面向云环境的多关键词密文排序检索研究综述

戴 华[1,2] 李 啸[1] 朱向洋[1] 杨 庚[1,2] 易 训[3]
(南京邮电大学计算机学院 南京 210023)[1] (江苏省大数据安全与智能处理重点实验室 南京 210023)[2]
(墨尔本皇家理工大学科学学院 墨尔本 3000)[3]

摘 要 随着云计算的广泛应用,面向数据或计算的外包服务模式越来越被业界所接受。为了保护数据拥有者外包数据的私密性,具备隐私保护能力的高效密文排序检索技术逐渐成为目前备受关注的一个研究热点。文中以面向云环境的多关键词密文排序检索技术为关注重点,介绍了现有研究工作的系统模型和威胁模型,并描述了模型中关于隐私保护、检索效率与准确率、检索结果完整性等的问题;全面分析了现有工作中典型的多关键词密文排序检索方法及相关扩展研究,讨论并梳理了这些方法的核心思想;最后,对现有研究工作进行了总结,并给出了该研究领域中待解决的关键性问题和未来的研究方向。

关键词 云外包,数据隐私,多关键词检索,密文检索,排序检索
中图法分类号 TP309.7 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.01.002

Research on Multi-keyword Ranked Search over Encrypted Cloud Data

DAI Hua[1,2] LI Xiao[1] ZHU Xiang-yang[1] YANG Geng[1,2] YI Xun[3]
(College of Computer Science & Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)[1]
(Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing 210023, China)[2]
(School of Science, Royal Melbourne Institute of Technology University, Melbourne 3000, Australia)[3]

Abstract With the extensive development of cloud computing, storage and/or computing outsourcing services are becoming more and more acceptable nowadays. To protect the privacy of outsourced data, the privacy-preserving multi-keyword ranked search scheme over encrypted cloud data is focused by researches, which turns to be a hot spot recently. This paper introduced the system model and threat model of existing work, and gave the problem description about privacy-preserving, search efficiency and accuracy, search result completeness, etc. Typical works and extended research about multi-keyword ranked search were studied, and the main ideas of those methods were discussed in detail. Finally, the conclusions about current works were given, and the future research directions were proposed simultaneously.

Keywords Outsourced cloud, Data privacy, Multi-keyword search, Encrypted data search, Ranked search

1 引言

IT资源服务化的思想日益普及,呈现“一切皆服务”的趋势,“服务”成为云计算的核心概念。然而在云计算蓬勃发展的同时,云安全也成为备受关注的问题[1-5]。在云环境中,用户由于无法直接控制放置在远程云服务器(Cloud Server,CS)中的数据,会担心自己的外包数据被云服务提供商(Cloud Server Provider,CSP)非法获取或滥用,尤其是对于私密性要求较高的敏感数据,例如电子病历[6-8]、银行交易数据、用户邮件等。尽管CSP宣称会采取适当措施(例如防火墙、访问控制和入侵检测等)来确保数据的安全和隐私,但是这种“尽力而为”的服务协议承诺很难消除用户的疑虑。除此之外,CSP内部员工职业操守带来的隐私泄露问题更是难以防范[9],例如瑞士银行泄露部分客户的资料,谷歌员工偷窥用户的Gmail信息[10]等。CSP内外部带来的数据安全问题在很大程度上阻碍了云计算的进一步发展[11-12]。

实现云环境中隐私保护最直接的方法是将数据先加密处理后再外包到云服务器。但是数据加密后带来了严重的数据可用性问题,例如在信息检索领域,现有的多关键词检索主要是面向明文数据的,无法直接应用到密文排序检索场景中。而将所有加密数据从云端下载到本地进行解密显然是一种不切实际且浪费资源的处理方法。因此,研究可搜索加密,寻找

到稿日期:2018-02-28 返修日期:2018-05-17 本文受国家自然科学基金项目(61402014,61572263,61672297,61472193),江苏省自然科学项目(BK20151511,BK20161516),中国博士后科学基金项目(2015M581794),安徽省自然科学基金项目(1608085MF127),江苏省博士后科研资助计划(1501023C),南京邮电大学自然科学基金项目(NY217119)资助。

戴 华(1982-),男,博士,副教授,主要研究方向为数据管理与安全、数据库技术,E-mail:daihua@njupt.edu.cn(通信作者);李 啸(1994-),男,硕士生,主要研究方向为隐私保护、密文检索;朱向洋(1993-),男,硕士生,主要研究方向为隐私保护、密文检索;杨 庚(1961-),男,教授,博士生导师,主要研究方向为大数据安全、隐私保护;易 训(1967-),男,教授,博士生导师,主要研究方向为密码学、信息安全。

一种新的密码学原语和协议,以确保用户数据的机密性和加密数据的可检索性,已经成为近几年的研究热点。

为了在保护数据隐私的同时确保密文数据的可用性,研究者采用可搜索加密技术(Searchable Encryption,SE)实现基于关键词的密文数据检索,包括对称可搜索加密(Symmetric Searchable Encryption,SSE)与非对称可搜索加密(Asymmetric Searchable Encryption,ASE)[13-14]。由于非对称可搜索加密难以应对密文的排序问题,现有的针对多关键词密文排序检索的研究主要采用对称可搜索加密机制,其核心问题是安全可搜索索引的数据结构、构建算法和检索算法。

本文针对基于对称可搜索加密的多关键词密文排序检索技术,从典型的多关键词密文排序检索技术及其扩展研究这两个角度出发,对现有工作进行综述研究,总结现有工作的核心思想,讨论各项工作的优缺点,并在此基础上分析未来研究的可能发展方向。基于本文的研究和分析,在当前以及未来的基于多关键词的密文排序检索研究工作中,如何构建结构简单、检索高效、更新方便且自身安全的索引机制是核心关键,而集群并行化处理检索请求以及检索结果的完整性验证方法是后续研究亟待解决的重要问题。

本文第 2 节为相关模型介绍以及问题描述;第 3 节针对典型的对称可搜索加密技术,从索引优化出发进行分析和总结;第 4 节介绍多关键词密文排序检索的扩展研究;最后对现有研究工作进行分析,并对未来可能的研究方向进行展望。

2　模型与问题描述

本节将从系统模型、威胁模型和问题描述这 3 个方面来阐述基于对称可搜索加密机制的多关键词密文排序检索技术的研究和应用背景。

2.1　系统模型

面向云环境的多关键词密文排序检索系统模型如图 1 所示,主要包含 3 个不同实体:数据拥有者(Data Owner,DO)、数据使用者(Data User,DU)和公有云服务器(Cloud Service,CS)。它们的协作方式如下:

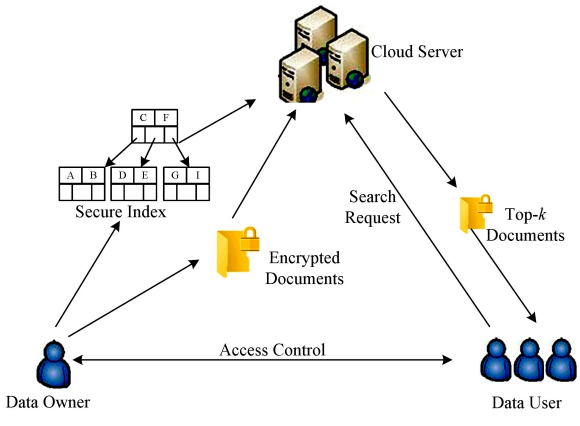

图 1　面向云环境的多关键词密文排序检索系统结构

Fig. 1　System architecture of multi-keyword ranked search over encryptedcloud data

(1)DO 负责原始文档的处理,将原始文档集合加密成密文并上传到公有云服务器 CS。为了实现针对密文文档的可检索,需要根据原始文档集合构造安全索引,并与密文文档一起外包存储至 CS 中。

(2)CS 接收并存储 DO 发来的数据;CS 在接收到 DU 的检索陷门指令后,根据安全索引执行检索处理,并返回符合检索要求的检索结果给 DU。

(3)DU 是指经过 DO 授权的用户,能够检索 CS 中存储的 DO 的数据;在执行检索请求时,DU 提交检索陷门指令至 CS,并等待检索反馈。

2.2　威胁模型

在面向云环境的多关键词密文排序检索技术的研究中,负责存储和处理外包数据的 CS 所面临的威胁模型主要有两类。

(1)诚实而好奇(Honest-but-Curious)模型

在诚实而好奇模型的假设下,CS 能够严格按照约定的协议存储外包数据并执行数据服务,但是出于好奇,CS 可能对外包数据实施窥探和分析,非法获取 DO 的敏感数据,进而造成数据隐私的泄露。该模型在部分研究工作中又称作“半可信”(Semi-Honest)威胁模型。

(2)恶意攻击(Malicious Attack)模型

在恶意攻击威胁模型的假设下,CS 对于 DO 完全不可信,CS 可能会对存储其中的外包数据或其提供的数据服务实施主动攻击(如数据篡改、数据伪造、数据丢弃等),这些攻击行为可能由 CS 内部管理员的权限滥用、黑客攻击等造成,从而导致 DU 获取的检索结果可能不满足完整性要求;同时,也有可能造成用户数据隐私的泄露。

现有的绝大多数多关键词密文排序检索研究方案普遍采用“诚实而好奇”模型。虽然在实际应用场景中针对外包数据存储和服务的恶意攻击行为确实有可能存在,但由于针对恶意攻击模型的防范技术难度大,目前完备的针对恶意攻击模型的多关键词密文排序检索方案尚未见报道。基于此,本文重点讨论现有的基于“诚实而好奇”威胁模型的多关键词密文排序检索技术。

2.3　问题描述

在上述威胁模型的基础上,基于对称可搜索加密的多关键词密文排序检索机制主要解决以下几方面问题。

(1)多关键词排序检索(Multi-keyword Ranked Search):CS 根据检索陷门,基于相似度度量指标,返回最相关的 k 个文件(Instead of Unrelated Files)。

(2)隐私保护(Privacy-preserving):该机制在运行过程中不能泄露文件机密性、索引机密性、检索陷门机密性、检索陷门关联关系机密性以及关键词隐私。

(3)检索效率和准确率(Efficiency and Accuracy):要在保证准确度的前提下尽量提高检索的效率,不能以降低准确度为代价来提高检索效率。

(4)完整性验证:对于 CS 返回的 Top-k 文件,其 Top-k 完整性能被检测,从而保证返回的文件是数据拥有者上传的文件,没有被篡改、伪造等。

3　典型的多关键词密文排序检索技术

现有的多关键词密文排序检索方法主要采用“诚实而好

奇”威胁模型，利用安全 KNN 技术，通过构建支持密文检索的索引实现多关键词密文排序检索。其中，索引是各方案的关键，直接影响检索的效率和效果。目前，较为典型的索引结构有线性索引、平衡二叉树索引、多维 B 树索引、关键词平衡二叉树索引、层次聚类树索引等。本节从索引和检索方法的角度介绍现有的典型多关键词密文排序检索技术。

3.1 基于线性索引的检索方法

文献[15]结合安全 KNN[16]实现了加密检索向量与加密文档向量之间的内积运算，首次给出了面向云环境的多关键词密文排序检索方法 MRSE。该方法将每一个文档进行 0/1 向量化处理，对于任意两个 0/1 向量文档，若两个向量中相同位取值为 1 的位数越多，则这两个文档的相关度就越高；通过引入安全 KNN 对文档向量及检索向量进行加密，从而将文档与检索多关键词之间的相关度计算问题转化成文档向量与检索向量的内积计算问题，通过计算相关度得分来确定检索结果。由于在检索时文档向量和检索向量需要逐一进行内积计算从而确定相关度，因此以该方法构建的索引被称为线性索引。

由于关键词词频在一定程度上表示了该关键词对文档的重要程度，文档检索应当考虑关键词的重要程度，因此 MRSE 提出的使用 0/1 向量标识关键词在文档中的存在性的检索方法在实际应用中有一定的限制。基于此，文献[17]对文档向量的表示方法进行了优化，针对关键词词频对文档的重要程度，引入 TF-IDF 与向量空间模型(VSM)，从而将文档和检索多关键词之间的相关度度量问题转化成两个向量之间的内积计算问题，提高了检索的准确度，但该方法本质上还是线性索引，其检索效率与 MRSE 相比并未提高。

3.2 基于平衡二叉树索引的检索方法

文献[18]使用平衡二叉树(Balanced Binary Tree，BB-Tree)和安全 KNN 实现多关键词密文排序检索 MKRS。该方法采用与文献[17]相同的 TF-IDF 和空间向量模型(VSM)，对文档和检索关键词进行归一化向量处理，利用向量之间的余弦距离表示文档和检索多关键词之间的相关度，并利用安全 KNN 对各向量进行加密处理，实现向量自身的隐私保护。为了提高检索效率，MKRS 将每个文档作为一个叶子节点，自下而上构建平衡二叉树 BB-Tree，叶子节点存储文档向量，中间节点存储辅助标识向量，每位的取值为 0/1，若以该中间节点为根的子树中所有叶子节点对应的文档均不包含关键词，则该位置取 0，否则取 1。在检索时，利用中间节点的辅助标识向量排除不包含检索多关键词的子树，从而提高文档的检索效率。但由于 BB-Tree 树在检索时，中间节点需要频繁加解密，且这种 0/1 的取值方式只能表示中间节点的该位对应的关键词是否存在，并不能反映关键词的词频，因此在检索时会造成大量的时间开销。

3.3 基于多维 B 树索引的检索方法

文献[19]利用 MDB-Tree 和安全 KNN 实现了支持隐私保护的多关键词密文排序检索技术 MTS。该方法将文档向量划分成多个有序的子向量，并将该子向量作为 MDB-Tree 中的独立属性值进行存储。检索向量的划分方法与文档向量相同，则文档向量和检索向量的相关度即为对应子向量点积的累加和；通过引入安全 KNN，即可实现密文状态下文档向量和检索向量之间的相关度计算。在检索时，采用深度优先的思想从根节点开始，每向下一层，选择未使用过的最大值进行检索，从而预测可能的最大相关度阈值(即预测门限值)。如果预测的最大相关度小于已经获取的候选检索结果中的最小相关度，则退回父节点；否则将继续检索进入下一层的孩子节点；检索过程递归执行，直到确定了排序检索结果。该方法通过预测门限值避免了检索时遍历整棵树，只需要访问部分节点就可以获得检索结果，从而提高检索效率。在 MTS 方案中，预测门限值是影响检索准确率和效率的重要参数，在检索过程中该阈值是动态变化的，它的确定是一个难点问题。

3.4 基于关键词平衡二叉树索引的检索方法

文献[20]提出利用关键词平衡二叉树(Keyword Balanced Binary Tree，KBB-Tree)索引结构实现安全动态的多关键词密文排序检索方法 DMRS。该方法同样采用 TF-IDF 和空间向量模型对文档和检索关键词进行向量化处理，通过安全 KNN 实现向量加密，并在此基础上提出基于 KBB-Tree 的贪婪深度优先搜索算法。该方法将每个文档作为一个叶子节点，自下而上构建 KBB-Tree，其中叶子节点存储文档向量，中间节点存储检索筛选向量，该筛选向量由其孩子节点向量对应位的最大值构成。在检索过程中，利用中间节点的筛选向量排除不可能包含检索结果文档的子树，从而有效地缩小搜索范围，提高文档的检索效率。但由于 KBB-Tree 在构建时将代表文档的叶子节点随机放置，并未考虑文档之间的内在关系，其检索效率依然存在一定的提升空间。

3.5 基于聚类树索引的检索方法

在信息检索(Information Retrieval，IR)[21]领域，若考虑文档之间的相关性，则可以显著提高检索效率。基于该思想，文献[22]提出基于层次聚类的多关键词密文排序检索方法 MRSE-HCI。该方法通过引入基于二分 k-means 的层次聚类方法自顶向下构建层次聚类索引树(Hierarchical Clustering Index Tree，HCI-Tree)：初始时将整个文档集合作为一个簇，利用二分 k-means 进行自顶向下的簇划分，直至划分出来的子簇中包含的文档数量恰好不超过给定阈值(称为最小分割子簇)。划分的过程即为 HCI-Tree 的构建过程，树中各节点代表一个聚类簇，其中存储着该聚类簇的簇中心向量和文档数量。叶子节点与该簇中的文档直接关联，文档向量同样采用 TF-IDF 和向量空间模型表示，安全 KNN 用于保护各向量的私密性。在检索时，根节点开始选取其孩子节点中的簇中心向量和检索向量相关度得分较大的子簇进行深度优先搜索，直到某个节点的任一子簇中的文档数量均小于检索目标数量时，该节点所代表的簇中的文档集合即为检索结果。MRSE-HCI 虽然检索效率较高，但可能存在一定的检索误差。

为了解决 MRSE-HCI 方法中的检索误差问题，文献[23]提出基于凝聚层次聚类方法实现精确的多关键词密文排序检索方法 MUSE。该方法采用自底向上的凝聚层次聚类方法

构建凝聚层次聚类索引树(Hierarchical Agglomerative Clustering Index Tree,HAC-Tree),其中叶子节点代表只包含一个文档的簇,并存储文档向量;中间节点代表以该节点为根的子树的叶子节点所构成的簇,并存储检索过滤向量,该过滤向量由其孩子节点向量对应位的最大值组成。在检索时,利用非候选剪枝深度优先算法(Noncandidate Pruning Depth-first Search Algorithm)排除确定不包含检索结果的子树,以在确保检索精度的同时提高检索效率。

4　多关键词密文排序检索的扩展研究

在多关键词密文排序检索领域,还有一系列重要研究成果扩展了传统的关键词密文排序检索,主要包含模糊多关键词密文排序检索、支持语义扩展的多关键词密文排序检索、支持个性化的多关键词密文排序检索和可验证的多关键词密文排序检索等,本节将对这些扩展工作进行综述和总结。

4.1　模糊多关键词密文排序检索

针对模糊检索问题,文献[24]提出面向云环境中加密数据的基于编辑距离(Edit Distance)的模糊检索算法,该方法利用通配符实现模糊关键词集合压缩,但由于采用的是基于枚举的检索算法,其检索效率不高。文献[25]在利用通配符和编辑距离计算索引的基础上,提出对索引关键词对应的符号集合进行"预定义",从而构造基于密文符号的语义树索引结构,树形索引结构提高了检索效率,但是语义树的检索效率极大地依赖于索引中关键词的重复率,因此算法对模糊关键词检索效率的提升空间有限,且这种模式很难抵抗字典攻击。文献[26]提出融合高编码效率的 Huffman 编码和具有高压缩存储优势的 Bloom Filter 编码,实现面向 DaaS 模式下支持隐私保护的模糊关键词的查询处理,但该方案的关键词密文匹配采用基于编辑距离的预定义集合,可扩展性较差,且检索效率不高。文献[27]提出基于 Simhash 的模糊关键词排序密文搜索方案,该方案将关键词进行 n-gram 分割[28],采用双因子排序,并利用倒排索引进行检索,提高了检索效率。

然而,上述方法都是针对单关键词的模糊检索,并不能直接应用于多关键词模糊检索场景。基于此,文献[29]首次解决了支持隐私保护的模糊多关键词密文排序检索问题(MFSE),该方案采用局部敏感哈希(LSH)技术,并基于 Bigram 方法[30]将各文档中的关键词映射到 Bloom Filter 编码中,通过索引向量和检索向量的安全 KNN 计算,实现了高效的模糊多关键词密文排序检索,并证明了该方案的安全性。文献[31]基于 MFSE 方案,针对关键词可能出现拼写错误等情况,提出比 n-gram 更有效的分割方案 uni-gram,并考虑了关键词的权重,提高了检索的准确率。文献[32]基于对偶编码函数、Bloom Filter 和距离可恢复加密算法,实现了面向多关键词的模糊密文搜索方法。此外,针对模糊多关键词检索的研究工作还有文献[33-35]。

4.2　支持语义扩展的多关键词密文排序检索

在实际应用中存在大量的同义词、上/下义词等现象,用户每次检索时不可能提供所有的同义词。例如,当用户检索"search"时,不考虑语义模糊的搜索方案只会返回包含"search"的文档,但是实际上包含"seek""find""retrieve"等的文档也可能符合检索要求,因此考虑语义模糊,实现支持语义扩展的多关键词密文排序检索方案同样具有现实意义。基于此,文献[36]对文档关键词进行同义词扩展,通过安全 KNN 技术,实现了支持同义词检索的多关键词密文排序搜索方案。文献[37]为文档集创建倒排索引,利用语义库扩展检索使用的关键词,基于倒排索引和保序加密实现了支持语义扩展的多关键词密文排序检索方案。文献[38]基于概念层次提出支持语义扩展的多关键词密文排序检索方案,根据文档集中域内的概念关系,构建层次概念树,对每个文档都构造两个向量,一个用于与检索请求进行概念匹配,另一个用于验证文档是否满足检索请求。文献[39]将概念图作为一种知识表示来代替传统的关键词,提出基于概念图的支持隐私保护的智能语义检索方案。近些年,一系列工作围绕支持语义扩展的多关键词密文排序检索展开[40-41]。

4.3　支持个性化的多关键词密文排序检索

现有研究工作中多关键词密文排序检索方案大多采用的是"普适"模型,并未考虑不同用户可能具有不同的兴趣爱好、文化背景等个性化特点;同时,不同用户对不同关键词感兴趣的程度可能也不尽相同。现有的基于"普适"模型的检索方法所得到的检索结果可能并不符合用户的独特需求。针对上述问题,文献[42]通过引入用户偏好,提出了个性化的关键词排序检索,并提供了一个规范的模型来整合优先排序和关键词检索;用户偏好的引入有助于实现在检索请求相同的情况下,不同的用户依然可以获得与其偏好和背景相匹配的不同检索结果。文献[43]提出了在密文状态下支持个性化的多关键词密文排序检索方案,通过在数据使用者端建立用户兴趣模型,在支持语义扩展的基础上,利用 WordNet 和用户检索历史分析用户的兴趣爱好,建立用户兴趣模型,提高用户对检索结果的满意度。目前支持个性化兴趣扩展的多关键词密文排序检索技术的研究工作还相对较少。

4.4　支持可验证的多关键词密文排序检索

在云环境中,文档检索除了需要考虑机密性之外,若云服务提供商不可信甚至可能破坏数据,此时即使用户获得了检索结果,检索结果的完整性也仍有待验证。针对多关键词密文检索可验证问题,文献[44]提出了一种可验证外包计算方法,该方法在确保数据隐私的情况下,将检索结果返回给用户,同时还返回一些证据信息,当用户收到检索结果及证据信息时,用户即可根据证据信息验证收到的检索结果是否准确无误,从而使得检索方法具备抗伪造和篡改的能力。针对多关键词密文排序检索的可验证问题,文献[45]提出了基于 Merkle 哈希树的可验证隐私保护的多关键词密文排序检索方案,该方案在保证隐私的同时,支持对检索结果正确性和完整性的验证。文献[22]在提出基于层次聚类方法实现多关键词密文排序检索的同时,还提出了利用哈希子树结构来实现对检索结果的完整性验证。然而,文献[22,44-45]中的检索结果验证方法都是建立在假设云服务提供商满足"诚实而好

奇”模型的基础上的，若云服务提供商完全不可信，甚至可能采取主动攻击来篡改或破坏检索结果数据时，现有的这些方法将无法进行验证；同时，对于云服务提供商为降低计算开销而采取的仅执行局部或抽样检索等“降级”服务行为，现有的方法也无法检测和发现。针对云服务端可能存在的不完全服务行为，文献[46]提出了基于同态 MAC 和随机挑战技术的可验证隐私保护多关键词密文排序检索技术(VPSearch)，该方法能够有效地检测云服务端是否发生了“降级”服务行为。

当前，如何对云服务端的主动攻击行为所造成的检索结果不正确或者不完整进行验证依然是一个难点，特别是在多关键词密文排序检索应用场景中，目前尚未出现较好的解决方案。

5 总结与展望

5.1 对现有工作的总结

本文主要介绍了多关键词密文排序检索及其相关扩展研究的现状，下面从现有典型研究工作所采用的核心基础技术和研究重点这两个角度进行总结。

(1)现有的典型多关键词密文排序检索方法中所采用的核心基础技术的分类情况如表 1 所列。由表 1 可知：现有的多关键词密文排序检索方法均采用安全 KNN 和对称加密技术实现针对文档、索引和检索关键词的隐私保护，提高检索准确率的方法基本是采用 TF-IDF 技术和空间向量模型 VSM，只有 MFSE 选择 Bloom Filter 来实现密文排序检索。

表 1 现有典型研究工作所采用的核心基础技术

Table 1 Key technologies adopted in existing typical works

研究工作	Bloom Filter	VSM	TF-IDF	安全 KNN	对称加密
MRSE[15]	No	Yes	No	Yes	Yes
DMRS[20]	No	Yes	Yes	Yes	Yes
MRSE-HCI[22]	No	Yes	Yes	Yes	Yes
MFSE[29]	Yes	No	Yes	Yes	Yes
PRSE[43]	No	Yes	Yes	Yes	Yes
VMTS[45]	No	Yes	Yes	Yes	Yes
VPSearch[46]	No	Yes	No	Yes	Yes

(2)现有的典型多关键词密文排序检索方法的研究重点分布情况如表 2 所列。由表 2 可知：MRSE，DMRS，VMTS 和 MRSE-HCI 从多角度优化多关键词排序检索的检索效率，同时 VMTS 和 MRSE-HCI 还给出了基于各自使用的树形索引结构的检索结果验证方法。MFSE 解决了多关键词密文排序检索中的模糊检索问题，PRSE 解决了多关键词密文排序检索中的个性化检索问题，VPSearch 解决了多关键词密文排序检索过程中云服务提供“降级”服务的可验证问题。

综合表 1 和表 2 可知，现有的多关键词密文检索方法多使用空间向量模型(布隆过滤器也是一种特殊的向量)、TF-IDF 技术和安全 KNN 技术，由于这些研究工作都需要使用安全索引机制来实现安全且高效的多关键词密文排序检索，因此，如何构建结构简单、检索高效、更新方便且自身安全的索引机制是现有研究工作的关键。

表 2 现有典型研究工作的研究重点

Table 2 Research focuses of existing typical research work

研究工作	多关键词检索	模糊处理	语义扩展	个性化	完整性
MRSE[15]	Yes	No	No	No	No
DMRS[20]	Yes	No	No	No	No
MRSE-HCI[22]	Yes	No	No	No	Yes
MFSE[29]	Yes	Yes	No	No	No
PRSE[43]	Yes	No	Yes	Yes	No
VMTS[45]	Yes	No	No	No	Yes
VPSearch[46]	Yes	No	No	No	Yes

5.2 对未来工作的展望

在云服务快速发展的今天，外包数据存储和计算引起了越来越广泛的关注，云计算快速发展带来的云端数据安全问题成为亟待解决的关键问题。针对多关键词密文排序检索技术的研究和应用仍然处于起步阶段，如何保护数据隐私、提高检索效率和准确率、保护检索结果的完整性等问题仍有待进一步研究。

(1)全同态加密技术在多关键词密文排序检索中的应用

目前，在云环境下的多关键词密文排序检索技术仍然以安全 KNN 加密方法为主，然而将每个文档表示为空间向量模型中的点，时空开销较大；同时，为了提高检索效率而建立索引树的过程会产生大量中间节点，进一步加剧了时空开销，因此基于安全 KNN 技术实现的多关键词密文排序检索方案对时空开销的要求都较高。针对这一问题，将全同态加密技术和倒排索引技术相结合，设计并实现类似于明文多关键词检索的解决方案，是一条可行的研究思路。

(2)集群并行化处理检索请求

在现有研究工作中，多数研究者都是在单机环境中进行多关键词排序检索技术的研究，随着数据量的爆发式增长，数据拥有者不可避免地将采用多级协同的集群提供数据服务；并且在云计算环境中，支持多关键词并行排序检索也是一种潜在的需求。然而，树形索引机制虽然能够在一定程度上提高检索效率，但也限制了检索过程的并行化处理。因此，研究适应集群环境的可并行的多关键词密文排序检索方法具有重要的现实意义。

(3)检索结果的完整性验证问题

当前研究者的工作焦点在于云环境中数据的隐私保护，而针对检索结果的完整性验证也是云安全的一个重点问题。虽然现有工作提出了一些验证多关键词密文排序检索结果完整性的研究方法，但这些方案并不能解决云服务器由管理员内部恶意篡改或伪造攻击所造成的检索结果不可信和不可用问题。因此，针对内部攻击行为的完整性可验证的多关键词密文排序检索技术，同样具有重要的研究和应用价值。

参 考 文 献

[1] LI H，SUN W H，LI F H，et al. Secure and Privacy-Preserving Data Storage Service in Public Cloud[J]. Journal of Computer Research and Development，2014，51(7)：1397-1409. (in Chinese)

李晖，孙文海，李凤华，等. 公共云存储服务数据安全及隐私保护

技术综述[J]. 计算机研究与发展,2014,51(7):1397-1409.

[2] ZHANG Y Q, WANG X F, LIU X F, et al. Survey on Cloud Computing Security[J]. Journal of Software, 2016, 27(6): 1328-1348. (in Chinese)
张玉清,王晓菲,刘雪峰,等. 云计算环境安全综述[J]. 软件学报,2016,27(6):1328-1348.

[3] HASHIZUME K, ROSADO D G, FERNÁNDEZ-MEDINA E, et al. An analysis of security issues for cloud computing[J]. Journal of Internet Services and Applications, 2013, 4(1): 1-13.

[4] FERNANDES D A B, SOARES L F B, GOMES J V, et al. Security issues in cloud environments: a survey[J]. International Journal of Information Security, 2014, 13(2): 113-170.

[5] ESPOSITO C, CASTIGLIONE A, POP F, et al. Challenges of Connecting Edge and Cloud Computing: A Security and Forensic Perspective[J]. IEEE Cloud Computing, 2017, 4(2): 13-17.

[6] MIAO Y, MA J, LIU X, et al. m2-ABKS: Attribute-Based Multi-Keyword Search over Encrypted Personal Health Records in Multi-Owner Setting[J]. Journal of Medical Systems, 2016, 40(11): 1-12.

[7] ZHANG L L, ZHANG Y Q, LIU X F, et al. Efficient Conjunctive Keyword Search over Encrypted Electronic Medical Records [J]. Journal of Software, 2016, 27(6): 1577-1591. (in Chinese)
张丽丽,张玉清,刘雪峰,等. 对加密电子医疗记录有效的连接关键词的搜索[J]. 软件学报,2016,27(6):1577-1591.

[8] YANG Y, MA M. Conjunctive keyword search with designated tester and timing enabled proxy re-encryption function for ehealth clouds[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(4): 746-759.

[9] WANG G F, LIU C Y, PAN H Z, et al. Survey on Insider Threats to Cloud Computing[J]. Chinese Journal of Computers, 2017, 40(2): 296-316. (in Chinese)
王国峰,刘川意,潘鹤中,等. 云计算模式内部威胁综述[J]. 计算机学报,2017,40(2):296-316.

[10] 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. (in Chinese)
田洪亮,张勇,李超,等. 云环境下数据库机密性保护技术研究综述[J]. 计算机学报,2017,40(10):2245-2270.

[11] KHALIL I M, KHREISHAH A, AZEEM M. Cloud computing security: a survey[J]. Computers, 2014, 3(1): 1-35.

[12] SINGH S, JEONG Y S, PARK J H. A survey on cloud computing security: Issues, threats, and solutions[J]. Journal of Network and Computer Applications, 2016, 75(1): 200-222.

[13] LI J W, JIA C F, LIU Z L, et al. Survey on the Searchable Encryption[J]. Journal of Software, 2015, 26(1): 109-128. (in Chinese)
李经纬,贾春福,刘哲理,等. 可搜索加密技术研究综述[J]. 软件学报,2015,26(1):109-128.

[14] DONG X L, ZHOU J, CAO Z F, et al. Research Advances on Secure Searchable Encryption[J]. Journal of Computer Research and Development, 2017, 54(10): 2107-2120. (in Chinese)
董晓蕾,周俊,曹珍富,等. 可搜索加密研究进展[J]. 计算机研究与发展,2017,54(10):2107-2120.

[15] CAO N, WANG C, LI M, et al. Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data[C]// 30th IEEE Annual Conference on Computer Communications (INFOCOM). IEEE, 2011: 829-837.

[16] WONG W K, CHEUNG D W, KAO B, et al. Secure knn computation on encrypted databases[C]// 2009 ACM SIGMOD International Conference on Management of Data. ACM, 2009: 139-152.

[17] CAO N, WANG C, LI M, et al. Privacy-preserving multi-keyword ranked search over encrypted cloud data[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(1): 222-233.

[18] FU Z, SUN X, LIU Q, et al. Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing [J]. IEICE Transactions on Communications, 2015, 98(1): 190-200.

[19] SUN W, WANG B, CAO N, et al. Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking[C]// 8th ACM SIGSAC Symposium on Information, Computer and Communications Security. ACM, 2013: 71-82.

[20] XIA Z, WANG X, SUN X, 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.

[21] SALTON G, HARMAN D. Information retrieval [M]. New York: John Wiley and Sons Ltd., 2003.

[22] CHEN C, ZHU X J, SHEN P S, et al. An Efficient Privacy-Preserving Ranked Keyword Search Method[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(4): 951-963.

[23] ZHU X Y, DAI H, YI X, et al. MUSE: An Efficient and Accurate Verifiable Privacy-preserving Multi-Keyword Text Search over Encrypted Cloud Data[J]. Security and Communication Networks, 2017, 2017(2): 1-17.

[24] LI J, WANG Q, WANG C, et al. Fuzzy Keyword Search over Encrypted Data in Cloud Computing[J]. International Journal of Engineering Research and Applications, 2014, 4(7): 441-445.

[25] WANG C, REN K, YU S, et al. Achieving usable and privacy-assured similarity search over outsourced cloud data[C]// 31st IEEE Conference on Computer Communications (INFOCOM). IEEE, 2012: 451-459.

[26] LI J G, TIAN X X, ZHOU A Y, et al. Privacy Preserving Fuzzy Keyword Search in Database as a Service Paradigm[J]. Chinese Journal of Computers, 2016, 39(2): 414-428. (in Chinese)
李晋国,田秀霞,周傲英,等. 面向 DaaS 保护隐私的模糊关键词查询[J]. 计算机学报,2016,39(2):414-428.

[27] YANG Y, YANG S L, KE M, et al. Ranked Fuzzy Keyword Search Based on Simhash over Encrypted Cloud Data[J]. Chinese Journal of Computers, 2017, 40(2): 431-444. (in Chinese)
杨旸,杨书略,柯闽,等. 加密云数据下基于 Simhash 的模糊排序搜索方案[J]. 计算机学报,2017,40(2):431-444.

[28] DEY A, JENAMANI M, THAKKAR J J, et al. Lexical TF-IDF: An n-gram Feature Space for Cross-Domain Classification of Sentiment Reviews[C]// International Conference on Pattern

Recognition and Machine Intelligence. Springer, Cham, 2017: 380-386.

[29] WANG B, YU S, LOU W, et al. Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud[C]// 33rd IEEE Conference on Computer Communications (INFOCOM). IEEE, 2014: 2112-2120.

[30] GÖGE C, WAAGE T, HOMANN D, et al. Improving Fuzzy Searchable Encryption with Direct Bigram Embedding[C]// Trust, Privacy and Security in Digital Business. Cham: Springer, 2017: 115-129.

[31] FU Z, WU X, GUAN C, et al. Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(12): 2706-2716.

[32] WANG K X, LI Y X, ZHOU F C, et al. Multi-keyword fuzzy Search over Encrypted Data[J]. Journal of Computer Research and Development, 2017, 54(2): 348-360. (in Chinese)
王恺璇, 李宇溪, 周福才, 等. 面向多关键词的模糊密文搜索方法[J]. 计算机研究与发展, 2017, 54(2): 348-360.

[33] SHI X J, HU S P. Fuzzy Multi-Keyword Query on Encrypted Data in the Cloud[C]// 4th International Conference on Applied Computing and Information Technology/3rd International Conference on Computational Science/ Intelligence and Applied Informatics/1st International Conference on Big Data, Cloud Computing, Data Science & Engineering (ACIT-CSII-BCD). IEEE, 2016: 419-425.

[34] KRISHNA C R, MITTAL S A. Privacy Preserving synonym based fuzzy multi-keyword ranked search over encrypted cloud data[C]// 2016 IEEE International Conference on Computing, Communication and Automation (ICCCA). IEEE, 2016: 1187-1194.

[35] WANG J, YU X, ZHAO M, et al. Privacy-Preserving Ranked Multi-keyword Fuzzy Search on Cloud Encrypted Data Supporting Range Query[J]. Arabian Journal for Science & Engineering, 2015, 40(8): 2375-2388.

[36] FU Z, SUN X, LINGE N, et al. Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query[J]. IEEE Transactions on Consumer Electronics, 2014, 60(1): 164-172.

[37] XIA Z, ZHU Y, SUN X, et al. Secure semantic expansion based search over encrypted cloud data supporting similarity ranking [J]. Journal of Cloud Computing Advances Systems & Applications, 2014, 3(1): 1-11.

[38] FU Z J, SUN X M, JI S, et al. Towards efficient content-aware search over encrypted outsourced data in cloud[C]// 33rd IEEE Conference on Computer Communications (INFOCOM). IEEE, 2016: 1-9.

[39] FU Z, HUANG F, REN K, et al. Privacy-Preserving Smart Semantic Search Based on Conceptual Graphs Over Encrypted Outsourced Data[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(8): 1874-1884.

[40] SAINI V, CHALLA R K, KHAN N S, et al. An Efficient Multi-keyword Synonym-Based Fuzzy Ranked Search Over Outsourced Encrypted Cloud Data[C]// 9th International Conference on Advanced Computing and Communication Technologies (ICACCT). 2016: 433-441.

[41] FU Z, SUN X, XIA Z, et al. Multi-keyword ranked search supporting synonym query over encrypted data in cloud computing [C]// 32nd IEEE International Performance, Computing, and Communications Conference (IPCCC). IEEE, 2013: 1-8.

[42] ZHAO R, LI H, YANG Y, et al. Privacy-preserving personalized search over encrypted cloud data supporting multi-keyword ranking[C]// 6th International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2014: 1-6.

[43] FU Z, REN K, SHU J, et al. Enabling personalized search over encrypted outsourced data with efficiency improvement [J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(9): 2546-2559.

[44] GENNARO R, GENTRY C, PARNO B, et al. Non-Interactive verifiable computing: outsourcing computation to untrusted workers [C]// 30th International Cryptology Conference (CRYPTO). 2010: 465-482.

[45] SUN W, WANG B, CAO N, et al. Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(11): 3025-3035.

[46] 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, PP(99): 1-12.