可验证对称加密搜索问题分析与研究

（申请清华大学工程硕士专业学位论文）

培 养 单 位: 计 算 机 科 学 与 技 术 系
学　　科: 计 算 机 技 术
申　请　人: 朱　洁
指 导 教 师: 李　琦　副　研　究　员

二〇一八年四月

Analysis and Research of Verifiable Searchable Symmetric Encryption

Thesis Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the professional degree of

Master of Engineering

by

Zhu Jie

(Computer Technology)

Thesis Supervisor : Professor Li Qi

April, 2018

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

（保密的论文在解密后应遵守此规定）

作者签名：_____　　导师签名：_____

日　　期：_____　　日　　期：_____

摘 要

云存储的发展使得用户可以方便地存储、获取与分享数据。但与此同时，云存储也带来了很多安全问题，例如，数据隐私泄露等等。对称加密搜索的提出解决了数据隐私泄露问题，同时也保证了数据的可搜索性。通过使用对称加密搜索方案，用户可以在上传数据到云服务器之前，对数据进行加密，同时云服务器可以在用户的加密数据上进行搜索，从而确保数据隐私。

然而，对称加密搜索的假定是云服务器是诚实且好奇的，即云服务器会遵守协议，但现实情况中云服务器往往是不可靠的。为了解决该问题，可验证对称加密搜索技术相应提出，它通过结果验证技术可以到检测云服务器的恶意行为。但是，现有的可验证对称加密搜索方案都不完善，例如，不支持用户数据动态更新，依赖于特定的对称加密搜索方案，只支持单用户读写等等。

针对以上的问题，本文提出了一种通用的可验证对称加密搜索框架，该框架普适于任何加密搜索方案，支持用户数据更新，并且能够同时在单用户和多用户的场景下工作。本文的主要工作和创新点包括：

- 提出了一种单用户场景下的可验证对称加密搜索框架，并在此基础上设计了结果验证算法，该算法能同时保证数据新鲜性和数据完整性。该框架支持用户数据动态更新，并且将验证索引从对称加密搜索方案中解耦，使其可以为任何加密搜索方案提供结果验证功能。
- 提出了一种多用户场景下的可验证对称加密搜索框架。该框架支持单用户写，多用户读，确保了多用户场景下的数据新鲜性，并实现了数据共享场景下的结果验证。
- 本文采用了一个开源数据集作为测试数据，在本地环境对该框架进行了实验测试。安全性分析和实验表明，本文提出的可验证对称加密搜索框架不泄露数据隐私，并且给对称加密搜索方案引入的额外计算开销和通信开销很小，几乎可以忽略不计。

关键词：对称加密搜索，结果验证，云存储

Abstract

Cloud storage allows users to retrieve and share their data conveniently. Meanwhile, cloud storage also brings serious data privacy issues, i.e., the disclosure of private information. In order to ensure data privacy without losing data usability, Searchable Symmetric Encryption (SSE) has been proposed. By using SSE, users can encrypt their data before uploading to cloud services, and cloud services can directly operate and search over encrypted data, which ensures data privacy.

However, most SSE schemes only work with honest-but-curious cloud services that do not deviate from the prescribed protocols. This assumption does not always hold in practice due to the untrusted nature in storage outsourcing. To alleviate the issue, there have been studies on Verifiable Searchable Symmetric Encryption (VSSE), which functions against malicious cloud services by enabling results verification. But to our best knowledge, existing VSSE schemes exhibit very limited applicability, such as only supporting static database, demanding specific SSE constructions, or only working in the single-user model.

In this paper, we proposed a generic verifiable SSE framework in both single-user model and single-writer multiple-reader model, which provides verifiability for any SSE schemes and further supports data updates. In summary, our contributions are three-fold:

- We proposed a verifiable SSE framework in single-user model and designed the result verification algorithms. The framework seperate the verification index from the SSE construction and can provides generic verification for any SSE schemes. The algorithms guaranteed both data freshness and data integrity with support of data updates.
- We proposed the first verifiable SSE framework in the single-writer and multiple-reader model, which ensures data freshness across multiple users and provides result verification under data sharing scenario.
- We implemented our framework in a local enviroment and fed it with an open-source data set. Rigorous analysis and experimental evaluations show that our shceme is secure and introduces small overhead for result verification.

Key words: Searchable Symmetric Encryption; Result Verification; Cloud Storage

目 录

主要符号对照表

SSE	加密搜索 (Searchable Symmetric Encryption)
VSSE	可验证加密搜索 (Verifiable Searchable Symmetric Encryption)
MPT	默克尔帕特里夏树 (Merkle Patricia Tree)
IH	增量哈希 (Incremental Hash)
$\mathcal{W}$	关键字集合
$\|W\|$	关键字集合大小
w_i	关键字，其中 $i \in \{1, \cdots, \|W\|\}$
$\mathcal{D}$	明文文件集合
D_{w_i}	包含关键字 w_i 的明文文件集合
C	密文文件集合
C_{w_i}	包含关键字 w_i 的密文文件集合
d	明文文件
c	密文文件
W_d	文件 d 包含的关键字集合
τ	搜索令牌
λ	验证索引
π	鉴别符

第 1 章　绪论

1.1　加密搜索的研究背景

云存储使得用户可以随时随地地存取数据，并且极大地方便了用户之间的数据共享。但与此同时，云存储带来了许多安全性问题，总体来说可以分为以下两类：（1）可用性（availability）。要求云服务器保证数据不丢失，用户可以将云端作为数据中枢进行数据备份和同步。目前，一般的云服务提供商都采用了多副本的方式保障数据的可用性，即将数据的多个副本分别写入其他的存储节点，当一个节点发生故障时，其他节点上的数据继续提供服务，同时通过其他节点中的数据副本，快速恢复故障节点上丢失的数据。目前，针对数据可用性的相关学术研究包括数据拥有证明（Proof of Data Possession, PDP）以及数据可恢复性证明（Proof of Retrievability, PoR）。（2）隐私性（privacy）。要求云服务器保证数据的隐私并且不泄露数据。目前，云服务提供商一般采用数据加密方式对隐私数据进行保护，但数据加密往往会导致数据可用性的降低，例如数据失去可搜索性，因此加密搜索（Searchable Encryption）应运而生。加密搜索技术主要分为两类，一是对称加密搜索（Searchable Symmetric Encryption, SSE），二是非对称加密搜索（Searchable Asymmetric Encryption，SAE）。由于非对称加密搜索的效率问题，我们在这里主要关注对称加密搜索。

对称加密搜索的模型如图 1 所示。用户自行对数据进行加密并上传到云端，与此同时，用户还需额外上传一个加密索引（index）使得云可以通过该索引来搜索数据。当用户需要搜索数据时，生成一个陷门（trapdoor），该陷门与关键字相关，使得用户可以在不暴露关键字内容的情况下进行内容搜索。加密搜索使得用户在保护数据隐私的同时，满足了其搜索需求，但加密搜索并不能保证搜索结果的正确性。也就是说，加密搜索的前提是云服务器是诚实的，即服务器会遵守与用户的协议来正确的执行搜索操作，然而实际应用中，云服务器往往是不可信的，例如，云服务器有可能为了节省计算开销和通信开销而返回少量搜索结果给用户，甚至有可能不返回搜索结果给用户。为了防止云服务器的不诚信行为，学术界又提出了可验证的对称加密搜索机制（Verifiable Searchable Symmetric Encryption, VSSE）。可验证的加密搜索允许用户对搜索结果进行验证，来检测服务器的不诚信行为，保障了加密搜索的正确性。

1.2　加密搜索的选题意义

在可验证加密搜索中，由于服务器不诚信导致的安全性攻击主要可以分为以下两种：重放攻击（Replay Attack）：在加密搜索中，重放攻击是指服务器（攻击者）试图返回旧的搜索结果，而不是最新的搜索结果。我们用 $_{n} = _{1,2,,n(}n +$ $1)_{i}1inDataIntegrityAttackF()G()G()F()G()$[5] – [12]

1.3　本文的主要内容

1.4　本文的结构安排

第 2 章　相关研究综述

现有的 CS2 论文 [5] 提出了搜索鉴别符方案，该方案利用了默克尔树（Merkle Hash Tree，MHT）来实现了对数据完整性的验证，然而该方案并没有考虑到服务器返回空结果来规避结果验证的情况。Kurosawa 等人提出了一系列可验证加密搜索方案 [6][8][12]，然而他们的方案要么在搜索效率上接近 O(n)，不适用于大型数据库，要么只支持静态数据库。除了 Kurosawa 等人的方案，其他还有许多只支持静态数据库的方案 [7][10][13]。虽然 Stefanov 等人的方案 [9] 通过消息验证码（Message Authenticated Code，MAC）实现了动态数据库下的可验证搜索，然而他们的方案仍然没有考虑到服务器返回空结果的情况。Bost 等人提出的方案是目前为止最完善的 [11]，但他们的方案在搜索时需要和服务器有两轮通信才能进行验证，无法并行进行验证，并且他们的方案同样也不支持多用户情况下的验证。各个方案的对比情况如表格 1 所示。表格 1 现有的可验证加密搜索方案比较动态性多用户防御重放攻击防御数据完整性攻击验证效率 KPR11[5] Œ Œ O(|W|) KO12[6] Œ Œ Œ Œ O(n) CG12[7] Œ Œ Œ O(log|W|) KO13[8] Œ Œ O(n) KW13[13] Œ Œ O(|W|) SPS14[9] Œ Œ O(rlog^3(n))BFP16[11]$O(r)OK$16[12]$O(1)OURSO(log|W|)$

2.2 可验证的非对称加密搜索第一个可验证的非对称加密搜索方案 [18] 由 Zheng 等人提出，他们的方案采用了基于属性的关键字（Attribute-based keyword，ABK），但是他们的方案也只适用于数据库静态的情况。基于他们的工作，Liu 等人又提出了一个更高效的可验证非对称加密搜索方案 [19]，然而，由于非对称加密本身的限制，他们的方案必不可少地需要引入一个可信第三方。2.3 不可验证的多用户加密搜索 Curtmola 等人在 2006 年即提出了一个基于广播加密的多用户加密搜索方案 [2]，该方案允许数据拥有者将数据分享给其他用户，并且数据拥有者具有对用户的访问控制权限，可以随时撤销或者新增用户。Jerecki 等人随后又提出了一个基于 Oblivious Cross Tag 的加密搜索方案 [21]，然而该方案需要数据拥有者和数据用户频繁的交互。2.4 总结综上所述，现有的可验证加密搜索方案都不能满足多用户场景下的安全性保证，并且现有的方案无法完善地解决重放攻击和数据完整性攻击。这需要我们设计合理的机制来防御多用户场景下重放攻击，并且需要我们利用新型的数据结构来完善对数据完整性攻击的防御，尤其是防御服务器返回空结果来规避结果验证的情况。

第 3 章　单用户下的可验证对称加密搜索方案研究

第 4 章　多用户下的可验证对称加密搜索方案研究

第 5 章　总结与展望

致　谢

衷心感谢导师 xxx 教授和物理系 xxx 副教授对本人的精心指导。他们的言传身教将使我终生受益。

在美国麻省理工学院化学系进行九个月的合作研究期间，承蒙 xxx 教授热心指导与帮助，不胜感激。感谢 xx 实验室主任 xx 教授，以及实验室全体老师和同学们的热情帮助和支持！本课题承蒙国家自然科学基金资助，特此致谢。

感谢 LaTeX 和 ThuThesis[?]，帮我节省了不少时间。

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____ 日　期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

xxxx 年 xx 月 xx 日出生于 xx 省 xx 县。

xxxx 年 9 月考入 xx 大学 xx 系 xx 专业，xxxx 年 7 月本科毕业并获得 xx 学士学位。

xxxx 年 9 月免试进入 xx 大学 xx 系攻读 xx 学位至今。

发表的学术论文

[1] Yang Y, Ren T L, Zhang L T, et al. Miniature microphone with silicon- based ferroelectric thin films. Integrated Ferroelectrics, 2003, 52:229-235. (SCI 收录, 检索号:758FZ.)

[2] 杨轶, 张宁欣, 任天令, 等. 硅基铁电微声学器件中薄膜残余应力的研究. 中国机械工程, 2005, 16(14):1289-1291. (EI 收录, 检索号:0534931 2907.)

[3] 杨轶, 张宁欣, 任天令, 等. 集成铁电器件中的关键工艺研究. 仪器仪表学报, 2003, 24(S4):192-193. (EI 源刊.)

[4] Yang Y, Ren T L, Zhu Y P, et al. PMUTs for handwriting recognition. In press. (已被 Integrated Ferroelectrics 录用. SCI 源刊.)

[5] Wu X M, Yang Y, Cai J, et al. Measurements of ferroelectric MEMS microphones. Integrated Ferroelectrics, 2005, 69:417-429. (SCI 收录, 检索号:896KM)

[6] 贾泽, 杨轶, 陈兢, 等. 用于压电和电容微麦克风的体硅腐蚀相关研究. 压电与声光, 2006, 28(1):117-119. (EI 收录, 检索号:06129773469)

[7] 伍晓明, 杨轶, 张宁欣, 等. 基于 MEMS 技术的集成铁电硅微麦克风. 中国集成电路, 2003, 53:59-61.

研究成果

[1] 任天令, 杨轶, 朱一平, 等. 硅基铁电微声学传感器畴极化区域控制和电极连接的方法: 中国, CN1602118A. (中国专利公开号)

[2] Ren T L, Yang Y, Zhu Y P, et al. Piezoelectric micro acoustic sensor based on ferroelectric materials: USA, No.11/215, 102. (美国发明专利申请号)