OBLIVIOUS MULTI-CLOUD FILE STORAGE

Tharushi Dinethri Pushpakumara

219387A

Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

July 2023

OBLIVIOUS MULTI-CLOUD FILE STORAGE

Elibichchiya Ralalage Tharushi Dinethri Pushpakumara

219387A

Thesis submitted in partial fulfillment of the requirements for the

Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

July 2023

DECLARATION

"I declare that this is my own work, and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature Date: 31st July 2023

The above candidate has carried out research for the Master's thesis under my supervision.

Name of the supervisor: Dr. Sunimal Rathnayake

Signature of the supervisor Date: 1st August 2023

ABSTRACT

Cloud storage facilities are now predominantly used to store outgrowing data. Information availability, improved performance and the trustworthiness are the key factors that the data owners mainly focus on, in storing data with a third party. With the multi-tenant concept on cloud computing, security threats have been evolved, as the trustworthiness of the neighbors has become a doubt. A malicious user could monitor the traffic between the client and the CSP. By analyzing the traffic attacker can get a clear picture regarding what kind of data has been passed or retrieved by the client and these questions the privacy level of stored data.

Critical, highly Sensitive and Personally Identifiable Information (PII) used in government organizations such as Defense Ministry, Person's Registration, Motor Traffic Department, Immigration and Emigration systems, among others, require data privacy, integrity and confidentiality which demotivate them in storing these highly sensitive data on cloud storage. But these organizations handle thousands of data records and adding more day by day and the physical storage expansion has become a huge challenge with the investments on infrastructure. The proposed solution would address both these challenges. The major security concerns the proposed solution focuses on is the data privacy, integrity, and confidentiality.

In this research we propose a novel approach to obfuscate the data distribution patterns in a multi cloud environment. The solution is to be implemented at the client side based on the systems' business requirements. So that a unified interface could be provided in storing/retrieving data in several cloud platforms. The uploaded file is encrypted with a public key, calculated the hash value, and divided into several small file chunks. Then the file chunks are scattered across several Storage accounts created on several CSPs randomly and hence, the confidentiality, integrity and privacy of data also can be achieved. The proposed solution consists of a central component through which all the communication between the client and the CSPs take place. Technology which is used within the central component is related to the ORAM concept. Further this facilitates dynamical scaling up of cloud storages.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Dr. Sunimal Rathnayake, for providing abundant guidance, support, and encouragement throughout this research. Also, I would like to thank my colleagues for sharing knowledge, support, and constant encouragement.

Last but not least, I express my love and gratitude to my beloved family for their continuous and unparalleled love, help, and support.

TABLE OF CONTENT

DECLARATION ..... *i*
ABSTRACT ..... *ii*
ACKNOWLEDGEMENTS ..... *iii*
TABLE OF CONTENT ..... *iv*
LIST OF FIGURES ..... *vi*
LIST OF TABLES ..... *vii*
ABBREVIATIONS ..... *viii*
1. INTRODUCTION *1*
1.1. Background **1**
1.2. Motivation **2**
1.3. Research Problem **3**
1.4. Research Objectives **5**
2. LITERATURE REVIEW / RELATED WORK *6*
2.1. Oblivious RAM (ORAM) **6**
2.2. Attacks on Memory **7**
2.3. Oblivious P2P **8**
2.4. Multi Cloud ORAM **9**
2.5. Oblivious Multi Cloud Storage **11**
2.6. Integrity Auditing **14**
2.7. ISSE-2 Encryption **15**
2.8. Bastion Algorithm **16**
2.9. Summary **16**
3. METHODOLOGY *17*
3.1. Overview of the proposed method **17**
3.2. Oblivious Distributor **18**
3.3. File uploading **19**
3.3.1. Encryption 20
3.3.2. Hashing 20
3.3.3. File Splitting 20
3.3.4. Uploading 21
3.4. Fetching a file **21**
3.5. File Upload Table **22**

3.6. File Chunk Table **22**

3.7. Summary **23**

4. EVALUATION ***24***

4.1. Overview **24**

4.2. Experiment Setup **24**

4.2.1. Development Language and Platform 24

4.2.2. Swagger UI 24

4.2.3. AWS and Azure 25

4.2.4. SQL Server Manager 25

4.2.5. Entity Framework 25

4.2.6. RSA Algorithm 26

4.2.7. MD5 Hashing 26

4.3. Implementation **27**

4.3.1. File Uploading 27

4.3.2. File Fetching 28

4.4. Validating Objectives **29**

4.4.1. Hiding Access Patterns to Cloud Storage 29

4.4.2. Confidentiality 30

4.4.3. Integrity 31

4.4.4. Unified Interface to Access Multiple Cloud Storages 32

4.5. Summary **34**

5. SUMMARY AND CONCLUSION ***35***

5.1. Summary **35**

5.2. Limitations **35**

REFERENCES ***36***

LIST OF FIGURES

Figure 1: operation of Path ORAM 7
Figure 2: Tracker updates on TagMap, and FileMap 8
Figure 3: Shuffling and Transferring Flow 9
Figure 4: Private + Public cloud scenario and Trusted Hardware in cloud scenario 10
Figure 5: Flow of Read and Write Partitions 12
Figure 6: Identity Based Integrity Auditing 14
Figure 7: Data Deduplication Factors 15
Figure 8: Connectivity of Oblivious Tracker 18
Figure 9: File Uploading 19
Figure 10: Retrieving a File 22
Figure 11: Hashing and Encryption 27
Figure 12: File Chunk Count Range 27
Figure 13: Random Cloud Storage Selection 28
Figure 14:File ID Mapping 28
Figure 15:Parallel Download of File Chunks 28
Figure 16: Decryption and File Downloading 29
Figure 17: Wireshark Capture 01 29
Figure 18: Wireshark Capture 02 30
Figure 19: Azure File Storage 30
Figure 20: AWS File Storage 31
Figure 21: Unreadable File Chunk Example 31
Figure 22: Generating Hash Values 32
Figure 23:Hash Value Table 32
Figure 24: Unified Interface (Landing Page) 33
Figure 25: Unified Interface (File Uploading) 33
Figure 26: Unified Interface (File Downloading) 33
Figure 27: Configuration File 34

LIST OF TABLES

Table 1: File Upload Table 22
Table 2: File Chunk Table..... 23

ABBREVIATIONS

CSP – Cloud Service Provider

PII – Personally Identifiable Information

CI – Critical Infrastructure

STaaS – Storage as a Service

API – Application Programming Interface

SaaS – Software as a Service

ORAM – Oblivious Read Only Memory

RORAM – Ring ORAM

RDIC – Remote Data Integrity Checking

DRAM – Dynamic Random Access Memory

MCOS – Multi Cloud ORAM Scheme

TCB – Trusted Computing Base

AA – Actual Address

LA – Logical Address

ISSE - Integrated Searchable Symmetric Encryption