程复杂,在初始化以及验证阶段都需要耗费大量时间和计算资源。文献[26]提出根据关键词与包含该关键词的文档的相关度得分,并利用哈希身份认证算法构造面向关键词的倒排文档偏序约束链,该偏序约束链本质上也是一种链式结构;在验证阶段,通过重构偏序约束链即可验证检索结果的一致性以及排序正确性。链式结构的原理相对简单,但是其特性决定了其在空间消耗和动态更新方面有着无法改变的局限性。

3.4 基于 RSA 累加器的可验证单关键词排序检索方法

RSA 累加器是基于强 RSA 假设[27]实现集合元素归属关系验证的机制。文献[28]利用 RSA 累加器实现了动态的关键字密文检索,在该方案中,每个关键词和包含该关键词的密文数据通过一个素数实现关联,并利用模幂运算将该素数加入到一个 RSA 累加器中。同时,该方案将密文数据与关键词的相关度得分的排序信息用矩阵的形式表示,随后以相同的方式将该排序信息加入另一个 RSA 累加器中。在执行搜索时,将不包含检索结果密文集的累加器作为证明发送给用户;用户在收到检索结果后,通过重构累加器来验证检索结果的准确性及完整性,同时验证相关度得分排序的准确性。但是,该方法所设计的用于一致性验证的矩阵在数据规模较大的应用场景中不具有实用性,同时该方案更新较为复杂。基于这些问题,文献[29]提出了改进方案,用删除表替代了可验证矩阵,在提高删除操作效率的同时减少了空间成本。此外,文献[30-32]也提出基于 RSA 累加器的可验证密文检索方案,但 RSA 累加器机制需要进行大量的模幂计算,导致初始化阶段耗时较长,而且在更新时也需要更加复杂的预处理。

4 可验证多关键词密文检索技术

由于检索关键词数量的增加,可验证多关键词密文检索的技术难度也随之增大,可验证多关键词的排序检索则是一个更具挑战性的问题。本节针对目前已有的技术方案,从不同的功能和方法出发进行研究和分析,其中与可验证单关键

词密文检索技术中通用的方法在本节不再赘述。

4.1　基于消息认证码的可验证多关键词密文检索方法

文献[33]提出了一种基于同态 MAC 和随机挑战技术的可验证多关键词密文检索技术，该方法将文档索引分为两个向量，利用矩阵加密保障索引的机密性，并且为索引以及陷门中每个元素计算一个添加了随机数和假关键词的认证标签。通过将索引的认证标签与陷门的认证标签相乘即可得到用于验证的证据信息，用户将结果文档的相关度得分与证据信息中的相关度得分进行对比，并且重构搜索函数实现对文档的验证。同时，该方法使用随机挑战技术对检索结果的排序进行验证，降低了排序结果被篡改的可能性。文献[34]同样利用同态 MAC 技术实现针对个人健康信息(Personal Health Information)的可验证密文检索。上述两种方案在 CS 存在惰性计算的情况下仍然存在检测失败的可能性，同时该方案在大型文档库中无法使用，实用性不高。文献[35]结合 HMAC 与基数树提出了一种支持关键词合取(Conjunctive)关系检索和模糊检索的可验证方案。该方法在索引构造阶段为每个关键词以及关键词-密文对计算一个 HMAC 值；在验证阶段，用户首先验证关键词-HMAC 的一致性，再验证关键词-文档 HMAC 的一致性，实现对检索结果的验证。此外，文献[36]通过引入 OXT 协议[37]也实现了具有亚线性复杂度的可验证关键词合取关系密文检索。这两种方法针对关键词合取关系检索都有较高的检索效率和验证效率，但无法满足大规模数据上的可验证检索要求。

4.2　基于多重集哈希函数的可验证多关键词密文检索方法

文献[38]提出了一种使用多重集哈希函数检测数据完整性的方法。多重集哈希函数是一种作用于集合或多重集合的函数，它可以将有限大小的多重集映射为固定长度的哈希值。该方案对内存数据的每次存取操作都添加时间戳以防止重放攻击，同时将存取数据、逻辑地址以及时间戳的三元数组更新至多重集哈希中。该方案通过检查读操作与写操作对应的多重集哈希值是否一致，来验证是否正常工作。基于多重集哈希的工作原理，文献[39]将其应用于针对密文数据的可验证关键词搜索，并实现了一种能够检测 CS 惰性计算的方法，该方法利用一个多重集哈希函数记录每个倒排列表中的所有元素，将其保存并做为本地证据。在发起检索后，CS 使用一个多重集哈希函数记录每个检索过的文档，并将该多级哈希值与搜索结果一同返回给用户。用户通过对比该哈希值与本地证据是否一致即可验证 CS 是否检索了整个搜索列表。该方法同时实现了前向和后向隐私保护。然而，由于在检索以及更新阶段都需要复杂的步骤与计算过程，该方案的效率较低，并且通过强制 CS 搜索整个搜索列表会造成计算资源的浪费。

4.3　基于多重集累加器的可验证多关键词密文检索技术

文献[40]使用多重集累加器实现了针对集值数据(Set-valued Data)的可验证聚合查询，该方法利用双线性映射累加器能够将多重集映射为一个单一值的特性，对数据进行预处理；同时该方法添加了对 CS 不可见的随机值，从而避免相同多重集映射结果也相同所带来的隐私泄露问题，结合 MG-tree(Merkle Grid Tree)根摘要不可变的特性实现对查询结果的验证。类似地，文献[41]基于多重集累加器提出了两种不同的构造方案，实现了可验证的布尔查询，该方法利用多重集累加器可证明集不相交的特性实现验证，同时提出了批量验证的方法，提升了查询性能。在这些方案中，CS 为构造证明而计算多重集累加器造成的效率降低的问题亟待解决，而多重集累加器技术在可验证密文检索领域仍值得深入探索。

此外，在可验证多关键词密文检索领域，除了上述方案之外，现有的研究工作中也存在其他相关解决方案。例如，文献[42]利用电子签名实现了基于属性的可验证关键词密文检索；文献[43]提出基于 Bloom 过滤器索引结构以及多样性均衡模型的可验证密文检索方案；文献[44]提出利用大整数因式分解复杂性特点实现可验证密文检索方案；文献[45]通过引入 B+树索引结构和计数型 Bloom 过滤器，提出支持动态更新和多用户查询的可验证密文检索方案；文献[46]提出基于双线性映射和安全 KNN 机制的 Full Secure 可验证密文检索方案；文献[47]提出一种面向可验证动态密文检索的有效容错解决方案，该方案将可验证密文检索视为黑盒，实现针对恶意威胁模型的有效容错。

5　总结与展望

5.1　总结现有工作

本节通过综合分析和对比现有的代表性研究工作的研究重点及其所使用的关键技术，对现有工作进行总结。

(1)现有的可验证关键词密文检索的研究重点和方向如表 1 所列，各方案在可验证单关键词检索或多关键词可验证检索方面都进行了优化，其中 VFKS，VPSearch 和 VCFE 还实现了对模糊关键词的可验证检索，而 DVSSE，GSSE，VDERS 和 VBS-FB 方案支持动态数据更新的可验证关键词检索，但对重放攻击具有防范能力的方案目前还较少。

表 1　现有的代表性研究工作的研究重点

Table 1　Research priorities of existing representative works

关键词数量	现有研究	排序检索	模糊处理	支持更新	隐私保护	正确性验证	完整性验证	新鲜度验证
单关键词	VFKS[16]	×	√	×	√	√	√	×
	DVSSE[17]	×	×	√	√	√	√	×
	GSSE[24]	×	×	√	√	√	√	√
	VDERS[28]	√	×	√	√	√	√	×
多关键词	VPSearch[33]	√	√	×	√	√	√	×
	VCFE[34]	×	√	×	√	√	√	×
	VBS-FB[36]	×	×	√	√	√	√	×
	PA^2[37]	×	×	×	√	√	√	×

(2)现有可验证关键词密文检索使用的关键技术如表2所列,这些方法都使用对称加密实现了对数据的隐私性保护。在实现一致性验证的技术手段上,Merkle哈希树是使用最广泛的技术,现有研究大多采用Merkle哈希树或基于Merkle哈希树的变型;RSA累加器作为提出较早、理论基础成熟的技术在最近的研究中也得到了应用;目前,消息认证码和位图技术被越来越多地作为辅助技术,用以提高检索性能和验证效果;而多重集哈希等技术应用较少,值得继续深入研究。

表2 现有代表性研究工作使用的关键技术

Table 2 Key techniques used in existing representative works

关键词数量	研究工作	Bitmap	Merkle哈希树	RSA累加器	消息认证码	多重集哈希	对称加密
单关键词	VFKS[16]	√	×	×	×	×	√
	DVSSE[17]	√	×	×	√	×	×
	GSSE[24]	×	√	×	×	×	√
	VDERS[28]	×	×	√	×	×	√
多关键词	VPSearch[33]	×	×	×	√	×	√
	VCFE[34]	√	×	×	√	×	√
	VBS-FB[36]	×	×	×	×	√	√
	PA²[37]	×	√	√	×	×	√

由表1和表2可知,目前可验证密文检索方案的大部分工作虽然基于不同技术,但基本上都是通过重构证据的形式来实现验证机制。其中,对排序检索中排序的准确性验证以及支持动态更新是目前研究的两大难点,构建安全稳定、便于检索的索引,同时实现动态数据更新的可验证关键词排序检索是现有研究工作的关键。

5.2 对未来工作的展望

云计算发展至今,基础的存储与计算服务已经无法满足日益发展的技术需求,同时越来越多的研究和应用开始聚焦云计算的安全问题,除了对数据隐私的保护外,存取数据的准确性和完整性也是需要保证的关键要素。当前,面向云环境的可验证密文检索技术经过多年的研究和发展,已经积累了一些具有实际应用价值的研究方案和技术成果,但在如何支持大规模数据的可验证密文检索、可并行化的可验证密文检索、基于新型索引机制的可验证密文检索、面向不同安全需求的可验证密文检索等方面,仍然存在不少亟待解决的难题。本文重点从以下4个方面阐述对可验证关键词检索研究的未来工作的展望。

(1)支持大规模数据的可验证密文检索方案。目前针对大规模数据的检索问题,已经提出了许多可行的方案,但这些方案并不支持可验证密文检索。同时,现有的可验证密文检索方案大多没有考虑大规模数据应用场景,因此也不适用于解决面向大规模数据的可验证密文检索问题,原因在于这些方案主要采用基于证据重构的方法实现,而构建证据信息的时间和空间消耗往往跟数据的规模呈正相关,这就导致在大规模数据检索应用场景中,在系统初始化时构建初始证据信息以及在检索过程中为检索结果重构证据信息的时间和空间消耗明显增大,直接影响可验证检索的执行效率。因此,面向大规模数据应用场景的兼顾时间和空间效率的可验证密文检索方案是具有现实意义的一个研究方向。

(2)支持并行处理的可验证密文检索方案。随着信息化进程的不断推进,数据的存储和计算量呈几何式增长,依托于云计算的多机协同并行处理技术正是应对这一问题的有效解决方案之一。然而,现有的可验证密文检索技术方案为了能够方便、可靠地获取和重构证据信息,往往会采取集中式证据信息存储和管理模式,这也使得现有的技术方案大多采用集中式处理模式,并不支持多机并行化可验证密文检索。此外,由于多机环境的安全信任、验证策略协同等现实问题的存在,使得现有的技术方案直接应用到并行处理上存在较大难度。因此,基于云计算技术,如何设计和实现支持并行处理的可验证密文检索方案是一个值得研究和探索的问题。

(3)基于新型索引机制的可验证密文检索方案。在现有的各种可验证密文检索方案中,索引是实现高效密文检索和验证的关键。索引结构设计的好坏,一方面直接影响着密文检索的执行效率,另一方面也会对构建与检索结果配套的证据信息产生直接影响。现有的研究方案由于验证技术的限制,在索引的创新上进展缓慢,无法适应技术发展的要求。现有方案中基于树形结构的索引随着数据规模的扩大逐渐失去了自身的优势,而其他如倒排索引等结构在搜索效率上大都无法与树形结构索引相提并论。因此,研究和探索新型索引机制,在支持检索结果一致性验证的基础上,同时满足证据构造简单、检索高效且便于更新等特性要求,是可验证密文检索研究中的一个重要研究课题。

(4)面向不同安全需求的可验证密文检索方案。随着计算机技术越来越多地被应用到各个领域中,对于安全的需求也呈现出多样化的特点,一方面安全需求的内涵具有差异性,如防篡改、防伪造、防懒惰计算、防隐私泄露等;另一方面安全需求的强度也存在着差异性,如国防领域的安全需求强度显然高于一般的企业和公司。现有的可验证密文检索方案针对的安全需求问题往往较为单一,并未考虑现实应用场景中的差异化安全需求。显然,在安全需求差异化的背景下,设计并实现能够同时解决所有安全问题且安全强度可调的可验证密文检索方案是不现实的,也是不经济的。因此,如何能够根据安全需求的多样性特点,研究面向不同安全需求的可验证密文检索方案也是一个值得研究和探索的方向。

参考文献

[1] SUN P. Security and Privacy Protection in Cloud Computing: Discussions and Challenges[J]. Journal of Network and Computer Applications, 2020, 160: 1-22.

[2] LU J, XIAO R, JIN S. A Survey for Cloud Data Security[J].

Journal of Electronics & Information Technology，2021，43(4)：881-891.

[3] PARAST F K，SINDHAY C，NIKAM S，et al. Cloud Computing Security：A Survey of Service-based Models[J]. Computers & Security，2022，114：1-14.

[4] DAI X，DAI H，RONG C，et al. Enhanced Semantic-Aware Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data[J/OL]. IEEE Transactions on Cloud Computing. https：//ieeexplore. ieee. org/document/9310281.

[5] DAI H，YANG M，YANG G，et al. A KGI-index Based Multi-keyword Ranked Search Scheme over Encrypted Cloud Data[J/OL]. IEEE Transactions on Sustainable Computing. https：//ieeexplore. ieee. org/document/9606613.

[6] MTHUNZI S N，BENKHELIFA E，BOSAKOWSKI T，et al. Cloud Computing Security Taxonomy：from an Atomistic to a Holistic View[J]. Future Generation Computer Systems，2020，107：620-644.

[7] ERMAKOVA T，FABIAN B，KORNACKA M，et al. Security and Privacy Requirements for Cloud Computing in Healthcare：Elicitation and Prioritization from a Patient Perspective[J]. ACM Transactions on Management Information Systems，2020，11(2)：1-29.

[8] TIAN H L，ZHANG Y，LI C，et al. A Survey of Confidentiality Protection for Cloud Databases[J]. Chinese Journal of Computers，2017，40(10)：2245-2270.

[9] ALJUMAH A，AHANGER T A. Cyber Security Threats，Challenges and Defense Mechanisms in Cloud Computing[J]. IET Communications，2020，14(7)：1185-1191.

[10] XIA Y，XIA F，LIU X，et al. An Improved Privacy Preserving Construction for Data Integrity Verification in Cloud Storage[J]. KSII Transactions on Internet and Information Systems，2014，8(10)：3607-3623.

[11] YU X，YAN Z，VASILAKOS A V. A Survey of Verifiable Computation[J]. Mobile Networks and Applications，2017，22(3)：438-453.

[12] XU Z，WU L，HE D，et al. Security Analysis of a Publicly Verifiable Data Possession Scheme for Remote Storage[J]. The Journal of Supercomputing，2017，73(11)：4923-4930.

[13] JIANG X，GE X，YU J，et al. An Efficient Symmetric Searchable Encryption Scheme for Cloud Storage[J]. Journal of Internet Services and Security，2017，7(2)：1-18.

[14] CHAI Q，GONG G. Verifiable Symmetric Searchable Encryption for Semi-honest-but-curious Cloud Servers[C]// 2012 IEEE International Conference on Communications(ICC 2012). New York：IEEE Press，2012：917-922.

[15] WANG J，MA H，TANG Q，et al. Efficient Verifiable Fuzzy Keyword Search over Encrypted Data in Cloud Computing[J]. Computer Science & Information Systems，2013，10(2)：667-684.

[16] LI J，WANG Q，WANG C，et al. Fuzzy Keyword Search over Encrypted Data in Cloud Computing[C]// Proceedings of the 29th Conference on Information Communications(INFOCOM 2010). New York：IEEE Press，2010：441-445.

[17] RAMASAMY R，VIVEK S S，GEORGE P，et al. Dynamic Verifiable Encrypted Keyword Search using Bitmap Index and Homomorphic MAC[C]// 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing(CSCloud 2017). New York：IEEE Press，2017：357-362.

[18] MERKLE R C. A Certified Digital Signature[C]// Conference on the Theory and Application of Cryptology (CRYPTO 1989). Berlin：Springer，1989：218-238.

[19] PANG H H，MOURATIDIS K. Authenticating the Query Results of Text Search Engines[J]. Proceedings of the VLDB Endowment，2008，1(1)：126-137.

[20] HU H，XU J，CHEN Q，et al. Authenticating Location-based Services without Compromising Location Privacy[C]// Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data(SIGMOD 2012). NewYork：ACM Press，2012：301-312.

[21] MOURATIDIS K，SACHARIDIS D，PANG H H. Partially Materialized Digest Scheme：an EfficientVerification Method for Outsourced Databases[J]. The VLDB Journal，2009，18(1)：363-381.

[22] WOOD G. Ethereum：A Secur e Decentralised Generalised Transaction Ledger[R/OL]. https：//files. gitter. im/ethereum/yellowpaper/VIyt/Paper. pdf.

[23] MATHIYALAHAN S，MANIVANNAN S，NAGASUNDARAM M，et al. Data Integrity Verification using MPT(Merkle Patricia Tree) in Cloud Computing[J]. International Journal of Engineering & Technology，2018，7(2)：500-503.

[24] ZHU J，LI Q，WANG C，et al. Enabling Generic，Verifiable，and Secure Data Search in Cloud Services[J]. IEEE Transactions on Parallel and Distributed Systems，2018，29(8)：1721-1735.

[25] PANG H H，JAIN A，RAMAMRITHAM K，et al. Verifying Completeness of Relational Query Results in Data Publishing[C]// Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data(SIGMOD 2005). NewYork：ACM Press，2005：407-418.

[26] DAI H，BAO J J，ZHU X Y，et al. Integrity-verifying Single Keyword Search Method in Clouds[J]. Computer Science，2018，45(12)：92-97.

[27] BARIC N，PFITZMANN B. Collision-free Accumulators and Fail-stop Signature Schemes without Trees[C]// International Conference on the Theory and Applications of Cryptographic Techniques(EUROCRYPT 1997). Berlin：Springer，1997：480-494.

[28] LIU Q，NIE X，LIU X，et al. Verifiable Ranked Search over Dynamic Encrypted Data in Cloud Computing[C]// 2017 IEEE/ACM 25th International Symposium on Quality of Service (IWQoS 2017). NewYork：IEEE Press，2017：1-6.

[29] LIU Q，TIAN Y，WU J，et al. Enabling Verifiable and Dynamic Ranked Search over Outsourced Data[J]. IEEE Transactions on Services Computing，2022，15(1)：69-82.

[30] CAMENISCH J，LYSYANSKAYA A. Dynamic Accumulators and Application to Efficient Revocation ofAnonymous Credentials[C]// Annual International Cryptology Conference(CRYPTO 2002). Berlin：Springer，2002：61-76.

[31] GOODRICH M，TAMASSIA R，TELALOVIC J H. An Effi-

cient Dynamic and Distributed RSA Accumulator[J]. arXiv: 0905.1307, 2009.

[32] ZHU X, LIU Q, WANG G. A Novel Verifiable and Dynamic Fuzzy Keyword Search Scheme over Encrypted Data in Cloud Computing[C]// 2016 IEEE Trustcom/BigDataSE/ISPA. New-York: IEEE Press, 2016: 845-851.

[33] WAN Z, DENG R H. VPSearch: Achieving Verifiability for Privacy-preserving Multi-keyword Search over Encrypted Cloud Data[J]. IEEE Transactions on Dependable and Secure Computing, 2016, 15(6): 1083-1095.

[34] LU H, CHEN J, ZHANG K. Verifiable Dynamic Searchable Symmetric Encryption with Forward Privacy in Cloud-Assisted E-Healthcare Systems[C]// 21st International Conference Algorithms and Architectures for Parallel Processing (ICA3PP 2021). Berlin: Springer, 2021: 645-659.

[35] SHAO J, LU R, GUAN Y, et al. Achieve Efficient and Verifiable Conjunctive and Fuzzy Queries over Encrypted Data in Cloud [J]. IEEE Transactions on Services Computing, 2022, 15(1): 124-137.

[36] GAN Q, LIU J, WANG X, et al. Verifiable Searchable Symmetric Encryption for Conjunctive Keyword Queries in Cloud Storage[J]. Frontiers of Computer Science, 2022, 16(6): 1-16.

[37] CASH D, JARECKI S, JUTLA C, et al. Highly-scalable searchable symmetric encryption with support for Boolean queries [C]// Proceedings of the 33rd Annual Cryptology Conference. Berlin: Springer, 2013: 353-373.

[38] CLARKE D, DEVADAS S, DIJK M, et al. Incremental Multiset Hash Functions and Their Application to Memory Integrity Checking[C]// International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT 2003). Berlin: Springer, 2003: 188-207.

[39] LI F, MA J, MIAO Y, et al. Towards Efficient Verifiable Boolean Search over Encrypted Cloud Data[J/OL]. IEEE Transactions on Cloud Computing. https://ieeexplore.ieee.org/document/9565340.

[40] XU C, CHEN Q, HU H, et al. Authenticating Aggregate Queries over Set-valued Data with Confidentiality[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 30(4): 630-644.

[41] XU C, ZHANG C, XU J. vChain: Enabling Verifiable Boolean Range Queries over Blockchain Databases[C]// Proceedings of the 2019 International Conference on Management of Data (SIGMOD 2019). NewYork: ACM Press, 2019: 141-158.

[42] ZHENG Q, XU S, ATENIESE G. VABKS: Verifiable Attribute-based Keyword Search over Outsourced Encrypted Data[C]// Proceeding of the 2014 IEEE Conference on Computer Communications (INFOCOM 2014). NewYork: IEEE, 2014: 522-530.

[43] LIU Y, PENG H, WANG J. Verifiable Diversity Ranking Search over Encrypted Outsourced Data[J]. Computers, Materials and Continua, 2018, 55(1): 37-57.

[44] ZHAO M, LIU L, DING Y, et al. Verifiable and Privacy-Preserving Ranked Multi-Keyword Search over Outsourced Data in Clouds[C]// 2021 IEEE 15th International Conference on Big Data Science and Engineering (BigDataSE 2021). New York: IEEE, 2021: 95-102.

[45] SHI Z, FU X, LI X, et al. ESVSSE: Enabling Efficient, Secure, Verifiable Searchable Symmetric Encryption[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(7): 3241-3254.

[46] NAJAFI A, JAVADI H, BAYAT M. Efficient and dynamic verifiable multi-keyword searchable symmetric encryption with full security[J]. Multimedia Tools and Applications, 2021, 80(17): 26049-26068.

[47] YUAN D, CUI S, RUSSELLO G. We Can Make Mistakes: Fault-tolerant Forward Private Verifiable Dynamic Searchable Symmetric Encryption[C]// 7th IEEE European Symposium on Security and Privacy (EuroS&P 2022). New York: IEEE, 2022: 587-605.

ZHOU Qian, born in 1983, Ph. D, lecture, master supervisor, is a member of China Computer Federation. Her main research interests include information security and privacy protection.

DAI Hua, born in 1982, Ph. D, professor, Ph. D supervisor, is a member of China Computer Federation. His main research interests include cloud computing security and privacy protection.

(责任编辑:何杨)