计算机科学
COMPUTER SCIENCE

云环境中语义感知密文检索研究综述

刘源龙, 戴华, 李张晨, 周倩, 易训, 杨庚

引用本文

刘源龙, 戴华, 李张晨, 周倩, 易训, 杨庚. 云环境中语义感知密文检索研究综述[J]. 计算机科学, 2024, 51(11): 298-306.

LIU Yuanlong, DAI Hua, LI Zhangchen, ZHOU Qian, YI Xun, YANG Geng. Research on Semantic-aware Ciphertext Rtrieval in Cloud Environments:A Survey [J]. Computer Science, 2024, 51(11): 298-306.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231000111

云环境中语义感知密文检索研究综述

刘源龙[1] 戴 华[1,2] 李张晨[1] 周 倩[3] 易 训[4] 杨 庚[1,2]
1 南京邮电大学计算机学院 南京 210023
2 江苏省大数据安全与智能处理重点实验室 南京 210023
3 南京邮电大学现代邮政学院 南京 210023
4 皇家墨尔本理工大学计算机工程学院 墨尔本 3001
(1021041407@njupt.edu.cn)

摘 要 随着云计算、大数据技术的不断发展,数据拥有者愈发倾向于将数据外包给云服务器。为了保证这些数据的安全,许多在云环境下进行的隐私保护密文检索技术被不断提出。但传统的隐私保护检索方案通常没有考虑关键词和文档之间的语义联系。为了解决这个问题,近年来,针对云环境的语义感知密文检索方案成为了研究的热点。针对基于云环境的语义感知密文检索问题,首先展示了现有研究工作主要采用的系统模型、安全模型和检索框架;接着按提取语义的核心技术对现有的语义感知密文检索方案进行分类并分别作研究和综述,阐述其优点与不足;最后,对现有研究工作进行总结,并对该领域未来的研究方向进行探讨。

关键词:云计算;隐私保护;语义感知;关键词检索;可搜索加密
中图分类号 TP391

Research on Semantic-aware Ciphertext Rtrieval in Cloud Environments:A Survey

LIU Yuanlong[1],DAI Hua[1,2],LI Zhangchen[1],ZHOU Qian[3],YI Xun[4] and YANG Geng[1,2]
1 School of Computer Science,Nanjing University of Posts and Telecommunications,Nanjing 210023,China
2 Jiangsu Key Laboratory of Big Data Security and Intelligent Processing,Nanjing 210023,China
3 School of Modern Posts,Nanjing University of Posts and Telecommunications,Nanjing 210023,China
4 School of Computing Technologies,RMIT University,Melbourne 3001,Australia

Abstract With the continuous development of cloud computing and big data technology,data owners are increasingly inclined to outsource their data to cloud servers. In order to ensure the security of these data,many privacy-preserving ciphertext retrieval techniques in cloud environments have been proposed. However,the traditional privacy-preserving search schemes usually do not consider the semantic relationship between keywords and documents. To address this problem,in recent years,semantic-aware ciphertext retrieval schemes for cloud environments have become a research hotspot.. This paper presents the existing research work on semantic-aware ciphertext retrieval in cloud environments,showcasing the system models,security models,and retrieval frameworks mainly adopted. It categorizes and summarizes existing semantic-aware searchable encryption schemes from the perspective of core technology for semantic expansion,illustrating their advantages and limitations. Finally,it concludes the existing research work and discusses future research directions in this field.

Keywords Cloud computing,Privacy protection,Semantic aware,Keyword retrieval,Searchable encryption

1 引言

云计算是一种IT基础设施模式,集成了各种IT资源和数据,并针对数据拥有者提供量身定制、可扩展的虚拟化服务[1-6]。这种模式可以让用户不受自身终端存储空间和处理能力的限制,对数据进行全面的分析和处理。然而,虽然云计算为用户提供了便利,但外包的云数据缺乏完善的安全保障,用户容易遭受隐私和安全风险[7-11]。在平台即服务(Platform as a Service,PaaS)[12]和基础设施即服务(Infrastructure as a Service,IaaS)的云服务模式中,数据在加密后

到稿日期:2023-10-18 返修日期:2024-02-07
基金项目:国家自然科学基金面上项目(62372244,61972209,61872197);江苏省研究生科研创新计划项目(KYCX22_0984)
This work was supported by the General Program the National Natural Science Foundation of China(62372244,61972209,61872197) and Postgraduate Research & Practice Innovation Program of Jiangsu Province(KYCX22_0984).
通信作者:戴华(daihua@njupt.edu.cn)

外包给云服务器，用户在云服务器上对数据进行搜索时，明文中经典的检索算法无法对加密后的数据进行搜索，这使得云环境下的密文检索成为一项挑战。基于这一挑战，许多有效的密文检索技术被陆续提出。

为了在保护数据隐私的前提下对加密数据进行检索操作，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]。目前针对多关键词的密文排序检索的研究主要采用对称可搜索加密机制，并且此类方案大都采用 TF-IDF 模型来创建表示文档和查询关键词的向量。TF-IDF 模型提取出的文档和关键词的特征信息来源于词频统计信息，但词频统计信息不考虑词与词之间的语义关联。可见，虽然 TF-IDF 模型能够有效地评估一组文档中关键词的重要性，但它忽略了词与词之间的语义关系以及词与文档之间的语义关系，导致搜索的结果可能偏离用户的实际要求。例如，当查询用户试图查询“体育”“运动”方面的文档时，现有的基于 TF-IDF 的检索方案将返回若干包含关键词{“体育”“运动”}的文档，而只包含“篮球”“足球”等与体育密切相关的关键词的文档则会被认为是低相关的，导致此类文档不会被纳入检索结果返回给用户，从而导致语义信息在搜索中的缺失。因此，具备语义感知能力的密文检索技术具有重要现实意义和实际应用价值。当前，面向云环境的语义感知密文检索已经成为可搜索加密研究领域中的一个研究热点。

本文重点关注对称可搜索加密中的语义感知密文检索问题，综述了面向外包云环境的语义感知密文检索方法。首先，对现有研究中普遍使用的系统模型、安全模型、问题描述和检索框架进行阐述；接着，从不同的角度对国内外已有的语义感知密文检索方法进行研究，介绍其使用的语义提取技术并分析其优劣；最后，归纳和总结现有的语义感知密文检索技术，研究并分析典型方案的实现方法与细节，并对未来的研究发展方向进行思考和展望。根据本文的研究和分析，怎样更加准确地提取文本之间的语义信息，并实现高效的语义感知密文检索，是目前基于云环境的语义感知密文检索进一步发展的主要方向。

本文第 2 章介绍了系统模型、安全模型、问题描述和检索框架；第 3 章从多个角度对现有的语义感知密文检索方案进行分析；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

当前普遍的基于云环境的语义感知密文检索的系统模型如图 1 所示，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）。各实体的主要工作方式如下。

1）数据拥有者：DO 负责为明文文档做处理。DO 在将明文文档集和索引外包给云服务器之前对它们进行加密和语义信息的提取，并执行算法来构建包含语义信息的安全索引。随后，DO 将加密文档集合以及安全索引外包给 CS。

2）云服务器：CS 为 DO 和 DU 提供数据存储和计算服务。CS 存储 DO 发送来的数据，并在接收到 DU 的搜索指令后，利用加密的索引进行隐私保护、排序的搜索操作。成功搜索之后，CS 将所需的目标文档返回给 DU。

3）数据使用者：DU 是获得 DO 授权的资料使用者，可查阅储存于个人资料中心的资料。DU 根据具体检索需求提取语义信息并创建包含语义信息的检索指令，然后将检索指令提交给 CS。一旦 CS 执行搜索并将结果文档发送给 DU，则对其进行解密以获得原始的明文信息。

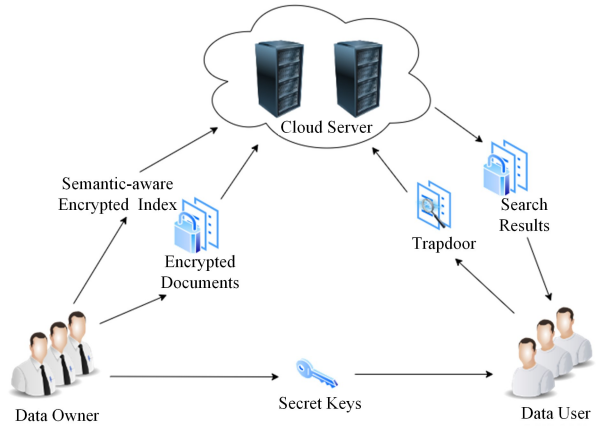


图 1 语义感知密文检索系统模型

Fig. 1 System model of semantic-aware search over encrypted data

2.2 安全模型

在基于云环境的语义感知密文检索技术的研究中，根据 CS 对存储数据的处理情况，存在如下两种常用的安全模型。

1）诚实而好奇（Honest-but-Curious）模型

在诚实而好奇模型下，数据拥有者和云服务器严格地遵守协议并执行所有任务，但是云服务器也将对查询和加密数据保持好奇，它试图获取加密数据的明文信息，并尽可能地对所有中间计算结果进行统计、分析和窥探来获得一些附加信息。根据云服务器所持有信息的不同级别，主要有两种安全模型[10]，即已知密文模型（Known Ciphertext Model）和已知背景模型（Known Background Model），这两个模型分别考虑了在不同的实际情况下如何确保数据的安全性。云服务提供商需要谨慎平衡好奇心和隐私保护，以确保用户数据得到充分的安全保护。

（1）已知密文模型。在该模型中，云服务器只掌握加密文档的集合，加密索引和检索陷门。在这种情境下，云服务器只具备对密文的访问权限，但无法直接获取明文信息。然而，通过纯密文攻击，云服务器可以尝试从密文中推断安全密钥，进而获取加密索引的明文。这种攻击方法依赖于对密文的分析和破解，以揭示底层安全机制的弱点。在应对这一挑战时，密钥的选择和安全索引的保护变得至关重要，以防止潜在的信息泄露。维护在已知密文模型下的安全性，需要采取有效的加密技术和密钥管理策略，以确保用户数据在云环境中的隐私得到可靠的保护。

(2)已知背景模型。在该模型中,云服务器能够掌握更多的背景信息,包括文档长度、数据集的词频统计信息和搜索请求之间的联系等。这些统计信息揭示了文档中特定关键词的词频分布,云服务器可能会滥用这些知识来进行统计攻击,通过分析相应的频率分布直方图或值范围甚至可以推断识别出某些关键词。例如,云服务器可以利用词频信息获取文档集中每个关键词的频率分布,并根据后台知识进一步发起统计攻击。

2)恶意攻击(Malicious Attack)模型

在恶意攻击模型下,云服务器可以任意地偏离原定的协议规则,如不执行原有协议,只执行部分协议以及提前终止协议等。同时,云服务器还可能在内部权限滥用和外部黑客攻击的情况下对外包的加密数据进行攻击,包括篡改数据、删除数据以及更改计算结果等。各参与方会恶意地破坏协议运行以节省自身的算力或获取其他参与方的数据。在该安全模型下,隐私泄露的风险较高。

诚实而好奇模型在实际环境中有着广泛的应用,尤其在云服务提供商对用户数据进行推荐时以及其他数据分析场景中。云服务提供商,如亚马逊的推荐系统,通常需要分析用户的行为和偏好以提供个性化的推荐服务。在这个过程中,诚实而好奇模型能够在保持用户隐私的同时,收集足够的信息进行有效的推荐。但在该模型下,云服务器只能获取对系统提供服务所必需的信息,而不会去窥探用户的敏感数据,同时还会正常执行与用户约定的义务。而恶意攻击在现实中通常发生在一些不道德的实体或黑客试图通过各种手段获取未经授权的信息或者破坏系统的完整性时。在实际环境中,这些攻击可能采用多种形式。为应对这些威胁,系统需要采用多层次的安全措施来防范恶意攻击。

总体而言,现有的绝大多数语义感知密文检索方案普遍采用诚实而好奇模型。然而,在实际情况下,恶意攻击模型所描述的场景也不是完全不存在,只是因为在这种模型下进行加密搜索的难度较大,所以现有研究基本未针对这种安全模型做过多讨论。本文的研究主要讨论诚实而好奇安全模型下的语义感知密文检索研究。

2.3 问题描述

基于云环境的语义感知密文检索技术主要是对云环境下的可搜索加密方案进行语义感知能力的扩展,主要解决如下关键问题。

1) 语义感知多关键词检索(Semantic-aware Multi-keyword Search):云服务器能够根据用户生成的搜索陷门,度量查询关键词与文档语义相似度,返回与查询的语义最相关的 k 个文档。

2)隐私保护(Privacy Preserving):技术方案要保证数据的安全性、索引的安全性、搜索陷门的安全性、搜索陷门的不可链接性以及关键词的安全性。

3)检索效率和准确率(Efficiency and Accuracy):技术方案在保证一定准确率的前提下,能够运用各种算法优化提高检索效率。

2.4 检索框架

当前实现语义感知密文检索的方法普遍采用以下框架结构,主要包含 4 个部分的算法。

1)Genkey:密钥生成算法,主要依据方法所使用的加密方法生成加密所需的各种密钥。该算法在 DO 端执行,执行完成后将密钥分享给 DU。

2)Setup:预处理算法,主要完成数据外包到 CS 前的预处理工作,包括加密文档、提取数据的语义信息、生成索引和加密索引等步骤。该算法在 DO 端进行实施,执行完成后将数据上传至 CS 端。

3)Trapdoor:搜索陷门生成算法,主要负责为查询关键词生成搜索陷门。该算法在 DU 端进行实施,DU 使用陷门生成算法将关键词构造为具体的搜索陷门,然后发送给 CS。

4)Search:关键词检索算法,主要负责语义感知密文检索。该算法在 CS 上进行实施,DU 将搜索陷门发送给 CS,CS 在接收到检索陷门后,在加密索引上执行密文检索,并将检索结果返回给 DU。

现有的多数语义感知密文检索方法大多采用上述检索框架,因此,该框架在语义感知密文检索方向最具有代表性。其中,在预处理算法和搜索陷门生成算法部分,每个具体方案所采用的技术和方法可能有所不同,但是主要完成的工作大致相同,比如都在预处理阶段完成对语义信息的提取,只是具体细节不同。其他方案可能根据具体情况增减特殊算法,在此不作详述。

3 语义感知密文检索技术

已有的密文多关键词排序加密检索方案多将关键词作为文档特征,然而,基于关键词的搜索方案忽略了用户检索的语义表示信息,不能完全匹配用户的搜索意图。因此,一些方案将可搜索加密技术与语义感知功能相结合,以实现更符合现实需求的密文检索。本章针对现有的语义感知密文检索方案,从不同种类的语义扩展方法出发对这些方案进行研究和分析。

3.1 基于同义词扩展的语义感知密文检索方法

在现实生活中,一个关键词拥有大量的同义词,而用户在进行检索时往往会忽略关键词的同义词,导致忽略了关键词暗含的语义信息,因此将关键词的同义词纳入考虑范畴,实现基于同义词扩展的语义感知密文检索方法能够符合现实需求。文献[15]使用 NARCT 词典对文档的关键词集合进行同义词扩展,使得搜索输入的关键词可以匹配到文档关键词的同义词,并通过安全 k 近邻(Secure k-Nearest Neighbor, SeckNN)技术[16]和向量空间模型实现了支持同义词检索的多关键词密文排序搜索方案;另外通过引入虚拟关键词,很好地保护了敏感的频率信息。然而该方案对整个关键词集合进行同义词扩展,且依旧使用 TF-IDF 来计算相似度,使得检索的时间和空间开销较大、语义准确率不高。

为了解决文献[15]的检索时间和空间开销较大、语义准确率不高的问题,文献[17-19]分别基于关键词的共现概率

构建了一个语义关系库(Semantic Relationship Library，SRL)来记录关键词之间的语义相似度，提高了检索的语义准确率。在搜索阶段，根据 SRL 可以将查询关键词进行扩展，在搜索算法中使用扩展后的查询关键词，以此实现语义扩展。其中，为了提高检索效率，文献[17]为文档集合创建倒排索引，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，提高了检索的效率。文献[18]提出的 RSS 方案，使用改进的一对多的保序加密(One-to-Many Order-Preserving Encryption，OM-OPE)方案，在保证总相关分数计算的同时，保护词频信息安全。文献[19]在文献[18]的基础上优化了检索效率，利用了私有云和公有云两种云的架构，将 SRL 存储在私有云中以保证安全，加密后的索引上传到公有云以提高搜索效率。为了在语义感知密文检索方案中支持模糊搜索功能，文献[20]提出了一种基于多关键词同义词的模糊排名搜索方案 MSFRS。该方案会在搜索时提示关键词的同义词，并根据相同排名的文档中关键字的出现频率评估相关分数并返回结果。文献[21]根据相关关键词通常具有相同根的特点，采用词干提取算法提取关键词根，用根代替关键词进行搜索。但这种方法的缺点是当语义相关的关键词具有不同的根时，效果不佳。

总的来说，基于同义词扩展的语义感知密文检索方法能够解决关键词检索中语义缺失的问题，但总体检索结果的语义准确度不高，消耗也较大。

3.2 基于 WordNet 扩展的语义感知密文检索方法

除了同义词以外，一个关键词的上义词和下义词也包含着该关键词的语义信息。WordNet[22]字典是一种基于认知语言学的英语词典，使用该词典不仅能够扩展关键词的同义词，也可以扩展关键词的上义词和下义词，因此也被引入到语义感知密文检索领域中。文献[23]提出了一种基于 WordNet 构建的 m-best 树和术语相似树对查询关键词进行语义扩展的语义感知密文检索方案 VKSS。VKSS 根据 WordNet 对查询关键词进行扩展，并使用 m-best 树构造术语相似树模型。m-best 树由关键词组成，根和叶之间的路径表示关键词之间的相似性。由 m-best 树构造的术语相似树模型可以表示与任意关键词之间的语义关系。为了证明方案有更高的安全性，VKSS 还提出了一种半诚实而好奇的安全模型。在此模型中，云服务器以一种自私的方式运行，不会完全遵循原定的协议。在这种安全模型下，为了防止结果不完整或不准确，VKSS 设计了一种基于符号的树结构索引，在该索引上进行搜索以支持搜索结果的可验证性。

针对多种语言的环境，文献[24]提出的跨语言的多关键词语义排序查询方案 CLRSE 在扩展查询语义的同时还可以实现语种转化。CLRSE 使用基于 OMW(Open Multilingual WordNet)构建的跨语言目标查询系统扩展语义；使用带阈值解密的 Paillier 密码系统(Paillier Cryptosystem with Threshold Decryption，PCTD)进行陷门不可链接的加密；并且使用相关度评分计算协议(Relevance Score Computation Protocol，RSCP)来计算语义相似度；还根据扩展关键字的语义自动计算偏好分数，实现智能和个性化排序搜索。文献[25]提出了一种有效的模糊语义搜索加密方案 FSSE。该方案利用关键词指纹生成算法来生成关键词字典的指纹集和查询关键词的指纹，并使用汉明距离来量化关键词的相似度，实现了模糊搜索。在此基础上，利用 WordNet 字典对查询关键词进行语义扩展，计算查询关键词与扩展词之间的语义相似度，实现语义搜索。同时，使用倒排索引提高检索效率，并使用向量相交匹配和短路匹配操作来有效地过滤不相关文档。为了保证索引的保密性，倒序索引采用改进的整数上全同态加密(Fully Homomorphic Encryption over the Integers，FHEI)[26]方案进行加密。FHEI 降低了加密的复杂度，提高了方案的效率。

基于 WordNet 扩展的语义感知密文检索方法进一步提高了语义信息的准确度，但由于仍然是基于传统的关键词检索，总体的语义准确度不高。

3.3 基于概念图的语义感知密文检索方法

为了解决语义感知问题，一些研究引入了概念图(Conceptual Graph，CG)[27]作为语义表示工具。概念图是一种基于第一逻辑的知识表示结构，是一个有限连通图，它能够自然、简单、细粒度地描述文本的语义信息，但是在加密形式的概念图上很难进行匹配。文献[28]提出了一种采用概念图表示语义的方案，将基于关键词的结构化加密方案 SSE[29]推广到概念图上，提出了一种称为消息查询的方案 MeQ。该方案预先计算一个索引表作为数据结构，用户为所有要存储的文档构造概念图并执行图同态转换，进而构建索引表，然后围绕索引表构建结构化加密方案。当查询开始时，将给定的查询短语或句子转化为一个结构为概念图的查询，在加密的文档数据库中通过概念图同态比较相似度来检索与查询匹配的加密文档。但该方案在加密之前执行了概念图同态，这意味着该方案无法对加密后的数据进行操作，无法实现真正意义上的可搜索加密。

文献[30]提出了一种能够在概念图上执行搜索的方法 SSCG，首先利用文本摘要从文档中提取出最重要和最简化的主题句，然后将这些简化的句子转换成概念图，再使用一种方法将概念图映射到向量上。该方法使用一对多保序加密方法对文档进行加密，查询时根据文本摘要得分对返回的结果进行排序并返回相关度最高的若干加密文档。但这是一个初始的方案，效率不高。在文献[30]的基础上，文献[31]提出了 PRSCG 方法，为了进行数值计算，将原始的概念图转换成线性形式。该方法将文件中的所有句子或最重要的句子整理成概念图，并将该图的每个部分看作一个整体，然后对概念图进行分割并获得它们的线性形式。该方法将一个概念图分成多个实体并将它们视为具有足够语义信息的“关键词”，根据向量空间模型生成向量来替换原始文档，进而实现语义信息的提取。该方案还使用虚拟关键词提高方案的隐私性，并且可以抵抗规模分析攻击。以 PRSCG 方案为基础，文献[32]和文献[33]提出了后续方案 ECSED。该方案采用了一个扩展的概念层次(Extended Concept Hierarchy，ECH)来表示概念之间的语义关系，使概念拥有属性，这些属性可以被赋予不同

的值。该方案通过扩展属性信息实现了更为精确的语义提取，从而提高了检索结果的准确率，因为它使搜索请求更加具体。最后，为了加快查询过程的速度，使用二叉树结构索引存储文档，使得检索过程的时间复杂度显著降低，提高了检索效率。

使用概念图形式扩展语义的方法不再使用传统的关键词作为知识表示工具，而是以加密的概念图等形式扩展语义，将概念图映射到向量并定量计算，进一步优化了语义搜索功能。

3.4 基于 Word2Vec 和 Doc2Vec 模型的语义感知检索方法

Word2Vec[34]是一个将单词转换成包含语义信息向量形式的工具。使用神经网络来训练数据集，Word2Vec 从周围的上下文单词窗口中精确地识别当前的单词并将其转化为向量空间中的向量，从而将文本之间的语义相似度比较转化为向量空间上的相似度计算，解决了词袋模型忽视词之间语义的问题。文献[35]提出了一种基于 Word2Vec 生成索引向量的查询方案 SSSW。其基本思想是将 Word2Vec 引入隐私保护搜索的研究中，选择 MRSE[10]作为基本框架。首先通过 Word2Vec 为每个文档构建的关键词集生成词向量，然后引入余弦距离来表示词向量之间的相似度，选择与查询词向量最为相似的文档作为查询结果。但是该方法直接对所有单词向量进行聚合，会破坏一些词嵌入的语义信息。文献[36]同样使用 Word2Vec 为语料库中的所有关键词创建字典，其中每个关键词与几个语义相关的关键词相关联，接着再将每个文档的关键词集进行扩展，将一些与旧关键词在语义上相关的新关键词加入字典。最后，通过将查询和索引关键词集分别转换为谓词和属性向量来计算相关度进行检索，提出了一种基于高效内积加密方案的 SPE-SMKS 方案。文献[37]提出了一种基于 Word2Vec 模型的上下文感知语义扩展可搜索加密方案 CASE-SSE。该方案以外包数据集为训练集，训练 Word2Vec 模型生成原始数据集的分布式词向量表示。为了进一步提高检索效率，在安全索引方面，CASE-SSE 提出了一种双层索引结构。首先利用 k-means 聚类算法对原始数据集进行分类；然后基于关键词类别构建平衡二叉树索引；最后基于类别对文档构造倒排索引，提高了检索的效率。为了实现模糊搜索功能，文献[38]提出了一种隐私保护的多关键字语义感知模糊搜索方案 SE-PPFM。该方案采用哈夫曼树中的 TD-IDF 作为文档中单词的关联权值，并通过 Word2Vec 创建关键词向量来实现同义词的模糊搜索，同时将非对称的向量积保持加密(Asymmetric Scalar-product Preserving Encryption，ASPE)与 Hadamard 积相结合，构建加密索引，从而实现语义感知密文检索。

Doc2Vec 是 Word2Vec 的扩展，使用与 Word2Vec 类似的方法将文档转化为包含语义信息的文档向量。由于 Word2Vec 模型只提取每个关键词的语义特征，而 Doc2Vec 模型使用文档向量来表示每个文档的语义特征，因此 Doc2Vec 更适合在加密文档上执行语义感知搜索。文献[39]提出了一种基于 Doc2Vec 模型的语义感知密文检索方案 DMRSE。DMRSE 使用 Doc2Vec 模型提取文档语义特征，并生成特征文档向量，采用安全内积运算对索引和查询向量进行加密；同时，通过索引与查询向量的加密形式，计算出索引与查询向量之间的相关性得分并给出结果。基于 Doc2Vec 模型的 infer 功能，该方案还可以支持动态更新，但是没有给出一个完善的动态更新流程和标准。

随着神经网络技术的发展，上述方案利用 Word2Vec 和 Doc2Vec 等神经网络训练词汇表示向量，不仅极大地加快了模型的训练速度，也改进了语义表示能力。但是，同其他神经网络模型一样，该模型仅利用局部上下文作为输入，缺乏对全局统计信息的利用，依然有进一步改进的空间。

3.5 基于 LDA 模型的语义感知检索方法

LDA 模型[40](Latent Dirichlet Allocation Model)是一个三层贝叶斯概率模型，是一种无监督学习算法。该模型将主题视为文档和关键词之间的连接，对具有主题联系的文档集进行分类，并且提取出文档中隐式的主题信息，以此来表示文档和关键词的语义。这一特性使得 LDA 模型能够应用于语义感知密文检索领域。

文献[41]提出的 LDA-MRSE 方案首次将 LDA 主题模型应用到可搜索加密方向，将用户搜索意图的潜在语义特征融入可搜索加密中。文档经过 LDA 主题模型训练后被转化为主题向量，生成文档-主题相关矩阵和查询主题向量；采用安全 kNN 方法对文档主题向量的索引和查询主题向量进行加密，通过对加密后的索引与查询之间进行语义相关性计算来获得与查询语义最相关的文档结果。该方法一方面能够更好地提取文档和关键词的语义特征；另一方面也减少了向量空间的维度，提高了检索效率。文献[42]在文献[41]的基础上，针对 LDA-MRSE 只考虑主题信息却忽略了关键词信息的问题，提出方案 LDA-ESSS。该方法使用基于 LDA 的信息增益(IG)和主题频率-逆主题频率(TF-ITF)模型来确保关键词信息不会被 LDA 模型忽略。

此外，为了提高检索效率，文献[43]在使用 LDA 语义感知模型的同时，对文献[41]中的文档存储索引进行优化，提出了精度优先密文检索方案 AF-SRSE 和效率优先密文检索方案 EF-SRSE。其中，AF-SRSE 引入二分 k-means 聚类算法，自底向上构建一棵精度优先过滤树索引 AFF-tree，使得潜在语义相关度高的向量存储在相邻的叶节点中；利用该索引，提出了一种能够保证精度的密文检索算法。在 EF-SRSE 方案中，通过双向链接 AFF-tree 中最下层的叶子节点，将 AFF-tree 改造成效率优先过滤树索引 EFF-tree，通过该索引的使用，提出了一种能够进一步提高效率但精度略有损失的排序检索算法。文献[44]提出的 SAPMS 方法与文献[43]的方法类似，也使用 LDA 模型来提取文档的语义特征，并将其表征为主题语义向量，然后通过分裂型层次聚类算法对文档向量进行自顶向下的聚类处理，在聚类过程中同步构建树形结构索引 BCI-tree，并将该索引应用到密文检索中。该方法将构建树形结构索引的过程与聚类过程相融合，提高了索引的构建效率。文献[45]在文献[44]的基础上，增加了针对密文检索算法执行过程中对于初始过滤阈值的预筛选机制，加速了基于树形结构索引的检索过程中的剪枝处理，在保证结果正确性的同时，进一步提高了密文检索的执行效率。

使用 LDA 主题模型发现文本中的潜在主题能够很好地提取文档的语义。然而，LDA 模型是一种无监督学习方法，对文档主题的提取可能不准确，并且忽略了关键词的顺序和上下文信息，会损失一定的语义精度。

3.6 基于 BERT 模型的语义感知检索方法

近年来，许多新语义模型不断被提出，也有一些方案采用了这些模型来进行语义的扩展。其中比较常用的就是 BERT 模型[46]。BERT 模型是一个预训练的表征模型，它不同于传统的采用单向语言模型进行预训练或者把两个单向语言模型进行浅层拼接进行预训练的方法，缓解了单向语言模型对预训练表征的限制，因此能够更好地对文本的语义特征进行分析。

文献[47]提出了基于 BERT 模型的语义感知加密搜索方案 SSRB-1 和 SSRB-2。其中，SSRB-1 首先将整个文本训练成一个包含文本语义信息的向量，通过注意模型机制使语义向量上下文相关，并通过 BERT 模型的掩模语言预训练使训练出来的向量更具语义性。最后，通过 MRSE[10] 框架，SSRB-1 实现了云环境下的密文检索。但在已知背景的诚实而好奇安全模型中，云服务器可以分析和推断特定信息，如关键词和文档的频率。SSRB-2 方案则通过随机插入更多的虚拟关键词来解决这个问题。通过将随机数加到向量上混淆了查询请求和查询结果之间的关系，增加了关键词推理的难度。又因为是随机数，所以具体生成的搜索陷门也是不同的，保证了搜索陷门的安全性和不可链接性。

文献[48]提出了一种基于改进的 BERT 模型（Sentence-BERT，SBERT）的语义感知密文检索方案 SBERT-SRSE。SBERT 在 BERT 模型的输出中添加一个池化操作，使训练后的语义特征更充足。数据所有者首先使用 SBERT 模型、均匀流形近似和投影降维算法 UMAP 以及基于密度的带噪声层次聚类方法 HDBSCAN 对文档进行训练，得到文档-主题关联矩阵，然后提出了主题频率-逆主题频率模型 TTF-ITF 来获得关键词-主题关联矩阵。在文档向量存储时，使用倒排索引增强了检索的效率。当开始搜索 k 个最相关的文档时，用户将查询的关键词转换为查询主题向量并加密传给云服务器。云服务器对加密索引执行搜索，并返回加密后的结果。使用 UMAP 降维一方面降低了计算复杂度，使损耗大大减少；另一方面在局部的语义特征提取上效果更好。

BERT 模型缓解了单向语言模型对预训练表征的限制，因此能够更好地对文本的语义特征进行分析。但该模型目前在密文检索中的使用还较少，可能存在一些还没有被发现的问题。

3.7 基于其他方法的语义感知检索方法

除了上面提到的几类语义感知密文检索的方案以外，还有一些其他的可以对文档语义进行检索的可搜索加密方案。这些方案所使用的方法使用的频率较低，因此单独在这一节列出。

文献[49]提出了一种复合概念语义相似度（Compound Concept Semantic Similarity，CCSS）计算方法来衡量复合概念之间的语义相似度，并基于该计算方法提出复合语义关键词搜索方案 SCKS。该方案在向量空间模型的基础上，结合主题信息来实现语义感知检索；将关键词向量的每个元素对应一个主题，使用 CCSS 相似度模型计算关键词与主题的语义相似度。CCSS 相似度模型使用复合概念计算相似度，将复合词分解为主题词和助词，综合考虑了本体的信息源，如分类特征、局部密度、路径长度和深度等，能够有效地提高检索结果的准确率。同时，该方案还消除了预定义的全局库，支持动态的数据更新。

为了支持结果的可验证性，文献[50]提出了一种安全可验证的语义搜索方案 SVSS。该方案使用词嵌入来将词转化为向量并保留两者之间的语义关系。针对密文的语义最优匹配问题，该方案以最小词传输代价 MWTC 作为查询与文档之间的相似度，并使用一种安全转换方法，将 MWTC 转化为随机线性规划（Linear Programming，LP）问题，从而获得加密后的 MWTC。在可验证性方面，利用 LP 的对偶定理，设计了一种将匹配过程中的中间数据与结果相互验证来保证准确性的验证机制。在安全模型方面，该方案提出了一种更具挑战性的安全模型，这种模型优于其他安全语义搜索方案中使用的诚实而好奇模型，能够应对不诚实的云服务器试图返回错误或伪造的搜索结果并学习敏感信息的情况。

4 总结与展望

4.1 对现有工作的总结

本文主要介绍了语义感知密文检索方法的研究现状，下文将从现有典型研究工作所采用的语义模型、安全模型、索引机制和隐私保护机制方面对现有工作进行总结。

现有的语义感知密文检索工作的研究总结如表 1 所列，各方案在扩展文档语义方面都给出了各自的方法。针对原有的关键词与文档之间语义关系被忽略的问题，其中一类方案对关键词和文档用同义词进行语义扩展，如使用 NARCT 词典扩展文档关键词的 SSERS，使用关系语义库扩展查询的 RSS 等。另外，使用 WordNet 字典扩展语义的 VKSS，CLRSE 和 FSSE 等方法也实现了语义感知密文检索，但是这些方法还是基于传统的关键词检索，语义准确率相对较低。另一类方案不再使用传统的关键词作为知识表示工具，而是使用概念图以加密的形式扩展语义，将概念图映射到向量并定量计算，实现了语义搜索功能，如 SSCG，PRSCG 和 ECSED 等。随着神经网络技术的发展，SSSW，CASE-SSE 和 DMRSE 等方案利用 Word2Vec 和 Doc2Vec 等神经网络训练词汇表示向量，不仅极大地加快了模型的训练速度，也改进了语义表示能力。但是，与其他神经网络模型一样，这类模型仅利用局部上下文作为输入，缺乏对全局统计信息的利用，依然有进一步改进的空间。还有一类方案使用 LDA 主题模型对词之间的关系行建模，以发现文本中的潜在主题，并通过主题表示匹配查询和文档，如 LDA-MRSE 和 SRSE 等。然而，LDA 模型是一种无监督学习方法，对文档主题的提取可能不准确，并且忽略了关键词的顺序和上下文信息，会损失一定的语义精度。还有使用 BERT 模型表示语义的 SSRE 和 SBERT-SRSE 等方案，BERT 模型缓解了单向语言模型对预训练表征的限制，因此能够更好地对文本的语义特征进行分析。

表 1 现有典型研究工作的研究总结

Table 1 Research summary of existing typical research works

研究工作	语义模型	安全模型	索引机制	隐私保护机制
SSERS[15]	NARCT	诚实而好奇	二叉平衡树索引	安全 kNN,对称加密
RSS[18]	SRL	诚实而好奇	倒排索引	OM-OPE
VKSS[23]	WordNet	半诚实而好奇	基于符号的索引	块加密
CLRSE[24]	OMW	诚实而好奇	完全二叉树索引	改进 Paillier 同态加密,RSCP 协议
FSSE[25]	WordNet	诚实而好奇	倒排索引	改进的 FHEI
SSCG[30]	CG	诚实而好奇	—	安全 kNN,OM-OPE
PRSCG[31]	CG	诚实而好奇	—	安全 kNN
ECSED[33]	ECH	诚实而好奇	二叉树索引	安全 kNN
SSSW[35]	Word2Vec	诚实而好奇	—	安全 kNN,对称加密
CASE-SSE[37]	Word2Vec	诚实而好奇	二叉平衡树索引+倒排索引	安全 kNN
DMRSE[39]	Doc2Vec	诚实而好奇	—	安全 kNN,对称加密
LDA-MRSE[41]	LDA	诚实而好奇	完全二叉树索引	安全 kNN,对称加密
SRSE[43]	LDA	诚实而好奇	二叉树索引	安全 kNN,对称加密
SSRB[47]	BERT	诚实而好奇	—	安全 kNN,对称加密
SBERT-SRSE[48]	SBERT	诚实而好奇	倒排索引	安全 kNN,对称加密
SCKS[49]	CCSS	诚实而好奇	散列索引	安全 kNN,对称加密
SVSS[50]	WordEmbedding	半诚实而好奇	前向索引	对称加密

分析表 1 可知,现有的多关键词密文检索方法大多针对诚实而好奇的安全模型进行研究。针对已知背景的诚实而好奇安全模型,SSERS,PRSCG,SSSW 和 SSRB 等还通过加入虚拟关键词等方法使得服务器无法从特定的频率信息中推断出其他信息,抵抗了统计分析的攻击。在索引机制方面,只有很少的方案没有对索引进行优化,其他大多数方案如 SSERS,ECSED 和 LDA-MRSE 等都使用各种树形索引对存储进行了优化;还有一些方案如 FSSE 和 SCKS 等使用倒排、散列等其他索引对索引进行优化。对索引进行优化,可以使方案在搜索效率和耗费空间上获得一定的性能提升。最后,在隐私保护机制方面,大多数方案都采用的是安全 kNN 技术。从上述角度分析,对文档包含的语义信息的提取方法以及数据存储的索引结构是目前研究的两大重点,如何更精确地提取出数据的语义信息,构建安全高效的索引,是研究工作的关键。同时,使方案能够适应不同的安全需求,改进隐私保护机制,也是研究的发展方向。

4.2 对未来工作的展望

云计算飞速发展的当下,在已经提出的各种可加密搜索方案中,支持语义感知的密文检索方案逐渐成为新的重点研究领域。当前已经提出了许多基于云环境的语义感知密文检索技术,其中一些已经具有实际的应用价值和成果,但在检索结果更加准确的语义感知密文检索、基于高效索引机制的语义感知密文检索、面向不同安全需求的语义感知密文检索、改进隐私保护机制的语义感知密文检索、支持结果可验证的语义感知密文检索和具备完善的动态更新功能的语义感知密文检索等方面,还有很大的研究空间。本文重点从以下几个方面对语义感知密文检索研究的未来工作进行展望。

1)检索结果更加准确的语义感知密文检索方案。现有的语义感知密文检索方案提出了各种不同的语义感知方法。在扩展文档和关键词的语义特征时,学术界给出了各种不同的方案。但一方面,现有的提取语义特征的方法还存在不足之处,如 LDA 模型会忽略关键词的信息,Word2Vec 模型缺乏对全局统计信息的利用等。另一方面,在明文环境下,不断有准确率更高的语义训练模型被提出,而这些模型还暂时没有被引入到密文检索方向中来。因此研究检索准确率更高的语义感知方法,是语义感知密文检索研究工作中的一个重要发展方向。

2)基于高效索引机制的语义感知密文检索方案。现有的各种语义感知密文检索方案,大多对文档在云服务器上存储的索引结构进行优化,以提高检索的效率。现有方案中基于树形结构的索引在大规模数据的情况下表现不佳,而其他索引结构的效率往往不如树形索引,还有一些新式索引还没有被应用到语义感知密文检索方案中。因此,设计和实验新的高效索引机制,实现检索高效且便于更新的索引结构,也是语义感知密文检索研究中的一个重要课题。

3)面向不同安全需求的语义感知密文检索方案。现有的语义感知密文检索方案通常只针对诚实而好奇的安全模型,有些方法甚至没有考虑到诚实而好奇模型下的已知背景模型,只考虑了已知密文模型。另外,针对另一种安全模型即恶意攻击模型的语义感知密文检索方案鲜有被提出。同时,还有许多新的安全模型不断被提出,语义感知密文检索方案也应该对这些新的安全模型进行适配。因此,如何根据安全需求的不同,研究面向各种安全模型的语义感知密文检索方案同样值得探索。

4)改进隐私保护机制的语义感知密文检索方案。现有的语义感知密文检索技术中使用的隐私保护机制多为安全 kNN 方法,但安全 kNN 方法在加密文档向量或检索向量时,需要大量的计算资源和存储资源,不适用于超大规模密文检索应用场景。此外,一些现有方案使用的加密方法随着研究的深入也被证明不够安全,如文献[51]证明了使用稀疏矩阵的对称加密不具有不可区分性。因此,如何适应现实隐私保护需要,研究采用具有更高安全等级隐私保护策略的语义感知密文检索方案也是未来的发展方向。

5)支持结果可验证的语义感知密文检索方案。语义感知密文检索能够保证数据的隐私性,但在数据的隐私性得到保证的前提下,由于内部权限滥用和外部黑客攻击,云服务器可能会为用户提供被篡改或伪造过的不完整或不准确的搜索结果,而在一些对检索准确性结果要求较高的情况下,数据

被篡改和伪造会导致严重的后果。因此，如何在语义感知可搜索加密的方法中确保检索结果的准确性，研究结果可验证的语义感知密文检索方案同样也是待研究的问题。

6）具备完善的动态更新功能的语义感知密文检索方案。现有的语义感知密文检索方案在将数据上传到云服务器之前通常对文档的语义信息进行提取，但是在对数据进行更新之后，许多目前的方案还需要重新对所有数据进行语义信息的提取，这显然不符合现实需求。而对于一些已经具备动态更新能力的方案，目前缺少完备的标准来判定其动态更新能力的好坏。因此，如何在语义感知可搜索加密的方法中加入完善的动态更新功能，研究具备完善的动态更新功能的语义感知密文检索方案同样也是待研究的方向。

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

[4] ALAHMAD S, KAHTAN H, ALZOUBI I, et al. Mobile Cloud Computing Models Security Issues: A Systematic Review [J]. Journal of Network and Computer Applications, 2021, 190: 1-17.

[5] WANG T, LIANG Y, TIAN Y, et al. Solving Coupling Security Problem for Sustainable Sensor-cloud Systems based on Fog Computing [J]. IEEE Transactions on Sustainable Computing, 2021, 6(1): 43-53.

[6] MTHUNZI 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, 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] HASSAN J, SHEHZAD D, HABIB U, et al. The rise of cloud computing: data protection, privacy, and open research challenges—a systematic literature review (SLR) [J]. Computational Intelligence and Neuroscience, 2022, 2022: 1-26.

[10] 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, 2013, 25(1): 222-233.

[11] SUN P J. Privacy protection and data security in cloud computing: a survey, challenges, and solutions[J]. IEEE Access, 2019, 7: 147420-147452.

[12] MELL P. The NIST Definition of Cloud Computing[J]. Communications of the ACM, 2010, 53(6): 50-50.

[13] LI J W, JIA C F, LIU Z L. Survey on the Searchable Encryption [J]. Journal of Software, 2015, 26(1): 109-128.

[14] DONG X L, ZHOU J, CAO Z F. Research Advances on Secure Searchable Encryption[J]. Journal of Computer Research and Development, 2017, 54(10): 2107-2120.

[15] 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.

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

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

[18] SUN X, ZHU Y, XIA Z, et al. Secure keyword-based ranked semantic search over encrypted cloud data[J]. Proceedings of the Advanced Science and Technology Letters (MulGraB 2013), 2013, 31: 271-283.

[19] SUN X, ZHU Y, XIA Z, et al. Privacy-preserving keyword-based semantic search over encrypted cloud data [J]. International Journal of Security and Its Applications, 2014, 8(3): 9-20.

[20] SAINI V, CHALLA R K, KHAN N S. An efficient multi-keyword synonym-based fuzzy ranked search over outsourced encrypted cloud data[C]// Advanced Computing andCommu-nication Technologies: Proceedings of the 9th ICACCT. Singapore: Springer, 2016: 433-441.

[21] MOATAZ T, SHIKFA A, CUPPENS-BOULAHIA N, et al. Semantic search over encrypted data [C] // International Conference on Telecommunications 2013. Casablanca: IEEE, 2013: 1-5.

[22] MILLER G A. WordNet: a lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41.

[23] FU Z, SHU J, SUN X, et al. Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data [J]. IEEE Transactions on Consumer Electronics, 2014, 60(4): 762-770.

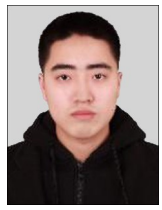
[24] GUAN Z, LIU X, WU L, et al. Cross-lingual multi-keyword rank search with semantic extension over encrypted data[J]. Information Sciences, 2020, 514: 523-540.

[25] LIU G, YANG G, BAI S, et al. FSSE: An effective fuzzy semantic searchable encryption scheme over encrypted cloud data[J]. IEEE Access, 2020, 8: 71893-71906.

[26] YU J, LU P, ZHU Y, et al. Toward securemultikeyword top-k retrieval over encrypted cloud data[J]. IEEE Transactions on Dependable and Secure Computing, 2013, 10(4): 239-250.

[27] MIRANDA-JIMENEZ S, GELBUKH A, SIDOROV G. Summarizing conceptual graphs for automatic summarization task[C]// Conceptual Structures for STEM Research and Education: 20th

International Conference on Conceptual Structures 2013. Mumbai: Springer, 2013: 245-253.
[28] POH G S, MOHAMAD M S, ZABA M R. Structured encryption for conceptual graphs[C]// International Workshop on Security. Berlin: Springer, 2012: 105-122.
[29] CHASE M, KAMARA S. Structured encryption and controlled disclosure[C]// Advances in Cryptology-ASIACRYPT 2010: 16th International Conference on the Theory and Application of Cryptology and Information Security. Singapore: Springer, 2010: 577-594.
[30] FU Z, HUANG F, SUN X, et al. Enabling semantic search based on conceptual graphs over encrypted outsourced data[J]. IEEE Transactions on Services Computing, 2016, 12(5): 813-823.
[31] 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.
[32] FU Z, SUN X, JI S, et al. Towards efficient content-aware search over encrypted outsourced data in cloud[C]// IEEE INFOCOM 2016 — The 35th Annual IEEE International Conference on Computer Communications. San Francisco: IEEE, 2016: 1-9.
[33] FU Z, XIA L, SUN X, et al. Semantic-aware searching over encrypted data for cloud computing[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(9): 2359-2371.
[34] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv: 1301.13781, 2013.
[35] LIU Y, FU Z. Secure search service based on word2vec in the public cloud[J]. International Journal of Computational Science and Engineering, 2019, 18(3): 305-313.
[36] ZHANG Y, WANG Y, LI Y. Searchable public key encryption supporting semantic multi-keywords search[J]. IEEE Access, 2019, 7: 122078-122090.
[37] CHEN L, XUE Y, MU Y, et al. CASE-SSE: Context-Aware Semantically Extensible Search-able Symmetric Encryption for Encrypted Cloud Data[J]. IEEE Transactions on Services Computing, 2022, 16(2): 1011-1022.
[38] ZHANG M, CHEN Y, HUANG J. SE-PPFM: A searchable encryption scheme supporting privacy-preserving fuzzy multikeyword in cloud systems[J]. IEEE Systems Journal, 2020, 15(2): 2980-2988.
[39] DAI X, DAI H, YANG G, et al. An efficient and dynamic semantic-aware multi keyword ranked search scheme over encrypted cloud data[J]. IEEE Access, 2019, 7: 142855-142865.
[40] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(Jan): 993-1022.
[41] DAI H, DAI X, YI X, et al. Semantic-aware Multi-keyword Ranked Search Scheme over Encrypted Cloud Data[J]. Journal of Network and Computer Applications, 2019, 147: 102442.
[42] DAI X, DAI H, RONG C, et al. Enhanced semantic-aware multikeyword ranked search scheme over encrypted cloud data[J]. IEEE Transactions on Cloud Computing, 2020, 10(4): 2595-2612.
[43] ZHOU Q, DAI H, HU Z, et al. Accuracy-first and efficiency-first privacy-preserving semantic-awa-re ranked searches in the cloud[J]. International Journal of Intelligent Systems, 2022, 37(11): 9213-9244.
[44] ZHOU Q, DAI H, HU Z, et al. SAPMS: A Semantic-Aware Privacy-Preserving Multi-keyword Search Scheme in Cloud[C]// Asia-Pacific Web(APWeb) and Web-Age Information Management(WAIM) Joint International Conference on Web and Big Data. Nanjing: Springer, 2022: 251-263.
[45] ZHOU Q, DAI H, LIU Y, et al. A novel semantic-aware search scheme based on BCI-tree index over encrypted cloud data[J]. World Wide Web, 2023, 6: 3055-3079.
[46] KENTON J D M W C, TOUTANOVA L K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of NAACL-HLT. Minneapolis: ACL, 2019: 4171-4186.
[47] FU Z, WANG Y, SUN X, et al. Semantic and secure search over encrypted outsourcing cloud based on BERT[J]. Frontiers of Computer Science, 2022, 16: 1-8.
[48] HU Z, DAI H, LIU Y, et al. CSMRS: An Efficient and Effective Semantic-aware Ranked Search Scheme over Encrypted Cloud Data[C]// 2022 IEEE 25th International Conf-erence on Computer Supported Cooperative Work in Design(CSCWD). Hangzhou: IEEE, 2022: 699-704.
[49] LANG B, WANG J, LI M, et al. Semantic-based compound keyword search over encrypted cloud data[J]. IEEE Transactions on Services Computing, 2018, 14(3): 850-863.
[50] YANG W, ZHU Y. A verifiable semantic searching scheme by optimal matching over encrypted data in public cloud[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 100-115.
[51] ZHAO L, CHEN L. A Linear Distinguisher and its Application for Analyzing Privacy-Preserving Transformation Used in Verifiable(Outsourced) Computation[C]// Proceedings of the 2018 on Asia Conference on Computer and Commu-nications Security. New York: Association for Computing Machinery, 2018: 253-260.



LIU Yuanlong, born in 2000, postgraduate. His main research interest is cloud computing security.

DAI Hua, born in 1982, Ph. D, professor, Ph. D supervisor, is a member of CCF(No. 40161M). His main research interests include cloud computing security and privacy protection.

(责任编辑:柯颖)