بسم الله الرحمن الرحيم

∞∞∞

تم رفع هذه الرسالة بواسطة / مني مغربي أحمد

بقسم التوثيق الإلكتروني بمركز الشبكات وتكنولوجيا المعلومات دون أدنى مسئولية عن محتوى هذه الرسالة.

ملاحظات: لا يوجد

Faculty of Computers and information Science
Information System Department
Ain Shams University

Confidentiality and Integrity of Query Results for Cloud Databases

A Thesis submitted in partial fulfillment of the requirements for the degree of
Master of Information System in Computer and Information Science

By

Mai Mustafa Adly Rady

B.Sc. in Information System,
Faculty of Computer and Information Sciences,
Arab Open University.

Under the Supervision of

Prof. Rasha Ismail

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

Dr. Tamer A. Mostafa

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

2022

Acknowledgements

First, I would like to express my sincere gratitude to my supervisors: Prof. Rasha Ismail and Dr. Tamer Abdelkader for their support, guidance and motivation during my work on this master's thesis. I cannot begin to express my thanks to my father, mother, brother and husband, for their support, encouragement and endless love to finish this master's thesis.

Abstract

The digital data amount is increasing at a phenomenal rate, outpacing the local storage ability of many organizations. Therefore, outsourcing data is considered a solution to store large data into efficient cloud servers. Cloud computing environment offers on demand access to several computing services providing benefits such as reduced maintenance, lower fundamental costs on different resources, global access, etc. Database as a Service (DBaaS) is one of its prominent services. Database Service Providers (DSPs) have the infrastructure to host outsourced databases at distributed servers and provide efficient facilities for their users to create, store, update and access databases anytime from any place through Internet. Outsourcing data into cloud servers have many features as flexibility, scalability and robustness, but in order to provide these, features as data confidentiality and data integrity are often sacrificed .

This thesis presents secure scheme to provide Confidentiality and Integrity of Query results for Cloud Databases (SCIQ-CD). The scheme architecture integrates different cryptographic techniques such as: AES, RSA, and SHA2 to achieve the data security properties. The scheme allows the data owner to outsource their database, which includes sensitive data to a DSP, and performs dynamic and static operations on the outsourced Database. In addition, the scheme uses trusted third-party server that serves as intermediate server between the DSP and users to check the data confidentiality and the data integrity. The security analysis demonstrates that our construction can achieve the desired security properties. The performance analysis shows that our proposed scheme is efficient for practical deployment, it imposes a small overhead for select, range, update and a reasonable overhead for insert and delete.

Table of Contents

List of Figures

List of Tables

Chapter (1)

Introduction

This chapter briefly introduces the background of the research topic. The outline of this chapter is the following: section 1.1 introduces the motivation of this work, Section 1.2 introduces the problem statement, section 1.3 introduces challenges that face storing database in CSP, section 1.4 stating the research objective, section 1.5 contains the approach that has been followed, section 1.6 describes the design science research methodology that is applied, section 1.7 provides the organization of the rest of our work and Section 1.8 concludes the chapter.

1.1Motivation

Cloud computing is a paradigm of computing that is shifting the way of thinking about IT industry. It provides different computing services remotely rather than locally and these services accessed through the Internet [17]. Since the invention of the internet in the 1970's, there have been different services offered to users; as they able to login remotely and transfer files via the FTP protocol, but the cloud environment took the online services to a new dimension [18]. The cloud computing services are hosted by cloud service providers (CSPs) and users of the CSP pay for these services according to the service type and their usage of the service [1].

The cloud computing services can be classified into three models: 1) Infrastructure as a service IaaS, 2) Platform as a service PaaS, and 3) Software as a service SaaS. In the infrastructure as a service, users being able to use servers, storage, network settings on-demand from the CSP and only pay per their usage. In the platforms as a service, users being able to build their own applications as the CSP offered all resources and development tools such as: databases, which are required to build the application and also maintain and secure this application. In the software as a service, the application is hosted as a service offered to users through the internet and the users doesn't have to worry about its maintenance, backups or security [20, 21].

The benefits of Cloud Computing for its users are various and significant as in [2, 19]:

1. Avoid big initial investments in hardware and software purchasing,
2. Flexibility and scalability: providing services to the user according to their needs,
3. Availability: different services accessed from anywhere through an internet connection,
4. Reduce maintenance and operational costs,
5. Reduce costs as users only pay for the service per their usage.

Alongside cloud computing services, the data storage is an important service that has been progressing, evolving and adapting [3]. The reason is that the digital data amount is increasing at a phenomenal rate, outpacing the storage ability of many organizations. Therefore, storing data in cloud servers is considered as a solution to store greater data into efficient distributed servers. By storing the data in the cloud, the fundamental costs of different resources such as software, hardware and even professionals hired to maintain the system are reduced [5]. The Storage as a Service is offered by different cloud service providers that allow users or organizations to store

data on remote servers than organization servers. The Storage as a Service is offered under the Infrastructure as a service model or Platform as a service model. In the Infrastructure as a service, the CSP offered the infrastructure for users to be able to upload their data, and the user maintains the data management storage system. In the Platform as a service model, the database to store data offered by CSP which maintains the data management storage system and the user manages the data .

Database outsourcing introduces a new paradigm, called database as a service (DBaaS) offered by different Database Service Providers (DSPs) which have the infrastructure to host the outsourced databases at distributed servers and provide efficient facilities for their users to create, store, update and access databases anytime from any place through Internet connection. The relational databases and non-relational databases can be outsourced, but our main concern in this work about relational databases. Amazon's SimpleDB, Amazon RDS, Google's BigTable, Yahoo's Sherpa and Microsoft's SQL Azure Databases are the commonly used databases in the Cloud [4].

Database Service Providers (DSPs) offer database cloud services in three different models: 1) Virtual Machine (VM) image, 2) Database as a service (DBaas), 3) Managed Hosting. In the virtual machine image model, the DSP offers infrastructure, which a Database Management System DBMS can run, to users that can upload or purchase DBMS. In DBaaS model, the DSP maintains the DBMS and the user manages the databases supported by the DBMS and paying for storage and resources. In the managed hosting model, the three phases of database implementation installs, maintains and manages is done by DSP [15, 29]. In our work we have implemented our scheme in the DBaaS model as we will explain later.

By storing database in the cloud using the DBaaS model, the data management becomes one of the DSP tasks to satisfy essential requirements that are inherent to these environments and the users can concentrate on their main tasks. Some of these requirements are related to: 1) scalability by handling the increase and decrease of active users and store the growing amounts of data, 2) flexibility by allowing the storage of large amounts of data, and 3) availability as the users can access the service from anywhere through an internet connection. The DSP offers these requirements, but doesn't guarantee the data security as the cloud servers and the networks are often targets of malicious attacks. Moreover, the cloud server, it might be malicious and may attempt to insert fake records into the database or modify existing records or even delete it [5]. To guarantee the data security, the database may be encrypted before outsource, but we still need to ensure that the query result is correct, complete and from the last version of the data. And this is the motivation of our work.

1.2 Problem definition

By storing the database in cloud service providers (CSPs), the users must ensure that the data is secured from inside and outside malicious. There are several issues must be taken in concern to store the database in CSPs [23], as follows:

1. Security issues: the main issue here is the data, by storing the database in the cloud, the data can be accessed by anyone and from anywhere; data can be stolen by unauthorized users or attackers. Besides, data can be lost, as the server is suddenly shutdown or during natural disasters, which cause data damaged. Moreover, some CSPs don't

provide their own server, which use another server because of the cost effectiveness and flexibility.

2. Privacy issues: the CSPs enforce their own policies to ensure the security of the database stored on their servers. The database stored in the CSPs accessed only by authorized users. Checking users' authorization done on the provider side and also the user side.
3. Application issues: the database monitoring and maintenance should be done by the CSPs to ensure that the cloud is secure and not infected by malicious code that have been uploaded to the cloud by hackers or attackers with the purpose of stealing sensitive information or even damaging the information of certain users.

Among the previously issued, the main question here is how we can protect the outsourced database from inside and outside malicious or attackers. Here are the important security issues to be addressed during our work, data confidentiality and data integrity. Data confidentiality can be guaranteed by encrypting the database by the data owners', using a certain key before outsourcing it to the DSP and only the authorized users can decrypt it [6]. Unauthorized users and the DSP will not have the key to decrypt the data. Data integrity is applied on query results, which are retrieved from the DSP. Data integrity includes three aspects: 1) Correctness, the returned results do exist in the outsourced database, 2) Completeness, the result is complete and there is not any missing parts, and 3) Freshness, the results are based on the recent version of the data .

Our main issue in this work is how to achieve the confidentiality and integrity of the query result of cloud databases.

1.3 Challenges

Among the increasing of the digital data amount and speed of threats to outsourced databases, there are some challenges meet the outsourcing database security as follows [22]:

1. Data Quality

   By storing database in the cloud, the database is under control the CSP. The data owner and users must ensure that the database is secure and the returned data is correct, complete and from the recent version of the data. To evaluate and assure the data quality, some methodologies and techniques are used.
2. Privacy preserving databases

   To achieve the database privacy, different approaches can be used such as data anonymization, which modify the data by removing all information that can directly link data items with users like names or ID numbers, but this may not be enough to anonymize the data. To overcome this problem, use data mining techniques that allow the recovering of the removed information. When the data is under a lot of modifications, the database quality may be affected [24].
3. Intellectual Property Rights

   Watermarking techniques are used to protect the content of the organization's data with from unauthorized duplication and distribution by enabling provable ownership of the content.

4. Database Survivability
   Database systems need to operate and continue their functions, even with reduced capabilities. Moreover, prevents attacks on data and detect if one happened. Recover corrupted or lost data and repair failed system functions to re-establish a normal level of operation.

1.4 Objectives

This thesis focuses particularly on the important security issues to be solved among outsource the database into the DSP. The most important security issues to be addressed are data confidentiality and data integrity.

1.5 Contribution

In order to achieve the aforementioned objective, we implement and test a secure scheme to provide Confidentiality and Integrity of Query results for Cloud Databases (SCIQ-CD) on the Microsoft Azure Cloud. Data confidentiality guaranteed by encrypting the sensitive attributes from the database before outsourcing it, and only the authorized users can decrypt it [70]. Data integrity guaranteed by converting the database to authenticated data structure and store it within the database in DSP, and data integrity, includes three aspects: correctness, completeness, and freshness. In this model, we use Trusted Third Party (TTP) to provide an indirect mutual trust between the data owner/users and the DSP, which checks the data confidentiality and verifies the query result integrity before sending it to the users. The performance of this solution is assessed in terms of response time for data outsourcing, and data retrieving.

1.6 Thesis Outline

The remainder of this thesis is organized as follows:

- Chapter 2 provides the background information of our research regarding Cloud Computing, database as a service, database outsourcing model, and database security issues.
- Chapter 3 describes the design of the cloud based storage scheme (SCIQ-CD).
- Chapter 4 describes the implementation of solutions in the Microsoft Azure Cloud.
- Chapter 5 details the experimental design used to measure the performance of the solutions.
- Chapter 6 presents the conclusions and further ideas to extend this research.

Chapter (2)

Background and Related Work

This chapter presents an overview of some of the significant topics relevant to this thesis. Section 2.1 describes cloud computing, its characteristics and the various services it provides. Section 2.2 gives a detailed description of the cloud database service as one of the key services provided in a cloud environment. Section 2.3 presents some of the challenges existing in the database as a service. Section 2.4 addresses database privacy as one of the crucial challenges in the DBaaS. Section 2.5 concludes the chapter.

2.1 Cloud Computing

Cloud computing is a major paradigm of computing that is shifting the way of thinking about IT industry by producing services and tools to be used in this industry. These services and tools offered to all different kinds of users and known as: X as a Service through internet connection without the need for costly infrastructure [8].

The National Institute of Standards and Technology (NIST) define Cloud Computing as [26]: A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics (On-demand self-service, broad network access, resource pooling, rapid elasticity, measured service), three service models (Software-as-a-Service, Platform-as-a-Service, Infrastructure-as-a-Service), and four deployment models (Hybrid Cloud, Private Cloud, Community Cloud, Public Cloud).

Cloud computing is similar to remote computing, as the users accessed other computers from their local machine through a network. A Cloud also can be defined as a parallel and distributed system which has a number of virtualized and interconnected computers. These are actively provisioned and presented as single or more united computing resources depending upon the service level agreement [13]. Cloud computing also offers the virtualization of applications and services, as the applications and services appear to users as it runs on their device rather than a remote cloud server. Thus, the users didn't have to install the actual software application to run the application, so both the expert and naive users didn't need to worry about the technical details and configurations to use these cloud services [9].

Cloud computing is based on a pay-as-you-go model, where users just pay for computing resources per the duration of their usage and didn't need to concern with how these computing resources are hosted or managed [10]. Applications and databases are stored in large data centers owned by companies such as Microsoft, Oracle, Google and IBM.

The architecture of cloud computing services has three main elements as shown in figure 2: 1) user: who uses any hardware or software application from cloud services as the front-end to perform their work, 2) web service: any software applications that are used to perform cloud computing, 3) the cloud server: that provides the service to the user.

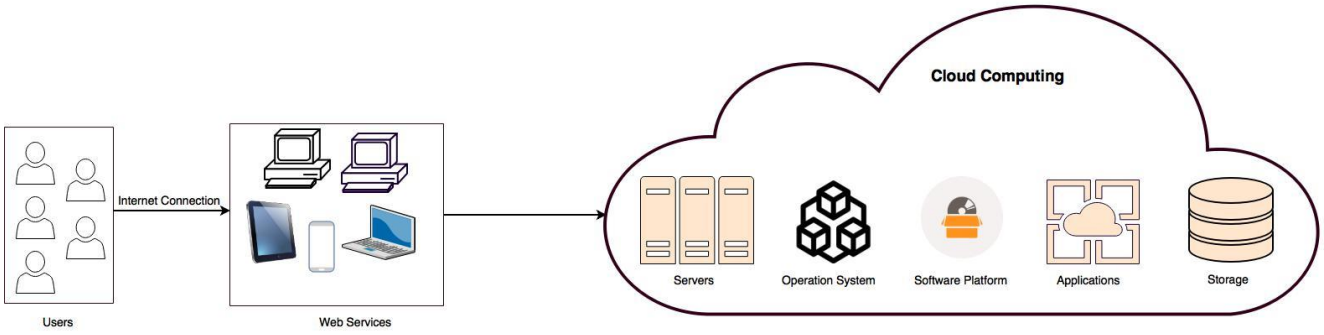

Figure 2.1: Cloud Computing Architecture

2.1.1 Cloud computing characteristics

The cloud computing essential characteristics is summarized in this subsection as follows [26, 28]:

- On demand self-service: Enables the users to access the computing capabilities as server and application automatically as needed without any interaction between user and service provider.
- Broad network access: The computing capabilities are available over the Internet and can be used by heterogeneous users' platforms as laptops, desktop and mobile phones.
- Resource pooling: the computing resources can be combined and dynamically assigned to serve multiple users based on a multi-tenant model. Examples of resources include storage, memory, and network bandwidth.
- Rapid elasticity: according to the user needs based on the demand, every computing capability can be provisioned rapidly, elastically and/or automatically to scale out or in, to meet the changes of the demand .
- Measured service: the cloud system automatically controls and optimize the resource usage to provide monitoring, controlling and reporting for providing transparency between the service provider and the user of the service.

2.1.2 Cloud computing service models

Cloud services can be classified into three models, Infrastructure as a service IaaS, Platform as a service PaaS, and Software as a service SaaS.

Figure 3 provides an overview of the corresponding service models for each cloud architectural layer [20, 21, 27, 28].

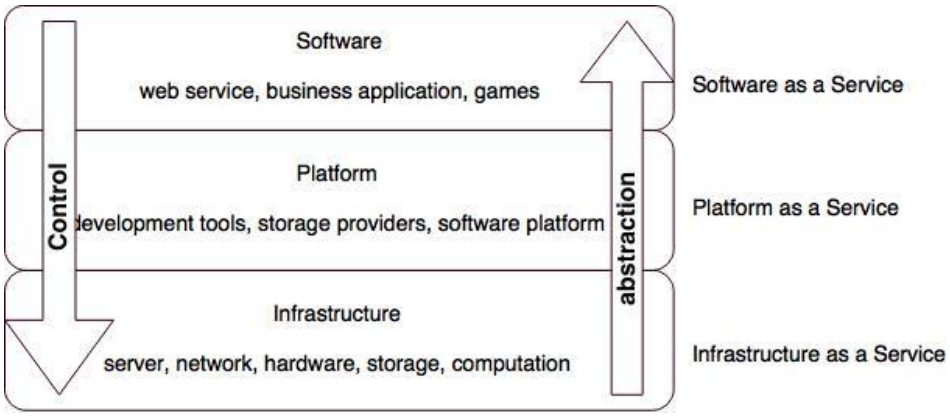

Figure 2.2: Cloud Computing Service Models

Infrastructure as a Service (IaaS): is the service where users can use the fundamental computing resources such as servers, network, storage, or hardware to create and run applications or even operating systems. The IaaS requires a virtualization platform where the users can install and configure a virtual machine which runs on the cloud servers. The CSP is responsible to control and manage the Cloud infrastructure while the user is responsible for the management of the virtual machine. Example of IaaS: Amazon's Elastic Compute Cloud (EC2), Windows Azure Virtual Machines, Google Compute Engine.

Platform as a Service (PaaS): is the service where a hardware or software platform is provided to users to be able to create their applications. A platform could be hardware equipment, software applications, or programming environment which can be used by users to create and run their own applications or to improve applications. The CSP is responsible to control and manage the Cloud infrastructure while the user can control and manage only the deployed applications. Examples of PaaS: Google App Engine, AWS (Amazon Web Services), Microsoft Azure.

Software as a Service (SaaS): is the service provided by the CSPs where users didn't have to install the software applications to maintain or support it. The CSP offers the computing capability which is deployed on a Cloud infrastructure and the users access the applications through a web browser or application interface. Examples of SaaS: Google apps (Gmail, Google Docs), YouTube, Facebook.

2.1.3 Cloud computing deployment models

Cloud computing is composed of four deployment models, private cloud, public cloud, community cloud, and hybrid cloud [26].

- **Private cloud:** this model used to protect sensitive data, as the infrastructure is provisioned for exclusive use by a single organization. The management and the operational tasks can be carried by the organization itself, a CSP or both.
- **Public cloud:** the cloud infrastructure is available for public use by everyone, including general public or a large industry group. The infrastructure is built and managed by the CSP and the users use this service within internet connection. This model has lack of security.