Ministry of Higher
Education and Scientific Research
University of Diyala
College of Science
Department of Computer Science

A Secure and Efficient Public Auditing System of Cloud Storage Based on BLS Signature and Automatic Blocker Protocol

**A Thesis Submitted to the Department of Computer Science/ College of Science/ University of Diyala
In Partial Fulfilment of the Requirements for the Degree of MASTER in Computer Science**

By

Baidaa Abdulrahman Jaleel

Supervised By

Prof. Dr. Taha Mohammed Hasan

2021AC **1442AH**

بسم الله الرحمن الرحيم

﴿ نَرْفَعُ دَرَجَاتٍ مَّن نَّشَاءُ وَفَوْقَ

كُلِّ ذِي عِلْمٍ عَلِيمٌ ﴾

صدق الله العظيم

سورة يوسف

الاية (76)

Acknowledgments

*First of all, praise is to **Allah**, Lord of all creation, for all the blessings he bestowed upon us for the fulfillment of this thesis.*

*I would like to express my thanks to my supervisor **Prof. Dr. Taha Mohammed Hasan** for his supervision of this research. I extend my thanks and gratitude to all my professors in the Department of the Computer Science/ University of Diyala.*

I express my thanks and gratitude to my husband and my children for their great support and constant encouragement. I thank everyone who has helped support me throughout this phase of the study.

Dedication

To...

My husband Ghassan

My father and my dear mother

My dear children

All our distinguished teachers those who paved the way for our science and knowledge

Baidaa Abdulrahman

Abstract

Cloud computing is a model that enables users to take advantage of remote data storage services in the cloud. With all the advantages of cloud computing, users lack physical possession of external data and this has made the process of maintaining the confidentiality of data stored in the cloud even more important. So that users can resort to cloud storage instead of local storage, along with reducing their concerns about the need to verify the integrity of their cloud data. Therefore, the concepts of confidentiality and data integrity have become two challenges that directly affect the security and efficiency of the performance of cloud systems. This is because one of the assumptions of the threat models is that Cloud Service Providers (CSPs) cannot be completely trusted.

This thesis focused on the goal of overcoming these challenges while paying attention to two aspects of data security concerns. The first aspect is concerned with preserving the confidentiality of data, so the data must be stored in the form of an encrypted file. While the second aspect is concerned with proof of data possession concern. Therefore, an effective auditing method must be relied upon to ensure periodic remote verification of the data of cloud computing users, while not keeping any copies of the data on local storage and this achieves the basic aspects to overcome these challenges in terms of the level of security, public auditing and effective performance. Since public cloud storage auditing is one of these basic aspects, this made cloud users rely on Third-Party Auditors (TPA) to verify the integrity of cloud data. But this audit process should not add any security gaps in the privacy of users' data nor provide them with any extra burden on the Internet. So a facility should be in place to increase the reliability of the TPA and maintain the privacy of user data stored in the cloud. Therefore, this thesis proposes an effective cloud data public auditing system based on Boneh-Lynn-Shacham (BLS)

signature, to ensure public auditing and maintain data privacy. The proposed system also realizes batch audits and dynamic data processes. Besides, the proposed system enhances the level of security authentication through Automatic Blocker Protocol (ABP) to protect the system from unauthorized TPA. As for the overall security and performance analysis, the proposed system is very secure and effective, as it is considered more efficient and secure compared to previous work. The proposed system used a data set (Berka), which is a collection of financial information from a Czech bank. The cloud data auditing rate was 100% and has the lowest costs in terms of computation and communication overheads.

Table of contents

List of Tables

List of Figures

List of Abbreviations

Symbols	Explanation
CSP	Cloud Service Provider
TPA	Third-Party Auditor
IT	Information Technology
AES	The Advanced Encryption Standard Encryption Algorithm
NIST	National Institute of Standards and Technology
AWS	Amazon Web Services
SLAs	Service-Level Agreements
RSA	Ron Rivest, Adi Shamir, and Leonard Adleman Algorithm
DES	Data Encryption Standard Algorithm
TEA	Tiny Encryption Algorithm
CDH problem	Computational Diffie-Hellman Problem
FEC	Forward Error Correction Code
BLS	Boneh, Lynn, and Shacham
SHA256	Secure Hash Algorithm
ABP	Automatic Blocker Protocol
S-PDP	Secure Provable Data Possession Schemes
E-PDP	Efficient Provable Data Possession Schemes

Chapter One

General Introduction

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Introduction

The cloud computing paradigm is the next development of an organization's Information Technology (IT) because it provides many unmatched services in IT include self-service on request, access to the network from anywhere, fast resource adaptability, location independence, payment based on usage and risk management [1,2]. Cloud computing is a good experience with deep implications for changing the way enterprise uses IT. One of the key aspects of this model is that the data is focused or intended for cloud computing. From the viewpoint of users, involving information technology enterprises and individuals, remote data storage in the cloud paradigm in a flexible on-request method brings good advantages such as reduce the load on storage management, overall data access to different geographic positions and decrease spending on devices, software, maintenance, etc. [3]. Cloud storage is one of the basic technologies in the cloud paradigm. Many systems discussed it for low cost and high efficiency of cloud storage; therefore cloud data storage will transform data centers into a large-scale computing service. Of course, the fast growth of bandwidth to the network combined with trust and flexible connection of the network will make users enjoy high-quality cloud services [4]. Cloud storage is dissimilar from traditional storage technologies. It affords large storage space for users and access to data through separate geographical locations. In other words, cloud users can easily access external data from any

device connected to the network and connected to the cloud model anytime and anywhere [5].

This chapter first introduces an introduction to cloud computing followed by motivation, problem statement, importance of the thesis, contributions, related work and finally, a thesis outline is presented.

1.2 Motivation

Despite the enormous benefits of the cloud model, there are also security challenges facing users through their use of outsourced data. Since the management entities of the CSP are separate, the users will relinquish control over their data. Thus, there are many reasons why data correctness on the cloud is at risk. First, cloud computing infrastructures face external and internal threats to data integrity, instances of service unavailable and security breaking of noticeable cloud computing services emerge from one interval to another [4,6]. Moreover, cloud service providers have many incentives to perform dishonestly to the status of cloud computing users concerning their external data. For instance, cloud service providers probably retrieve storage related to financial issues by ignoring unused data or rarely used, or even through hiding data accidents to maintain their reputation [7,8]. In short, the cloud offers no guarantee of data integrity, despite the economic advantages offered by the cloud to store data on a large scale and in the long term. This problem may prevent the successful application of the cloud architecture, if not addressed correct manner. On the other hand, users of cloud computing do not possess their data storage, so the use of traditional encryption methods to protect their data may not be used directly [8]. Particularly, because of the high cost of input and output over the network, it is not possible to provide a workable solution to

download and verify all data. Also, detect the data damage, is mostly inadequate only when data is accessed, as it does not allow users to ensure the authenticity of their data that has not been accessed and it may be too late to detect the data corruption or loss. Data verification on the cloud can be costly to cloud users compared to the large size of data that can be obtained from external sources in addition to the user's limited resource capacity [9,10].

Thus, to ensure data integrity, it is important to activate the public auditing service for the data stored in the cloud computing model, so that cloud users can authorize a third party as an independent auditor to audit the external data when necessary. Cloud users do not have experience auditing cloud data integrity, so there is a need to use TPA on behalf of users to accomplish this task; this is an easy and effortless way for users to make sure the data stored in the cloud is correct [11]. In addition to helping cloud users, TPA work will also be beneficial to cloud service providers in terms of upgrading the cloud services platform [7,12]. As the result, enabling the services of public auditing will play a significant function in making this newly emerging technique fully established, as users need means to evaluate risks and earn confidence in the cloud.

Recently, several systems have been proposed [14,16,13,15,4] to verify data integrity in a cloud computing environment without recovering the whole data. Some of these systems use integrity checks through random sampling, which are limited to queries, while others use integrity checks without the possibility of public verification. Also, some of the schemes mentioned above are not fit for third-party auditing and other systems do not support dynamic data operations or batch audits. The systems suggested by [17, 18, 8] include ensuring that privacy-preserving even though [19, 18] is insecure while [18,8]

is ineffective. The system suggested by [18] is a general audit system with privacy in the cloud model environment but is not secure. The reason for the insecurity of this system is improper identification and the use of private and public parameters during the signature creation process. Therefore, users' mistrust of the cloud and the issues mentioned in the systems above, are those factors that motivated us to suggest this thesis.

1.3 Problem Statement

Storing data in cloud computing bring many problems, largely due to a loss of control over physical control. These issues greatly affect data security as well as the performance of cloud computing systems. This means that data in the cloud is vulnerable to many attacks. This thesis focus on the major security issues in the cloud environment, including data confidentiality and data integrity issue.

a. **Data Confidentiality Problem:** Due to data transfer between cloud service providers (CSP) and users, the data confidentiality problem is increasing. This is because consumers of the cloud outsource their data from managed and somewhat unreliable servers.
b. **Data Integrity Problem:** Because users lose physical control over their data in cloud computing, involuntary security breaches may occur. For example, the CSP may lose users' data due to hardware failure, unintended bugs, or outside intrusion. Or that CSP tries to hide this incident to preserve its reputation. Also, unhonest CSP will unintentionally remove redundancy, resulting in major data errors in terms of their ability to recover or cause data loss.

1.4 Importance of the Thesis

The leading U.S. market research firm Gartner released a report "Assessing the Security Risks of Cloud Computing", this report demonstrates that cloud computing has many risks to data integrity, data recovery, privacy, etc. [20].

Therefore, the main importance of this thesis arises from:

a. This thesis should alleviate the fear of many cloud computing users.
b. The integrity check can be performed without having to save a copy of the data locally.

1.5 Contributions of the Thesis

To address the above challenges and resolve gaps in previous works, this thesis introduces a system for storing dynamic cloud data in a standard security model based on BLS signature. The main contributions of this thesis are:

a. Propose a remote data integrity auditing system based on the BLS signature, that guarantees public auditing, privacy-preserving and batch auditing. Moreover, the proposed system also supports data dynamic operations.
b. The proposed system supports data confidentiality in cloud storage environments using the AES encryption algorithm.
c. Extend the proposed scheme to enhance the level of authentication for security by the ABP to protect the proposed scheme from unauthorized TPA.
d. Evaluate the proposed system through security analysis and have been proven secure in a random oracle model assuming the stability of the CDH problem. The system also has been proven efficient via performance analysis, as well as compare with related systems. The communication overhead of the proposed system is $O(n)$.

This thesis aims to design a secure and effective data auditing system in a cloud computing environment. Based on security and performance analysis, it can be concluded that data confidentiality and data integrity are important factors that must be taken into account when designing a data protection system in cloud environments. The results of the proposed system indicate that the confidentiality of user data enjoys greater privacy and security with lower computational costs. Moreover, data auditing and privacy are maintained efficiently and securely with lower communication costs.

1.6 Related Work

Recently, the integrity and privacy of cloud data have been addressed in many schemes. There are some schemes such as, Ateniese et al. [14] who proposed the first public audit scheme relying on RSA signatures and random sampling methods to present probabilistic evidence of data possession by a remote server. They have given two protocols for data possession which are S-PDP (Secure Provable Data Possession Schemes) and E-PDP (Efficient Provable Data Possession Schemes). S-PDP offers a robust data possession assurance; however, it is less efficient than E-PDP. One limitation of this scheme is that it is valid for static data only. The restrictions on dynamic data processes in a PDP were partially addressed in [37] by permitting modification, deletion and append processes. But, the insertion process is still not maintained in this scheme. The efficiency and scalability are realized in this scheme by using symmetric cipher and hash functions. Whilst this scheme lacks the randomness in challenges and therefore has been shown to be unreliable because CSP can cheat by temporarily storing responses. And for this, there were restrictions on the number of times a data owner could query about the integrity of their data.

While Shacham et al [13]. They proposed two systems POR for static data. The first one is based on short BLS signatures with public auditing. Whereas the second one relies on Pseudo random functions for longer queries and short responses, with special support only for verification. Other properties included non-blocking verification, unforgeability and unlimited queries, however, these systems fail to avoid data leakage. Wang et al. [8] focused their system on the use of utilized bilinear pairing, RSA and Merkle Hash Tree construction for BLS signatures to realize public auditing as well as data dynamics. Though, this system lacks privacy-preserving, a prerequisite for the possibility of public auditing. And after the development of these systems, introduced Wang et al. [19] a verifiable public scheme that could support dynamic operations on the data. Additionally, they have supported batch auditing, which lets multiple auditing be achieved simultaneously. However, this scheme incurred additional performance overheads upon receiving failed audit responses and anyone could modify the data without being detected.

In addition, Ateniese et al. [38] improved PDP by introducing robustness to its scheme. But it still does not support dynamic operations and public validation. Forward Error Correction Code (FEC) is used in the mentioned scheme to achieve data recovery. In this regard proposed Lee et al. [39] a scheme implemented using RSA-based algorithms and used Hash functions (SHA-1 and SHA-512), pseudo-random functions and pseudo-random permutations. This scheme is efficient in terms of data storage and as well as in its utilization of bandwidth. However, the mentioned system is only applicable to static data, although it supports public verification and privacy protection.

In light of this development in systems, Q. Wang et al. [17] introduced a system that could trust the task of allowing TPA to authenticate the integrity of

dynamic data stored in cloud computing. The first recognized the potential security issues of direct extensions with dynamic data and then introduced how to create a verification system to achieve data dynamics. In addition, they explored the bilinear aggregate signature to extend their results into a multi-user setting, in which a TPA can fulfill multiple audit tasks simultaneously. However, they didn't address the issue of privacy-preserving. While Hao et al. [40] proposed a scheme that supports public auditing without TPA. In addition, this scheme does not leak information to any party. A disadvantage of this system is that it does not support batch auditing. While maintaining on preserves privacy the scheme proposed by Wang et al. [41] Supports public verification of remote outsourced storage shared by multiple users, in a manner that preserves privacy. To support public auditing, group signatures are used for homomorphic authenticators and homomorphic MACs form for storage efficiency to support the public audit. Although privacy is provided, it is unable to identify small corruption cases and the costs of sampling and calculation increase with the increase in the masses from which samples are taken. Provide traceability for a large number of users without affecting verification performance.

To support public auditing, C. Wang et al. [42] suggested a cloud storage scheme supporting privacy-preserving public auditing. They extended their result to enable the third-party auditor to achieve audits for various users simultaneously. Shuang Tan and Yan JIA [43] combined the ID-based collective signature and public validation to create the data integrity schema. Therefore, the third-party auditor not only authenticates the integrity of external data on behalf of cloud users but also improves the burden of verifying tasks with the help of user identity. Likewise, S.G Worku et al. [44] proposed a

public auditing system with a third-party auditor who would audit data on behalf of users but for the static data only.

In the feature of supporting static data, Ren et al. [45] proposed a mutual verifiable provable data possession system, which uses the Diffie-Hellman key to construct a homomorphic authenticator. In particular, the verifier in their system is stateless and independent of a cloud storage service. The limitations of this system are that it does not support batch auditing also. Whereas, Yu et al. [46] introduced a zero-knowledge privacy scheme to guarantee that the third-party auditor knew nothing about customer data from all available information. The main limitation of this scheme is that it does not support batch auditing for multiple users at the same time [47]. Zhang et al. [48] proposed two ID-based public auditing systems by combining Waters' signature and public auditing of cloud data. In [49] they suggested a data audit scheme based on Merkel's relatively indexed and time-stamped hash for cloud computing. They guaranteed that the external data was not polluted and that they also included restoring the last copy of the data. This scheme supports public data auditing and supports dynamic data operations. However, this scheme does not support batch auditing. Y. Li et al. [50] proposed a data integrity auditing scheme based on the Fuzzy identity of trusted cloud systems. Sookhak et al. [51] suggested a remote data integrity scheme that utilizes algebraic characteristics of external files in the cloud. They utilize the Divide and Conquer table data structure. The drawback of this scheme does not support batch auditing. Shen et al. [52] proposed a public auditing scheme that supports blockless verification and batch auditing. The dynamic structure of this scheme involves a doubly linked information table and a location array. So, the computational and communication overheads substantially can be

reduced. The limitations of this scheme do not support privacy preservation. F. Wang et al. [53] suggested a dynamic structure and created an identity based non-repudiable dynamic provable data possession system for cloud computing by using index logic tables. This system resisted may attack and avoids the synchronization issue. Moreover, in dynamic operations, this system had lower storage costs and computation costs. However, this theme cannot satisfy the definition of the strictest privacy protection model.

To support privacy preserving, Luo et al. [54] proposed an integrity verification scheme of cloud data based on BLS signature, which ensures public auditing and data privacy preserving. The limitations of their scheme are only static data supported. Shane et al. [55] proposed a scheme that addresses auditing concerns. They used a utilized method based on the use of signatures based on fuzzy identity. Although this scheme only takes into account static data and does not take into account the dynamic operations on the data. Fan [56] suggested a system using identity-based aggregate signatures as the data integrity checking system which resorts trusted execution environment as the auditor to check the outsourced data on the local side. They also achieved secure key management in a trusted execution environment, and it is extended to fully support dynamic data operations. In addition, they considered the situation of multiple file requests for outsourcing concurrently, which can significantly improve the efficiency of integrity checking. With all these advantages of this system, they did not use public auditing and replace the third party auditor with one secure environment on the client side.

For data dynamics, T. Shang et al. [57] proposed a dynamic scheme for identity-based data auditing. To achieve dynamic operations in their scheme, they used Merkle's hash data structure to validate the block tag and thus helped

to update the data while ensuring integrity. Shinde et al. [58] supported a private auditing system, as this system used a double level encryption algorithm for the data file using 256-bit AES algorithm and BLOWFISH algorithm to protect data in the cloud. However, the computational cost of the user-side is high and this system does not support data dynamics as well. Likewise, Ge et al. 2019 [36] supported private auditing by suggesting exploring search investigation using keywords encrypted dynamic cloud data with symmetric key-based validation. They designed a new cumulative authentication tag based on symmetric key cryptography to create a cumulative authentication tag for each keyword. The disadvantage of this system does not support batch auditing. The scheme proposed by Lu et al. [15] supported private auditing and they have proposed a cloud data sharing system for mobile devices. Their system can ensure authorized access to data with security checks before sharing data with users, to avoid incorrect calculation. Their system also achieved lightweight mobile device operations on both the data owner and data requester sides. However, their system did not support dynamic data operations and batch auditing. Whereas Ping et al. [4] suggested a system for public data integrity based on algebraic signature and elliptic curve coding. This system allowed TPA to authorize users to verify the integrity of external data, and additionally resist malicious attacks such as replay attacks, and replace attack and forgery attacks. Data privacy is assured by symmetric encryption. Moreover, they created a data structure called divide and conquer hash list, which can realize data updating processes. They did not consider the situation of multiple files requested for outsourcing concurrently. The scheme proposed by Sun et al. 2020 [5] supported the public auditing and used an authenticated data structure named privacy-preserving adaptive trapdoor hash authentication tree by introducing trapdoor hash and BLS signature to the Merkle hash tree.

They supported data integrity in their proposed scheme, but they did not achieve batch auditing and cloud data confidentiality.

However, previous studies have not highlighted all of the key features of data integrity (public auditing, supporting data dynamics, blockless verification, privacy preservation, unrestricted challenge repetition, recoverability, batch auditing). Little attention has been paid in previous studies to selecting a large-scale and appropriate method for ensuring the integrity of data stored in the cloud. Therefore, a good system should focus on these important features of cloud model security and provide solutions to the issue of providing data confidentiality and data integrity, to be a secure and efficient system at the same time. Table 1.1 explains the comparison of previous works mentioned in this related works.

Table 1.1: Functionality comparison of auditing schemes

Schemes	Public auditability	Data dynamics	Privacy preserving	Blockless verification	Unbounded use of queries	Batch Auditing
[14]	Yes	No	No	Yes	Yes	No
[37]	Yes	Yes	Yes	No	No	Yes
[13]	Yes	No	No	Yes	Yes	Yes
[8]	Yes	Yes	No	Yes	No	Yes
[19]	Yes	Yes	No	Yes	Yes	Yes
[38]	No	No	Yes	Yes	Yes	Yes
[17]	Yes	Yes	No	Yes	Yes	Yes

[40]	Yes	Yes	Yes	Yes	Yes	No
[41]	Yes	No	Yes	Yes	Yes	Yes
[42]	Yes	Yes	Yes	Yes	No	Yes
[43]	Yes	No	No	Yes	Yes	No
[44]	Yes	No	Yes	Yes	No	Yes
[45]	Yes	No	No	Yes	Yes	No
[46]	Yes	Yes	Yes	Yes	Yes	No
[47]	Yes	No	Yes	No	Yes	No
[48]	Yes	No	No	No	No	No
[49]	Yes	Yes	No	No	Yes	No
[50]	Yes	No	Yes	No	Yes	No
[51]	Yes	Yes	No	No	Yes	No
[52]	Yes	Yes	No	Yes	Yes	Yes
[53]	No	Yes	Yes	No	Yes	No
[54]	Yes	No	Yes	Yes	Yes	Yes
[55]	Yes	No	No	Yes	Yes	No
[56]	No	Yes	Yes	No	Yes	Yes
[57]	Yes	Yes	Yes	No	Yes	No
[36]	No	Yes	Yes	Yes	Yes	No
[15]	No	No	Yes	Yes	Yes	No

[4]	Yes	Yes	Yes	Yes	Yes	No
[5]	Yes	Yes	Yes	No	No	No

It is evident from Table 1.1 that the current systems were not able to achieve all the mentioned goals. Therefore, this thesis aim to build a secure and efficient system that can achieve all goals with a lower cost of computation and communication.

1.7 Thesis Outline

This thesis is organized as follows:

Chapter Two: '*Background'* this chapter provides background on cloud computing, along with a detailed related works analysis, and a summary of the chapter.

Chapter Three: *'Proposed System'* introduces a secure and efficient public auditing scheme based on BLS signature and ABP, to ensure public scrutiny and privacy preservation of data integrity in a cloud computing environment.

Chapter Four: '*Evaluation*' evaluates the proposed system through analyzing security and performance, as well as comparing it with related works.

Chapter Five: *'Conclusions and Future works'* this chapter provides concluding observations and suggestions for future work.