Ain Shams University

Faculty of Computer & Information Sciences

Information Systems Department

Integrated Security-as-a-Service Model for Cloud Data Storage

A Thesis submitted to Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University,
in partial fulfillment of the requirements for
the degree of PhD in Computer and Information Sciences

By

Alshaimaa Abo-alian Ahmed

Masters Degree in Computer and Information Sciences,
Assistant Lecturer at Information Systems Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Under Supervision of

Prof. Dr. Mohammed Fahmy Tolba

Professor, Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Prof. Dr. Nagwa Lotfy Badr

Professor, Information Systems Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Cairo 2016

Acknowledgement

First and foremost, I am grateful to Almighty Allah for His immense blessings and graciously helping me to complete this thesis.

This thesis owes its existence to the help, support, and inspiration of many people. In the first place, I owe my deepest gratitude to my main supervisor Prof. Dr. Mohamed Fahmy Tolba whose sharp sense of research direction have provided invaluable feedback to improve the quality of this thesis. This thesis would not have been possible without his sound advice and encouragement. I would like to express my sincere appreciation and gratitude to my associate supervisor, Prof. Dr. Nagwa Lotfy Badr for her tremendous amount of support, insightful comments, and invaluable assistance.

Last, but definitely not least, I would like to thank all my friends and family members for their endless love and support. My love and heartfelt thank to my parents for their lifelong support in all my endeavors. My deepest gratitude goes to my husband who shared with me all the ups and downs and he is a great source of inspiration all along. My sincere thanks and love are extended to my two precious daughters for making every moment in my life meaningful.

Abstract

Cloud computing is an emerging paradigm that delivers a large pool of virtual, on-demand and dynamically scalable resources to users via Internet technologies, following the notion of pay-as-you-go. Examples of these resources include computational power, storage capabilities, hardware platforms and applications. The key advantages of cloud computing are immense flexibility and monetary savings through minimization of infrastructure and software investments as well as management and maintenance costs. Besides popular cloud infrastructure and platform providers, such as Amazon, Google, and Microsoft, there are many cloud storage providers which offer more accessible and user friendly data storage services to cloud customers. Examples of these services include Dropbox, SkyDrive, Box.net, Zoho, Ubuntu One or Apple iCloud.

Along with the widespread interest on cloud computing, however, there are still concerns that hinder the proliferation and the adoption of cloud services. One of the main concerns is data security in cloud storage environments. Numerous research problems belonging to the cloud storage security have been studied intensively before. However, addressing the three dimensions of outsourced data security (i.e., confidentiality, integrity and availability) as a cloud service is still a challenge in cloud storage. As there is always a tradeoff between maintaining security and obtaining efficiency, it is difficult but nevertheless essential to explore how to efficiently address security challenges over dynamic cloud data.

The thesis first addresses the security requirements for cloud storage as identified from the literature, given the difficulty that data are no longer locally possessed by data owners. Then it aims to design an integrated **Security-as-a-Service** model for data storage in the cloud that provides authentication, access control, auditing and data management services. We propose a new keystroke authentication system for verifying the identity of cloud users. The proposed keystroke authentication system removes redundant or irrelevant features from the large scale keystroke dynamics by combining different feature selection methods and different fusion rules which, in turn, achieve higher authentication accuracy and performance. Moreover, it eliminates the tradeoff between the authentication accuracy and the elapsed time of the verification process by clustering the user profile templates in the keystroke dataset.

Then, a dynamic access control system is proposed to ensure data confidentiality in cloud computing. The proposed access control system supports automatic user role assignments so that it relieves the data owner from the online and computational burdens of user role assignment processes, especially for large scale systems with a huge number of users and continuously changing user role policies. Additionally, the proposed access control system tackles the key escrow and key management problems in a decentralized cloud environment by defining roles in a hierarchy and supporting key delegation.

Finally, a public auditing system is proposed to delegate the integrity verification of outsourced data in the cloud storage to a third party auditor. The proposed auditing system is privacy preserving so that keeps the data confidential/invisible to the auditor during the auditing process. Moreover, a data management system is proposed to support data dynamics for replicated and single-copy data files with variable sized blocks on the cloud storage. So, the proposed system supports updates with a size that is not restricted by the size of file blocks. It thereby offers extra flexibility and scalability compared to existing systems. To address the efficiency problem in verifying variable-size updates for cloud storage with multiple replicas, the proposed system incorporates a new authenticated data structure, namely Modified Rank based Authenticated Skip List (MRASL). The proposed MRASL supports verification of all dynamic data replicas at once. It thereby reduces the computation and communication costs. Moreover, the proposed auditing system supports efficient data recovery to repair the corrupted data in the case of single copy data files. Additionally, the proposed auditing system supports batch auditing where multiple auditing tasks with different data files can be performed simultaneously. Extensive experiments and performance analysis demonstrate the effectiveness and efficiency of the proposed model.

TABLE OF CONTENTS

Acknowledgment . . . . **i**

Abstract . . . . **ii**

Table of Contents . . . . **v**

List of Figures . . . . **x**

List of Tables . . . . **xiv**

List of Abbreviations . . . . **xvi**

List of Publications . . . . **xx**

Chapter 1: Introduction

1.1 Overview 1

1.2 Problem Definition 4

1.3 Contributions 6

1.4 Thesis Organization 7

Chapter 2: Background and Preliminaries

2.1 Overview 9

2.2 Cloud Background 10

2.2.1 Cloud Basics 10

2.2.2 Cloud Storage Services. 15

2.2.3 Security Issues of Cloud Storage 17

2.3 Cryptographic Background 19

2.3.1 Bilinear Maps 19
2.3.2 Hash Functions. 20
2.3.3 Erasure Correction Codes 21
2.3.4 Probabilistic Encryption 22
2.3.5 Homomorphic Verifiable Tags. 23
2.4 Authenticated Data Structures 24
2.4.1 Merkle Hash Tree 24
2.4.2 Balanced Update Tree 25
2.4.3 Skip List 26
2.4.4 Index Table 28

Chapter 3: Literature Review

3.1 Overview 30
3.2 Authentication Systems. 30
3.2.1 Keystroke Dynamics based Authentication Systems 32
a. Fixed text keystroke authentication systems . . . 33
b. Free text keystroke authentication systems. . . . 35
3.3 Access Control Systems. 37
3.3.1 Traditional Encryption 38
3.3.2 Broadcast Encryption. 39
3.3.3 Identity-Based Encryption. 40
3.3.4 Hierarchical Identity-Based Encryption 42
3.3.5 Attribute-Based Encryption 43
a. Key policy attribute-based encryption 44
b. Ciphertext policy attribute-based encryption 47
3.3.6 Hierarchical Attribute-Based Encryption 48
3.3.7 Role-based Access Control 51

3.4 Data Auditing Schemes 52

3.4.1 Auditing Schemes for Single-Copy Data 54

3.4.2 Auditing Schemes for Multiple-Copy Data 56

3.4.3 Auditing Schemes for Static Data 57

3.4.4 Auditing Schemes for Dynamic Data 58

Chapter 4: The Proposed Security-as-a-Service Model for Cloud Environment

4.1 Overview 61

4.2 The Proposed System Architecture 62

4.3 The Proposed Keystroke based Authentication Subsystem 64

4.3.1 Data Acquisition 66

4.3.2 Keystroke Feature Extraction 67

4.3.3 Clustering 68

4.3.4 Classification. 68

4.4 The Proposed Access Control Subsystem. 68

4.5 The Proposed Auditing and Data Management Subsystems 74

4.6 Summary 81

Chapter 5: The Proposed System Implementation

5.1 The Detailed Construction of the Proposed Keystroke based Authentication Subsystem 83

5.1.1 Keystroke Feature Extraction 83

a. Outlier detection 84

b. Feature selection and reduction 84

5.1.2 Clustering 88

5.1.3 Classification 90

a. Support vector machines based classification . . 90

b. Naive Bayesian based classification 90

c. Multilayer perceptron based classification 91

5.2 The Detailed Construction of the Proposed Access Control Subsystem 92

5.2.1 The First Access Control Subsystem Construction 92

5.2.2 The Second Access Control Subsystem Construction 97

5.2.3 Performance Analysis of the Proposed Access Control Subsystem 101

5.3 The Detailed Construction of The Auditing and Data Management Subsystems 104

5.3.1 The Data Management Subsystem Construction . . . 104

5.3.2 The Auditing Subsystem Construction. 111

5.3.3 Supporting Batch Auditing 113

5.3.4 Performance Analysis of the Proposed Auditing and Data Management Subsystem. 114

5.4 Security Analysis 117

Chapter 6: Experimental Results and Discussions

6.1 Overview 121

6.2 Experiments on the Keystroke Authentication Subsystem 122

6.2.1 Datasets Description 122

6.2.2 Evaluation Criteria 125

6.2.3 Results and Discussion 126

a. The impact of feature selection and fusion methods. 126

b. The impact of number of clusters. 129

c. The impact of classification methods, user sample size and password length 129

d. Comparison between the keystroke authentication subsystem and the state-of-the-art 135

e. Other accuracy criteria 137

6.3 Experiments on the Access Control Subsystem 137

6.3.1 Experiments on the First Construction 138

a. The effect of the role depth and the number of users per role 138

b. The effect of number of user attributes 140

c. The effect of the file size 140

6.3.2 Experiments on the Second Construction 141

6.4 Experiments on the Auditing and Data Management Subsystems 143

6.4.1 The Impact of Corruption Rate and Challenge Size. 144

6.4.2 The Impact of File Size and Block Size 145

6.4.3 The Impact of Number of Replicas. 146

6.4.4 The Impact of Number of Users 149

6.4.5 The Advantages of Batch Auditing 150

6.4.6 Comparison between the Auditing Subsystem and the State-of -the-art 152

Chapter 7: Conclusions and Future Work

7.1 Conclusions 156

7.2 Future Work 158

References. 160

LIST OF FIGURES

2.1 Cloud computing essentials. 11
2.2 The top cloud benefits. 14
2.3 The top challenges within cloud environment. 15
2.4 Merkle Hash Tree. 25
2.5 Example of balanced update tree operations 26
2.6 Rank-based authenticated skip list 27
2.7 An example of index table during different file operations. 28

3.1 Symmetric key encryption versus public key encryption. . . 39
3.2 Example of an identity-based encryption. 41
3.3 Hierarchical identity-based encryption 43
3.4 Key policy attribute-based encryption in a healthcare system 45
3.5 Hierarchical attribute-based encryption model 49
3.6 Structure of an auditing scheme 54

4.1 The proposed security-as-a-service model 63
4.2 The proposed keystroke based authentication subsystem . . 66
4.3 The proposed access control subsystem 70
4.4 An example of role hierarchy 71
4.5 Example of an access policy tree 72
4.6 The proposed auditing scheme 77
4.7 The modified rank-based authenticated skip list (MRASL) 79

5.1 The clustering process in the offline mode 89
5.2 The proposed "Create Role" process of the first construction 94
5.3 The proposed "Create Role" process of the second construction 99
5.4 The key generation process 105
5.5 The tag generation process 108
5.6 An Insertion after the 1st block into the MRASL in Fig. 4.7 109
5.7 A Deletion of the 2nd block for the MRASL in Fig. 4.7 . . . 110
5.8 The Key Modification process 110

6.1 Some data samples from the CMU keystroke dynamics benchmark dataset 123
6.2 The impact of different cluster sizes 130
6.3 The impact of different classifiers and sample size on equal error rate (EER) using CMU and AndroidKeystroke datasets 131
6.4 The impact of different classifiers and sample size on the elapsed time using CMU and AndroidKeystroke datasets . 132
6.5 The impact of different classifiers and password length on EER using three datasets 133
6.6 The impact of different classifiers and password length on elapsed time using three datasets 134

6.7 The effect of the role depth and number of users per role on decryption, encryption and user revocation times 139

6.8 The effect of the number of user attributes 140

6.9 The effect of file size 141

6.10 The effect of role depth and number of users/role on the decryption time (in the second construction) 141

6.11 The effect of role depth and number of users/role on the ciphertext (in the second construction) 142

6.12 The user secret key size of the proposed system in comparison with the TimePRE scheme 143

6.13 The effect of different corruption rates and detection probabilities on the computational time of the proposed auditing subsystem. 144

6.14 The effect of different file sizes and block sizes on the computational time of the proposed auditing subsystem .. 145

6.15 The effect of the size of the blocks modified on the verification time 146

6.16 The effect of different number of replicas and file sizes on the computational time of the proposed auditing subsystem 147

6.17 The effect of the number of replicas on the communication costs of dynamic operations 148

6.18 The effect of the number of replicas on the storage overheads 149

6.19 The effect of the number of users on the proposed scheme scalability 149

6.20 The effect of the number of auditing tasks on the verification time 150

6.21 The effect of the number of auditing tasks on the communication costs 151

6.22 The effect of the number of auditing tasks on the verification time and the detection probability 152

6.23 The computational time of the proposed auditing scheme (multiple-replica version) and [48] using different numbers of replicas 153

6.24 The computational time of the proposed auditing scheme (single-copy version) and [160] using different numbers of challenged blocks 154

LIST OF TABLES

2.1 Cloud SPI services delivery vendors. 12

2.2 Commercial cloud storage providers 16

2.3 Comparison of different cryptographic hash functions 20

2.4 Comparison between different probabilistic encryption schemes 23

4.1 Example of keystroke features 67

5.1 Feature fusion rules. 88

5.2 Performance analysis of the proposed access control subsystem and the state-of-the-art 102

5.3 The advantages of batch auditing. 114

5.4 Performance analysis of the proposed auditing subsystem and the related work 116

6.1 Keystroke features in the AndroidKeystroke dataset 124

6.2 Summary of the keystroke datasets used 125

6.3 A Simple confusion matrix 126

6.4 The impact of feature selection and fusion methods using the CMU dataset 127

6.5 The impact of feature selection and fusion methods using the GP dataset 128