面向云环境的一致性可验证单关键词检索方法

戴 华[1,2] 保静静[1] 朱向洋[1] 易 训[3] 杨 庚[1,2]
(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理学院 墨尔本 3000)[3]

摘 要 在云环境资源外包服务模式下,数据拥有者不再参与对其外包数据的直接管理,这就使得验证数据使用者获得的检索结果是否满足一致性成为具有挑战性的问题。现有的研究工作重点聚焦于解决云服务提供商满足"诚实而好奇"模型假设下的隐私保护问题,但并不能解决恶意攻击威胁模型下的检索结果一致性验证问题。针对云服务提供商恶意攻击威胁模型,提出了一种面向云环境的基于偏序约束链的一致性可验证单关键词检索方法——IVSKS。数据拥有者根据文档与关键词的相关度的偏序关系,构造用于检索结果一致性验证的偏序约束链验证编码信息,并将该信息与文档集共同外包存储至云端;数据使用者在执行单关键词检索时,云端返回检索结果文档集以及相应的验证编码;最后,数据使用者根据获得的检索结果重构验证编码,实现针对检索结果的一致性验证。实验表明,与同类方法相比,IVSKS 在检索结果冗余度以及一致性验证时间开销上具有更好的表现。

关键词 云计算,关键词检索,一致性验证,Top-k,偏序约束链
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.12.014

Integrity-verifying Single Keyword Search Method in Clouds

DAI Hua[1,2] BAO Jing-jing[1] ZHU Xiang-yang[1] YI Xun[3] YANG Geng[1,2]
(College of Computer Science, 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 The service of outsourcing resources in clouds makes the data out of control from its owner and generates many security issues. It has become a serious threat to data users to verify the integrity of search results received from clouds. In the area of secure keyword search for cloud computing, current works mainly focus on the privacy-preserving issues which adopts the honest-but-curious threat model, but they are not able to solve the problem of integrity verification of the search result while the malicious attack threat model is adopted. This paper proposed a method of verifiable single keyword searching based on the partially ordered constraint chain, called IVSKS. According to the partial order relation of the relevance between keywords and files, data owner constructs the partial ordered constraint chains as verification objects of files, which are generated by hash functions. The verification objects and the corresponding files are subsequently outsourced together to clouds. When data users search the top-k relevance files by an interested keyword, the clouds will return the qualified files along with the corresponding verification objects. Data users can reconstruct the verification objects by these files and make a comparison to determine whether the result files are complete or correct. The experimental results indicate that IVSKS performs better on search result redundancy and completeness verification efficiency compared with the existing methods.

Keywords Cloud computing, Keyword search, Integrity verification, Top-k, Partial order constraint chain

到稿日期:2017-11-08 返修日期:2018-02-06 本文受国家自然科学基金项目(61872197,61572263,61672297,61472193),江苏省自然科学项目(BK20151511,BK20161516),中国博士后科学基金(2015M581794),江苏省高校自然科学研究项目(15KJB520027),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);保静静(1993—),女,硕士生,主要研究方向为数据管理与安全;朱向洋(1993—),男,硕士生,主要研究方向为隐私保护、密文检索;易 训(1967—),男,教授,博士生导师,主要研究方向为信息安全、分布式数据处理;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。

1　引言

随着云计算技术的成熟,以 Amazon EC2,Google App Engine 等为代表的云服务模式得到了快速发展并被广泛应用。越来越多的企业和个人将存储、计算等资源外包至云服务提供商(Cloud Service Provider,CSP),以降低运营成本。IT 资源服务化的思想日益普及,呈现“一切皆服务”(X as a Service,XaaS)的趋势,服务成为了云计算的核心概念。然而,在云计算蓬勃发展的同时,云安全[1]也成为被广泛关注的问题。在云计算环境中,数据拥有者(Data Owner,DO)失去了对放置在云服务器(Cloud Server,CS)中的数据和计算的控制,对数据是否受到保护、计算任务是否被正确执行都无法确定,因此需要设计相应的安全机制来保护其数据的机密性和一致性(Integrity)[2-5]。其中,前者的目标是保护数据隐私的不可窥探性;后者的目标则是要实现检索结果的一致性验证,即判断检索结果是否完整且正确。

目前,针对云环境中关键词检索的安全保护已经成为现有研究关注的热点问题。文献[6-13]采用“诚实而好奇”(Honest-but-Curious)模型[9],即假设 CS 能够严格按照约定的协议为客户提供数据检索服务,但存在窥探私密信息的意图。因此,检索过程以及检索结果的隐私保护问题是现有研究的重点。然而,CS 并不一定总是遵守“诚实而好奇”模型,甚至在某些情况下,拥有数据管理权限的 CS 可能成为恶意攻击(Malicious Attacks)的实施者,例如 CS 内部管理员的权限滥用(Authority Abuse)、成功入侵 CS 的黑客攻击等。这种恶意攻击行为可能造成返回给用户的检索结果不满足一致性要求(即检索结果不正确或不完整),进而影响建立在相关检索结果基础上的上层决策的正确性。当前,在云环境中,针对恶意攻击的一致性可验证关键词检索方法的研究尚处于起步阶段。

本文重点关注针对恶意攻击模型的云环境单关键词检索结果的一致性验证问题。事实上,在某些特定领域中,检索结果的一致性要求非常重要,而数据的私密性有时却不是必须的。例如,在证券交易应用中,任何一只股票的对价成交数据都是公开的,每个用户都可以查看这些数据;而能否返回符合特定检索模式要求的正确且完整的检索结果才是用户最关心的。随着越来越多的政府、企业和个人将自己的数据托管到 CS,由 CS 内部存在的恶意攻击而造成的检索结果的不一致性问题也引起了业界越来越多的关注,研究针对恶意攻击模型的云环境一致性可验证单关键词检索方法具有实际意义。

针对云环境中由于恶意攻击而导致的检索结果的一致性验证问题,本文提出了一种基于偏序约束链(Partial Order Constraint Chain,POCC)的一致性可验证单关键词检索方法(Integrity-Verifying Single Keyword Search,IVSKS)。在数据外包存储阶段,DO 在将文档外包存储至 CS 之前,根据文档与关键词的相关度的偏序关系,为每一个文档构造针对各个关键词的偏序约束链,该偏序约束链将作为验证编码(Verification Object,VO)用于检索结果的一致性验证;然后,再将文档和相应的偏序约束链一同外包存储至 CS。数据使用者(Data User,DU)向 CS 提交单关键词检索请求后,CS 根据检索关键词确定检索结果文档集合和唯一的偏序约束链,并将这些数据作为检索结果返回给 DU;DU 根据收到的检索结果重构偏序约束链,最终通过校验偏序约束链实现针对检索结果的一致性验证。基于上述思想,本文给出基于偏序约束链的数据上传协议和关键词检索及验证协议,并分析协议的安全性和性能。最后,基于真实数据集进行实验分析,实验结果表明,IVSKS 在一致性验证时间开销、检索结果冗余度方面具有明显优势。

2　相关工作

DO 将自身拥有的数据外包存储至 CS,由 CS 负责处理授权 DU 的检索请求。该外包数据服务模式存在两类安全威胁:1)数据脱离 DO 的直接管辖,委托给 CS 托管所带来的隐私泄露或隐私窥探问题;2)针对 CS 的外部或内部的恶意攻击行为(篡改、伪造等)所造成的 DO 获取检索结果一致性的可验证性问题。

目前,针对文档检索中隐私保护问题的研究是当前的一个重要研究方向,现有的研究工作大部分聚焦于此,重点解决具有隐私保护能力的关键词文档检索问题。其中,文献[14]首次提出了可搜索加密机制,通过加密文档中的每个关键词实现对密文的检索;随后出现的可搜索加密机制大多是基于索引的方案[15-17],应用布隆过滤器、倒排索引等技术提出安全的索引模型;文献[18]提出了有效的支持密文存储结构的范围查询方法;此外,针对云环境中密文数据的单关键词[9-10,19-20]和多关键词[7-8,11-13]的隐私保护搜索机制也得到了广泛的研究和关注。然而,现有的这些工作并不能解决由恶意攻击行为带来的检索结果的一致性可验证问题。

针对检索结果的一致性验证问题,不同研究者采用不同的威胁模型。在“诚实而好奇”威胁模型的研究背景下,文献[8]针对可验证的多关键词文档检索问题,提出在多维 B 树[21]的基础上通过哈希子树和签名技术实现检索结果的验证;文献[7]提出基于层次聚类加速 CS 对检索过程的处理,并应用 MerkleHash Tree 和密钥签名技术验证了检索结果的一致性。这两种方法所提出的一致性验证了解决方案并不能直接应对恶意攻击问题。在“部分诚实”(Partially Honest)威胁模型的研究背景下,文献[22]在一致性验证方面加强了威胁模型,提出通过同态消息认证编码和随机挑战技术实现多关键词检索结果的一致性验证,但该方式也不能用于解决 CS 内部对外包数据的恶意攻击行为。在恶意攻击威胁模型的研究背景下,文献[23-25]提出了支持一致性可验证的多关键词全文档布尔检索方法,即根据给定的检索关键词,并不区分文档与目标关键词的相关度,直接返回所有包含该目标关键词的全部文档,但该类方法由于未考虑文档与关键词的相关度问题,无法用于解决一致性可验证 Top-k 检索问题。文献[9]提出 CS 在提供服务时可能不遵循“诚实而好奇”模型,甚至可能进行主动攻击,并针对一致性验证问题给出了解决方案的初步构思,但没有给出具体的实现方法和相关实验验证。

本文采用的是更具有现实意义的恶意攻击威胁模型。在基于这种威胁模型的前提下,本文提出了一种基于偏序约束链的一致性可验证单关键词 Top-k 检索方案。在本方案中,

DU 向 CS 提交检索请求后，能够获得与检索关键词最相关的 k 个文档，并能准确验证 CS 返回的结果与 DO 外包的数据是否是一致的。

3 问题定义

3.1 系统模型

本文采用与文献[7-13]相同的系统模型(见图 1)。该模型主要包含 3 类实体，即数据拥有者(Data Owner，DO)、数据使用者(Data User，DU)和云服务器(Cloud Service，CS)，它们的协作方式如下：

(1)DO 负责原始文档的处理，并将处理后的数据(如文档、索引、验证信息等)上传存储到 CS；

(2)CS 为 DU 提供检索服务，即根据 DU 提交的指令进行检索处理，并返回符合要求的检索结果；

(3)DU 根据自身的应用需求定制检索指令，并将其提交至 CS，然后等待接收检索结果。

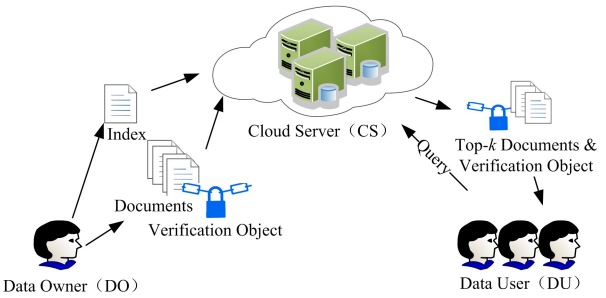

图 1 系统模型结构图

Fig. 1 Structure diagram of system model

3.2 单关键词检索模型

本文讨论的单关键词检索是一种以获得目标文档集合 D 中与被检索关键词 w_q 最相关的前 k 个文档为目的的处理方法，可以形式化为如下三元组：

$$Q=(D,w_q,k) \tag{1}$$

其中，D 为所有文档构成的集合，$D=\{d_1,d_2,\cdots,d_n\}$；w_q 为关键词集合 $W=\{w_1,w_2,\cdots,w_m\}$ 中的用户检索关键词；k 为检索需返回的最相关的文档数量，一般而言，在实际应用中 $k\ll n$。为方便讨论，我们假设对于任一关键词 $w_q\in W$ 而言，D 中包含该关键词的文档数量均不小于 k。

对于关键词 w_q 和文档 d_i，采用与文献[9-10,19-20]相同的方法计算 w_q 和 d_i 的相关度，记为 $RS(w_q,d_i)$：

$$RS(w_q,d_i)=\frac{1}{|d_i|}(1+\ln f_{q,i}) \tag{2}$$

其中，$|d_i|$ 是文档 d_i 中包含的关键词数量，$f_{q,i}$ 是 w_q 在 d_i 中出现的次数。

对于任意两个不同文档 d_i 和 d_j，如果考虑文档的 ID、创建时间等信息，能够确保 d_i 和 d_j 对于同一检索关键词的相关度不相等。因此，不妨设 D 中各文档对于任意一条关键词 w_q 满足：

$$\forall d_i,d_j\in D\wedge d_i\neq d_j\rightarrow RS(w_q,d_i)>RS(w_q,d_j)\vee RS(w_q,d_i)<RS(w_q,d_j) \tag{3}$$

基于该假设，易知对于任一检索指令 Q 而言，其检索结果具有唯一性。设检索结果文档集合记为 $D_{q,k}$，则：

$$|D_{q,k}|=k\wedge\forall d_i,d_j(d_i\in D_{q,k}\wedge d_j\in(D-D_{q,k}))\rightarrow RS(w_q,d_i)>RS(w_q,d_j) \tag{4}$$

3.3 威胁模型与问题描述

与现有研究工作中所采用的“诚实而好奇”威胁模型不同，本文采用安全要求更高的恶意攻击威胁模型，即假设 CS 对于 DO 不可信，CS 可能会对其存储的 DO 的外包数据进行恶意攻击(如篡改、伪造、丢弃等)，这些攻击行为可能是由 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一致性要求(不正确或不完整)，最终影响到依赖于检索结果的上层决策的正确性。本文重点关注由恶意攻击导致的检索结果不满足一致性要求的检测和验证问题，而对数据的私密性保护不是本文的研究重点。

与文献[8,23]相同，本文讨论的检索结果一致性的验证问题包含两方面：1)验证检索结果的正确性，即验证检索结果中是否存在数据被篡改；2)验证检索结果的完整性，即检查检索结果中的数据是否恰好满足检索条件的所有数据项。对于检索 $Q=(D,w_q,k)$，当且仅当检索结果 $D_{q,k}$ 同时满足如下两个条件时，$D_{q,k}$ 为满足一致性要求的检索结果。

(1)正确性要求：$D_{q,k}$ 中任何文档均未被篡改，即：

$$\forall d_i\in D_{q,k}\rightarrow d_i\in D \tag{5}$$

(2)完整性要求：D 中与 w_q 最相关的前 k 个文档都包含在 $D_{q,k}$ 中，没有遗漏，即严格满足式(4)的要求。

本文给出的针对检索结果的一致性验证过程正是基于上述条件的判断方法实现的，我们将在第 4 节给出具体的验证过程。此外，为了对检索方法的执行性能进行定量评估，我们采用如下度量指标。

(1)检索结果一致性验证时间开销，记为 δ_v，表示 DU 在接收到检索结果文档集后，完成检索结果验证的时间开销；

(2)检索结果冗余度，记为 δ_r，表示在 CS 的返回结果中，验证编码长度占返回数据总长度的比值，即：

$$\delta_r=\frac{l_c}{\sum_{i=1}^{k}l_i+l_c}\times 100\% \tag{6}$$

其中，l_c 是验证编码的长度，l_i 是检索结果集中文档 d_i 的长度。

4 一致性可验证单关键词检索方法

本节重点讨论一致性可验证单关键词检索方法的具体过程，主要包含基于偏序约束链的数据外包协议和关键词检索及验证协议。

4.1 基于偏序约束链的数据外包协议

为了实现针对检索结果的一致性验证，DO 在将数据外包至 CS 时，需同时上传用于验证检索结果一致性的编码信息，从而使得 DU 在获得检索结果时，能够利用该编码信息验证检索结果是否满足一致性要求，进而确认结果文档集是否可用。这里的编码信息即为偏序约束链，其构成方式如定义 1 所示。

定义 1(偏序约束链，Partial Order Constraint Chain，POCC) 设有文件 $d_1,d_2,\cdots,d_i$，对于关键词 w_q 满足 $RS(w_q,d_1)>RS(w_q,d_2)>\cdots>RS(w_q,d_i)$，则将 HMAC 的计算结果 $H_g(d_1\|d_2\|\cdots\|d_i\|w_q)$ 称为 d_i 关于 w_q 的偏序约

束链,记为 $h_{q,i}$。其中,$H_g(\cdot)$表示 Hash 身份认证编码算法(如 HMAC-SHA1 等),g 为密钥,由 DO 和 DU 共享,$\|$ 为连接运算符。

DO 发起的数据外包处理过程如协议 1 所示。

协议 1　基于 POCC 的数据外包协议

(1)设在 D 中包含关键字 w_q 的文档集合为$\{d_1,d_2,\cdots,d_v\}$,各文档与关键词 w_q 的相关度满足 $RS(w_q,d_1)>RS(w_q,d_2)>\cdots>RS(w_q,d_v)$,根据如下计算方法为文档$\{d_1,d_2,\cdots,d_v\}$构造关于 w_q 的偏序约束链 POCC,即 $h_{q,1}$,$h_{q,2},\cdots,h_{q,v}$。

$$h_{q,1}=H_g(d_1\|w_q)$$

$$h_{q,2}=H_g(d_1\|d_2\|w_q)$$

$$\cdots$$

$$h_{q,v}=H_g(d_1\|d_2\|\cdots\|d_v\|w_q)$$

(2)根据(1)中所述方法,对于关键词集合 $W=\{w_1,w_2,\cdots,w_m\}$中的每一个关键词,分别构造针对 D 中每一个文档的 POCC,并为文档集合 D 构建针对 W 的包含 POCC 信息的倒排索引[9],然后将文档和索引一起外包存储至 CS。

(3)CS 接收并存储 DO 发来的文档、索引和验证编码信息。

由协议 1 可知,对于 W 中的任一关键词 w_q,需要计算 D 中各文档与 w_q 的相关度,并进行排序,进而为每一个文档生成对应于 w_q 的验证编码 POCC;对于 D 中的任一文档 d_i,最多将生成 m 个 POCC,并与该文档一起外包存储于 CS 中,用于实现针对检索结果的一致性验证。因此,协议 1 中构造 POCC 的时间复杂度为 $O(m\cdot n\cdot\log_2 n)$。

4.2　关键词检索及验证协议

关键词检索及验证主要由 DU 和 CS 协作完成:DU 提交检索指令至 CS,然后 CS 根据指令要求执行检索处理过程,并返回相关结果,最后 DU 验证返回结果的一致性。具体过程如协议 2 所示。

协议 2　关键词检索及验证协议

(1)DU 将检索指令 $Q=(D,w_q,k)$提交至 CS,然后等待 CS 的返回结果。

(2)CS 接收到检索指令 Q 后,利用倒排索引查找包含 w_q 的文档,并通过计算与 w_q 的相关度,确定相关度最高的前 k 个文档,构成检索结果文档集 $D_{q,k}$,不妨设 $D_{q,k}=\{d_1,d_2,\cdots,d_k\}$,且满足 $RS(w_q,d_1)>RS(w_q,d_2)>\cdots>RS(w_q,d_k)$。然后,CS 将 $D_{q,k}$和唯一的 POCC 编码信息 $h_{q,k}$返回至 DU。

(3)DU 接收到 CS 返回的结果消息$\langle ds,h\rangle$,其中 ds 为检索结果文档集合,h 为 POCC 编码信息。DU 执行如下针对检索结果的验证过程:

1)检查 ds 中是否包含 k 个文档,若$|ds|=k$ 成立,则转验证步骤 2),否则 $D_{q,k}$即为不满足一致性要求的检索结果,检索失败退出;

2)设 $ds=\{d_1,d_2,\cdots,d_k\}$,对于 w_q,有 $RS(w_q,d_1)>RS(w_q,d_2)>\cdots>RS(w_q,d_k)$成立,此时 DU 根据与 DO 共享的密钥 g 重构 POCC 编码信息 $H_g(d_1\|d_2\|\cdots\|d_k\|w_q)$。若 $H_g(d_1\|d_2\|\cdots\|d_k\|w_q)=h$ 成立,则 ds 即为满足一致性要求的正确检索结果,检索成功退出;否则 ds 不满足一致性要求,检索失败退出。

由协议 2 可知,CS 根据 DU 的检索指令返回 k 个检索结果文档和 1 个 POCC 编码数据,DU 根据收到的检索结果文档重构 POCC,并检查是否与收到的 POCC 相一致,从而实现对检索结果一致性的验证。由于针对检索结果的验证过程只需重构和比较 1 个偏序约束链即可,因此协议 2 中实现检索结果一致性验证的渐进时间复杂度为 O(1)。

4.3　协议的安全性分析

在 IVSKS 中,POCC 是基于 HMAC(Hash Message Authentication Code)构造的验证编码。HMAC 的雪崩效应(Ava-lanche Effect)[26]以及单向不可逆性,确保了 POCC 的安全性。对于外部攻击,从最终的攻击效果角度来看,对检索结果的影响主要表现为两个方面:(1)检索结果的正确性;(2)检索结果的完整性。从这两种情况出发,对 IVSKS 协议进行安全性分析。

定理 1　IVSKS 协议能够验证检索结果的正确性,即能够发现检索结果中的数据是否被篡改。

证明:假设 DU 接收到 CS 返回的检索结果为$\langle ds,h\rangle$,$ds=\{d_1,d_2,\cdots,d_k\}$,满足 $RS(w_q,d_1)>RS(w_q,d_2)>\cdots>RS(w_q,d_k)$,$h$ 为验证编码 POCC。若其中 d_i 的内容被篡改,不妨设篡改后的文档为 d_i'。DU 接收到检索文档集后,重构 POCC 编码 $h'=H_g(d_1\|d_2\|\cdots\|d_i'\|\cdots\|d_k\|w_q)$,显然 $h'\neq h$,此时 DU 根据协议 2 步骤(3)中的 2)即可判断 ds 不满足一致性要求。同理,若验证编码 h 被篡改,在重构和比对验证编码的过程中,DU 也将检测到 ds 不满足一致性要求。因此,本文提出的协议能够验证检索结果的正确性。

定理 2　IVSKS 协议能够验证检索结果的完整性,即能够验证检索结果中的所有数据是否恰好是满足检索条件的所有数据。

证明:外部攻击导致的检索结果不满足完整性主要有两种情况:(1)返回的文档个数小于 k,即$|D_{q,k}|<k$,根据协议 2 步骤(3)中的 1)可知,该情况下返回的检索结果文档集显然不满足一致性要求;(2)返回的检索结果文档集不是与 w_q 最相关的前 k 个文档,即丢弃了检索结果中与 w_q 更相关的文档,而用其他文档替代。假设 CS 应返回的正确检索结果为$\langle ds,h\rangle$,若 CS 丢弃 ds 中的一个或多个文档,而用其他文档替代,不妨设替代后的检索结果为 ds',$ds\neq ds'$。根据 HMAC 的雪崩效应机制可知,ds 变化为 ds',必然将导致基于 ds'重构出来的 POCC 编码 h'与 h 不相等,此时根据协议 2 步骤(3)中的 2)即可判断出 ds'是不满足一致性要求的检索结果。此外,由于 POCC 采用带密钥的 HMAC 构造,且密钥仅由 DU 和 DO 共享,而对 CS 保密,因此,CS 无法在替代 ds 中的部分文档之后成功伪造出与之相匹配的正确的 POCC,这也使得 CS 无法成功实施丢弃攻击而不被发现。因此,本文提出的协议能够验证检索结果的完整性。

综合上述定理可知,本文提出的 IVSKS 方法能够从检测检索结果是否满足正确性和完整性这两个方面实现针对检索结果的一致性验证。

4.4　协议实例说明

为了更清晰地描述协议的执行过程,结合图 2 所示的实

例进行描述。设文档集合 $D=\{d_1,d_2,d_3,d_4\}$，关键词集合 $W=\{w_1,w_2\}$，对于 w_1，有 $RS(w_1,d_1)>RS(w_1,d_2)>RS(w_1,d_3)>RS(w_1,d_4)$ 成立，对于 w_2，有 $RS(w_2,d_2)>RS(w_2,d_3)>RS(w_2,d_1)>RS(w_2,d_4)$ 成立。根据协议 1，对于 w_1 而言，d_1，d_2，d_3 和 d_4 对应的 POCC 分别为 $H_g(d_1\|w_1)$，$H_g(d_1\|d_2\|w_1)$，$H_g(d_1\|d_2\|d_3\|w_1)$ 和 $H_g(d_1\|d_2\|d_3\|d_4\|w_1)$；对于 w_2 而言，d_1，d_2，d_3 和 d_4 对应的 POCC 则分别为 $H_g(d_2\|d_3\|d_1\|w_2)$，$H_g(d_2\|w_2)$，$H_g(d_2\|d_3\|w_2)$ 和 $H_g(d_2\|d_3\|d_1\|d_4\|w_2)$。在执行检索和验证时，假设 DU 向 CS 发送检索请求 $(D,w_1,3)$，则 CS 应返回检索结果文档集 $\{d_1,d_2,d_3\}$ 和唯一的验证编码 $H_g(d_1\|d_2\|d_3\|w_1)$；DU 接收到 CS 返回的结果，通过执行协议 2 中步骤(3)的验证过程即可判定检索结果是否满足一致性要求。

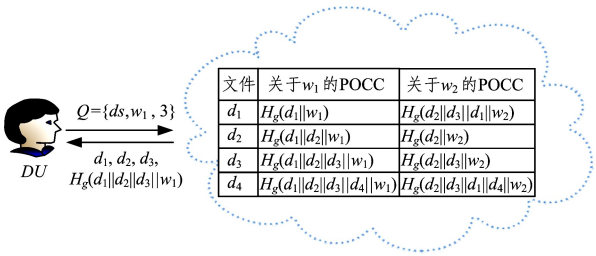

图 2 检索实例

Fig. 2 Example of search

针对上述实例，对于检索结果而言，本文提出的 IVSKS 只需返回 3 个文档和 1 个验证编码，而文献[9]提出的 SERKS 方案则需要返回 3 个文档和 3 个验证编码，由此可见 IVSKS 的检索结果的传输代价小于 SERKS，同时检索结果的冗余度也低于后者。对于检索结果的验证过程而言，IVSKS 只需重构 1 个验证编码即可完成整个验证过程，而 SERKS 则需要重构 3 个验证编码，因此，IVSKS 完成检索结果验证的时间消耗也低于 SERKS。

此外，当外包文档数据量较大时，通过如下扩展即可实现本文协议的 Map-Reduce 并行化处理，从而充分发挥云环境的优势特点。在数据处理阶段，首先对原始文档集进行分片处理，然后将每个分片交给一个 MapTask 进行处理，并将处理后的数据外包存储至 CS 端。在关键词检索与验证阶段，DU 将检索指令 $Q=(D,w_q,k)$ 提交至 CS，启动并行化处理过程：在 Map 阶段，每个 MapTask 根据其获得的分片计算与检索目标关键词 w_q 最相关的前 k 个文档以及对应的验证编码；在 Reduce 阶段，ReduceTask 将所有 MapTask 返回的信息进行融合处理，并最终确定最相关的前 k 个文档和相应验证编码，并将其作为检索结果返回给 DU；最后，DU 利用相应的验证编码对检索结果文档进行一致性验证。

5 实验和结果分析

为了对协议的性能进行比较和分析，分别实现了本文提出的 IVSKS 和文献[9]中给出的 SERKS 方案，并从检索结果一致性验证的时间开销和检索结果冗余度两个指标上进行对比实验分析。实验采用的测试数据集为 UCI 的 NSF Research Award Abstracts[27]；实验硬件环境为 3.4 GHz 主频 4 核 CPU，16 GB 内存，1 TB 硬盘；软件环境为 64 位的 CentOS 操作系统，Java 开发工具。

5.1 检索结果验证的时间开销

当 DU 检索文档数量 k 变化时，IVSKS 和 SERKS 的检索结果一致性验证的时间开销 δ_v 的变化情况如图 3 所示。

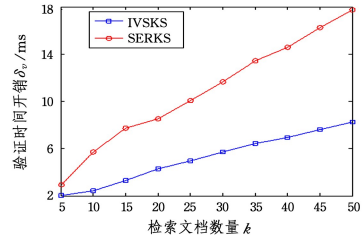

图 3 k 对检索结果一致性验证时间开销 δ_v 的影响

Fig. 3 Influence of k on δ_v

由图 3 可知，IVSKS 和 SERKS 的 δ_v 都随着检索文档数量 k 的增长而增长，但 SERKS 的增长速度显著快于 IVSKS。在实验设置的条件下，SERKS 的 δ_v 平均比 IVSKS 的 δ_v 高 103.24%。主要原因是：在 IVSKS 中，DU 在接收到 CS 返回的检索结果文档集上，重构 POCC 的次数要显著少于 SERKS，IVSKS 只需要重构 1 个 POCC，而 SERKS 需要重构 k 个 POCC。

5.2 检索结果的冗余度

(1)当 DU 检索文档数量 k 变化时，IVSKS 和 SERKS 的检索结果中验证编码长度 l_v 的变化情况如图 4 所示。

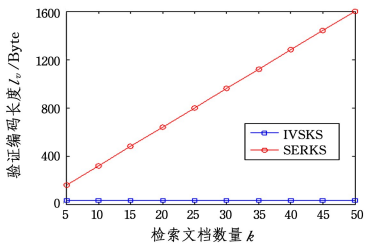

图 4 k 对验证编码总长度 l_v 的影响

Fig. 4 Influence of k on l_v

由图 4 可知，SERKS 的检索结果中验证编码长度 l_v 与检索文档数量 k 成正比例线性关系，即 l_v 随着 k 的增大而快速增长；而 IVSKS 的检索结果中验证编码长度与 k 无关，在 k 变化的过程中始终保持不变，且远小于 SERKS 的验证编码的长度，前者约为后者的 $1/k$，其原因与 5.1 节类似。

(2)当 DU 检索文档数量 k 变化时，IVSKS 和 SERKS 的检索结果的冗余度 δ_r 的变化情况如图 5 所示。

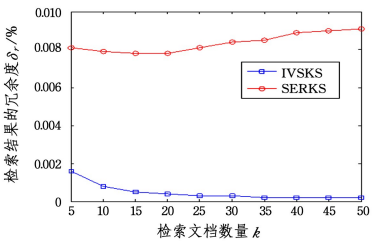

图 5 k 对检索结果冗余度 δ_r 的影响

Fig. 5 Influence of k on δ_r

由图 5 可知，IVSKS 的 δ_r 随着 k 的增加而逐步减小，并趋近于 0；而 SERKS 的 δ_r 变化不明显，在 0.8%附近震荡，但

显著高于 IVSKS。在实验设置的条件下,SERKS 的 δ_r 平均是 IVSKS 的 29.17 倍,其原因在于:SERKS 根据文档与关键词的相关度进行排序后对相邻文档进行了哈希计算,从而生成验证编码,当 DU 检索的目标文档数量为 k 时,CS 返回 k 个验证编码,而 IVSKS 只需要返回 1 个验证编码。由式(6)易知,δ_r 与返回文档的长度和验证编码的长度密切相关,因此,在短文本应用场景中,本文提出的 IVSKS 在检索结果的冗余度指标上的优势也将更加突出。

结束语　随着云计算外包服务模式的广泛应用,针对脱离数据拥有者直接管辖范围的外包数据的安全保护问题受到越来越多的关注。在面向外包数据的关键词检索研究中,如何验证检索结果的一致性是一个亟待解决的关键性安全问题。本文提出了一种面向云环境的一致性可验证单关键词检索方法——IVSKS。该方法利用文档与关键词间相关度得分的偏序关系构造偏序约束链作为验证编码,并与对应的文档数据一起外包存储至云端;在执行关键词检索时,通过返回特定的验证编码信息使得数据使用者能够判断其获得的检索结果是否满足一致性要求。实验结果表明,与现有的方法相比,本文提出的 IVSKS 方法在检索结果冗余度和一致性验证过程的时间开销方面具有显著的优势,更适用于云计算外包服务模式中存在恶意攻击行为且对文档检索结果一致性要求较高的应用场景。在后续的工作中,我们将重点研究协议的并行化方法和一致性可验证多关键词检索方法。

参 考 文 献

[1] DING Y,WANG H M,SHI P C,et al. Trusted Cloud Service [J]. Chinese Journal of Computers,2015,38(1):133-149. (in Chinese)
丁滟,王怀民,史佩昌,等. 可信云服务[J]. 计算机学报,2015,38(1):133-149.

[2] ZHANG M,HONG C,CHEN C. Server Transparent Query Authentication of Outsourced Database [J]. Journal of Computer Research and Development,2010,47(1):182-190. (in Chinese)
张敏,洪澄,陈驰. 一种服务器透明的外包数据库查询验证方法[J]. 计算机研究与发展,2010,47(1):182-190.

[3] ARORA R,PARASHAR A. Secure User Data in Cloud Computing Using Encryption Algorithms [J]. International Journal of Engineering Research and Applications,2013,3(4):1922-1926.

[4] WANG Y D,YANG J H,XU C,et al. Survey on access control technologies for cloud computing [J]. Journal of Software,2015,26(5):1129-1150. (in Chinese)
王于丁,杨家海,徐聪,等. 云计算访问控制技术研究综述[J]. 软件学报,2015,26(5):1129-1150.

[5] TIAN X X,WANG X L,GAO M,et al. Database as a services Security and privacy preserving[J]. Journal of Software,2010,21(5):991-1006. (in Chinese)
田秀霞,王晓玲,高明,等. 数据库服务——安全与隐私保护[J]. 软件学报,2010,21(5):991-1006.

[6] FU W,YAN B,WU X P. Data Possession Provability on Semi-trusted Cloud Storage[C]// Cloud Computing－4th International Conference. 2013:199-209.

[7] 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.

[8] SUN W H,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.

[9] WANG C,CAO N,REN K,et al. Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data [J]. IEEE Transactions on Parallel and Distributed Systems,2012,23(8):1467-1479.

[10] WANG C,CAO N,LI J,et al. Secure Ranked Keyword Search over Encrypted Cloud Data[C]// 2010 International Conference on Distributed Computing Systems. 2010:253-262.

[11] WANG D S,FU S J,XU M. A Privacy-Preserving Fuzzy Keyword Search Scheme over Encrypted Cloud Data[C]// IEEE 5th International Conference on Cloud Computing Technology and Science. 2013:663-670.

[12] CAO N,WANG C,LI M,et al. Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data [J]. IEEE Transactions on Parallel & Distributed Systems,2014,25(1):222-233.

[13] NA H Y,YANG G,SHU X W. Multi-keyword Ranked Search Method Based on B+ Tree [J]. Computer Science,2017,44(1):149-154. (in Chinese)
那海洋,杨庚,束晓伟. 基于 B+树的多关键字密文排序检索方法[J]. 计算机科学,2017,44(1):149-154.

[14] SONG D,WAGNER D,PERRIG A. Practical Techniques for Searches on Encrypted Data[C]// 2000 IEEE Symposium on Security and Privacy. 2000:44-55.

[15] GOH E J. Secure Indexes[OL]. http://www.researchgate.net.publication/2889193_Secure_Indexes.

[16] CHANG Y C,MITZENMACHER M. Privacy preserving keyword searches on remote encrypted data [C] // International Conference on Applied Cryptography and Network Security. Springer Berlin Heidelberg,2005:442-455.

[17] CURTMOLA R,GARAY J,KAMARA S,et al. Searchable symmetric encryption: improved definitions and efficient constructions[C]// 13th ACM Conference on Computer and Communications Security. ACM,2006:79-88.

[18] HORE B,MEHROTRA S,CANIM M,et al. Secure multidimensional range queries over outsourced data[J]. The International Journal on Very Large Data Bases,2012,21(3):333-358.

[19] LI J G,TIAN X X,ZHOU A Y. 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. 面向 DaaS 保护隐私的模糊关键字查询[J]. 计算机学报,2016,39(2):414-428.

[20] YANG C,YANG S L,KE M. 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. 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.

[21] SCHEUERMANN P,OUKSEL A M. Multidimensional B-trees for associative searching in database systems[J]. Information Systems,1982,7(2):123-137.

(下转第 110 页)

数据包接收位图中的数据在保证链路质量不变的前提下进行篡改达到对源节点的邻居节点之间的相关性进行欺骗攻击的目的。针对这一攻击算法,本文还设计并提出了基于 Watchdog 机制的恶意节点检测机制,利用节点的转发行为得出真实的数据包接收位图。仿真实验的结果表明了链路相关性欺骗攻击算法的有效性以及基于 Watchdog 的恶意节点检测机制对链路相关性欺骗攻击算法的保护效果。

参 考 文 献

[1] SONG M K,WANG S,HE T. Exploiting causes and effects of wireless link correlation for better performance [C]// Computer Communications. IEEE,2015:379-387.

[2] WANG S,KIM S M,LIU Y H,et al. CorLayer:A transparent link correlation layer for energy efficient broadcast [J]. In MobiCom,2013,23(6):1970-1983.

[3] SINGH G. A Survey of Multicast Routing Protocols in MANETS[C]// International Conference on Futuristic Trends in Engineering & Management. 2014.

[4] WANG S,BASALAMAH A,SONG M K,et al. Link-Correlation-Aware Opportunistic Routing in Wireless Networks [J]. IEEE Transactions on Wireless Communications,2015,14(1):47-56.

[5] WANG H,LIU Y,XU S. An Opportunistic Routing Protocol Based on Link Correlation for Wireless Mesh Networks [M]// Wireless Communications, Networking and Applications. Springer India,2016.

[6] SALEHI M,BOUKERCHE A,DAREHSHOORZADEH A. Modeling and Performance Evaluation of Security Attacks on Opportunistic Routing Protocols for Multihop Wireless Networks [J]. Ad Hoc Networks,2016,50(C):88-101.

[7] LI L,RAMJEE R,BUDDHIKOT M,et al. Network Coding-Based Broadcast in Mobile Ad-hoc Networks[J]. Proceedings - IEEE INFOCOM,2007,5(8):1739-1747.

[8] WANG S,KIM S M,YIN Z,et al. Encode When Necessary:Correlated Network Coding Under Unreliable Wireless Links [J]. Acm Transactions on Sensor Networks,2017,13(1):1-22.

[9] ZHU T,ZHONG Z,HE T,et al. Exploring link correlation for efficient flooding in wireless sensor networks[C]// Proceedings of the First USENIX/ACM Symposium on Network Systems Design and Implementation (NSDI). 2010.

[10] WANG S,BASALAMAH A,SONG M K,et al. A Unified Metric for Correlated Diversity in Wireless Networks [J]. IEEE Transactions on Wireless Communications,2016,15(9):6215-6227.

[11] SRINIVASAN K,JAIN M,CHOI J I,et al. The κ-factor:Inferring protocol performance using inter-link reception correlation [C]// Proceedings of ACM MOBICOM. 2010.

[12] ZHAO Z,DONG W,BU J,et al. Exploiting link correlation for core-based dissemination in wireless sensor networks [C]// Eleventh IEEE International Conference on Sensing,Communication,and Networking. IEEE,2014:372-380.

[13] TIAN B. Attack Detection and Defense Technologies in Wireless Sensor Networks[D]. Beijing:Beijing University of Posts and Telecommunications,2012. (in Chinese)
田斌. 无线传感器网络中攻击检测和防御技术研究[D]. 北京:北京邮电大学,2012.

[14] CHO Y,QU G. Detection and Prevention of Selective Forwarding-Based Denial-of-Service Attacks in WSNs [J]. International Journal of Distributed Sensor Networks,2013,2013(3):264-273.

[15] BORA S,SINGH S,MOHAMAD A S,et al. Watchdog:A Study on Examining and Eliminating Misbehaviour[J]. International Journal of Computer Applications,2014,87(3):1-3.

[16] AGRAWAL S,JAIN S,SHARMA S. A Survey of Routing Attacks and Security Measures in Mobile Ad-Hoc Networks [J]. 2011,4(1):41-48.

[17] MACKE J H,BERENS P,ECKER A S,et al. Generating spike trains with specified correlation coefficients [J]. Neural Computation,2009,21(2):397-423.

[18] HIGHAM N J. Computing the near est correlation matrix—a problem from finance [J]. Ima Journal of Numerical Analysis,2002,22(3):329-343.

[19] HEINZELMAN W B,CHANDRAKASAN A P,BALAKRISHNAN H. An Application Specific Protocol Architecture for Wireless Microsensor Networks[C]// IEEE Transactions on Wireless Communication. 2002:660-670.

(上接第 97 页)

[22] WAN Z G,DENG R H. VPSearch:Achieving Verifiability for Privacy-Preserving Multi-Keyword Search over Encrypted Cloud Data[J]. IEEE Transactions on Dependable & Secure Computing,2016,PP(99):1-1.

[23] ZHU X Y,HAO R P,JIANG S R,et al. Verification of Boolean Queries over OutsourcedEncrypted Data Based on Counting Bloom Filter[C]// IEEE Global Communications Conference. 2015:1-6.

[24] SUN W H,LIU X F,LOU W J,et al. Catch You If You Lie to Me:Efficient Verifiable Conjunctive Keyword Search over Large Dynamic Encrypted Cloud Data[C]// IEEE Conference on Computer Communications. 2015:2110-2118.

[25] JIANG S R,ZHU X Y,GUO L K,et al. Publicly Verifiable Boolean Query Over Outsourced Encrypted Data[C]// IEEE Global Communications Conference. 2016:1-6.

[26] PALLIPAMU V,REDDY K,VARMA P. ASH-160:A novel algorithm for secure hashing using geometric concepts [J]. Journal of Information Security and Applications,2014,21:52-63.

[27] LICHMAN M. UCI Machine Learning Repository [OL]. http://archive.ics.uci.edu/ml.