TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO.: 075MSICE017

SECURE DATA CLASSIFICATION AND MOBILITY MODEL FOR CLOUD DATA GOVERNANCE

BY

RAMESH PAUDYAL

A THESIS SUBMITTED TO THE DEPARTMENT OF ELECTROICS AND COMPUTER ENGINEEING IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTROICS AND COMPUTER ENGINEEING

August, 2021

SECURE DATA CLASSIFICATION AND MOBILITY MODEL FOR CLOUD DATA GOVERNANCE

BY

Ramesh Paudyal

075MSICE017

Thesis Supervisor

Prof. Dr. Subarna Shakya

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Information and Communication Engineering

Department of Electronics and Computer Engineering

Institute of Engineering, Pulchowk Campus

Tribhuvan University

Lalitpur, Nepal

August, 2021

COPYRIGHT ©

DECLARATION

I declare that the work hereby submitted for Master of Science in Information and Communication Engineering (MSICE) at IOE, Pulchowk Campus entitled "**SECURE DATA CLASSIFICATION AND MOBILITY MODEL FOR CLOUD DATA GOVERNANCE**" is my own work and has not been previously submitted by me at any university for any academic award. I authorize IOE, Pulchowk Campus to lend this thesis to other institution or individuals for the purpose of scholarly research.

Ramesh Paudyal

PUL075MSICE017

August, 2021

RECOMMENDATION

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for acceptance, a thesis entitled **"SECURE DATA CLASSIFICATION AND MOBILITY MODEL FOR CLOUD DATA GOVERNANCE"**, submitted by **Ramesh Paudyal** in partial fulfilment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**".

Supervisor: Dr. Subarna Shakya
Professor,
Department of Electronics and Computer Engineering
Pulchowk Campus
Institute of Engineering, Tribhuvan University

External Examiner: Mr. Om Bikram Thapa
CTO-Vianet Communication Pvt. Ltd.

Committee Chairperson: Dr. Basanta Joshi
Assistant Professor,
Program Coordinator, M.Sc. in Information and Communication Engineering, Department of Electronics and Computer Engineering, Institute of Engineering, Tribhuvan University

Date of approval: August, 2021

DEPARTMENTAL ACCEPTANCE

The thesis entitled **"SECURE DATA CLASSIFICATION AND MOBILITY MODEL FOR CLOUD DATA GOVERNANCE",** submitted by **Ramesh Paudyal** in partial fulfilment of the requirement for the award of the degree of "**Master of Science in Information and Communication Engineering**" has been accepted as a bonafide record of work independently carried out by him in the department.

Dr. Ram Krishna Maharjan
Head of the Department
Department of Electronics and Computer Engineering,
Pulchowk Campus,

Institute of Engineering, Tribhuvan University, Nepal

ACKNOWLEDGEMENT

I express my deepest gratitude to my thesis supervisor **Prof. Dr. Subarna Shakya**, for providing me the valuable supervision and feedback during this research work. I also express my sincere gratitude to **Prof. Dr. Shashidhar Ram Joshi**, **Dr. Surendra Shrestha, Dr. Nanda Bikram Adhikari, Dr. Babu Ram Dawadi and Dr. Aman Shakya** for providing me the valuable guidance and all feedback regarding entire thesis work.

I am indebted to our M.Sc. coordinator **Asst. prof. Dr. Basanta Joshi** for the valuable suggestion and guidance during this thesis work**.** I am extremely thankful to our department for providing me the opportunity to conduct the research in the field of Electronic and Computer engineering.

Ramesh Paudyal
075MSICE017

ABSTRACT

In the digital era, cloud computing has emerged as a successful computing paradigm which proposes the on-demand delivery of IT resources such as: computing power, storage and database over the internet. The adoption of cloud computing has created operational and security challenges. Cloud data governance security checklist and parameter is an essential element for measuring the computing security in a cloud. It also helps to manage data with proper security measurements and identify the security requirements. Handling all data with the same level of security measurements is not a sustainable security solution. Data centric security solutions offer sensitivity based security requirements during the mobility of data. In this thesis, a fuzzy logic based data classification model has been proposed for security management and mobility that uses cloud data governance security parameters confidentiality, integrity and availability as an input features. In addition, this model automatically identifies the appropriate security algorithm on the basis of sensitivity level, i.e. confidential, sensitive and public classes. This integrated method of data classification and securing significantly improves cloud data mobility.

Keywords: Cloud computing, Cloud Data Mobility, Data governance, Computing security, Sensitivity, Fuzzy logic.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ICT	Information and Communication Technology
IS	Information System
CC	Cloud Computing
DG	Data Governance
DLP	Data Loss Prevention
RIL	Risk Impact Level
HDFS	Hadoop Distributed File System
IT	Information Technology
MAV	Metadata attribute Value
AWS	Amazon Web Service
AHP	Analytical Hierarchical Process
HFL	Hierarchical Fuzzy Logic
RRP	Random Rotation Perturbation
NRA	National Reconstruction Authority
ISO	International Organization for Standard
MOM	Mean of Maxima
FKB	Fuzzy Knowledge Base
FDB	Fuzzy Data Base
FRB	Fuzzy Rule Base
FRBCS	Fuzzy Rule Base Classification System
ECC	Elliptical Curve Cryptography
AES	Advanced Encryption Standard
KNN	K-Nearest Neighbors
COBIT	Control Objectives for Information and Related Technology
NIST	National Institute of Standard and Technology
CSA	Cloud Security Alliances
CPU	Central Processing Unit
GPU	Graphics Processing Unit

GB	Giga Byte
vCPU	virtual CPU
EBS	Elastic Block Storage
SQL	Structure Query Language
CCP	Cloud Service Provider
API	Application programming Interface
DTRT	Data transmission Response Time
DTDT	Data Transmission Delay Time
DTT	Data Transmission Throughput
ISMS	Information Security Management System
APTs	Advance Persistence Threat

CHAPTER 1: INTRODUCTION

1.1 Background

In the era of digital world, most of the organizations adopt Information and Communication Technology (ICT) to deliver their services with more informative, efficient and secure approach [1]. Data Governance is the process of managing the availability, usability, integrity and security of the data. The organization Information System (IS) is operated by internal data standards and policies [2]. The effective data governance ensures that data is consistent and trustworthy and doesn't get misused. Organization's administrative system interacts with people, processes and technology that cause the security threat in data and information due to the complexity of technology. Migration of traditional data governance architecture into secure cloud based data governance architecture is new challenges in the field of data management and information security. Adoption of new emerging technology will be the opportunity to protect the value of any organization [3].

Implementation of cloud computing for data governance is a new paradigm but loss of data governance is the global issue. Data security is a major concern to achieve the Confidentiality, Integrity and Availability (CIA) of data due to malicious data threats that need to be used for data protection techniques. Mobility of valuable and confidential data was not sufficient in traditional security mechanisms, thus the cloud data governance required a sustainable security solution [4].

Massive amount of data is generated by the use of ICT to fulfill the organization goal and guide the direction for data driven decision making. Due to the diverse nature of data we could implement the data governance strategy for generating, managing, processing and transmission of data to achieve the Data Loss Prevention (DLP) in the cloud computing environment [5]. If data governance is not effectively implemented then the organization loses their profile, assets and goodwill. Each organization has generated sensitive data along with demanding a different level of security will be the viable solution for securing data in distributed environments. Data governing policy is essential for protecting sensitive data from breaches, loss, theft and misusing [6]. Due to the expanding trend of data

mobility in the cloud environment, the key challenge is to protect the valuable and confidential data. Automated discovery, detection and classification of sensitive data helps to reduce the risk of data loss and enable the appropriate security control based on the sensitivity level of data. The implementation of risk based metadata management enhances the security of data. Security parameters help to maintain the quality, security and privacy of an organization by defining the appropriate security classes [7].

Cloud computing is a new paradigm for providing assorted IT services that refers to on-demand network access, shared pool of configurable computational outsourcing and multi tenancy concept. It increases the privacy and security distresses and challenges besides the all services [1], [2]. Cloud storage supports data store, process and access service remotely by the use of the network [6]. Before executing private and confidential data in cloud storage we need to know about the cloud security issues or threats and cloud data governance system. Cloud data governance system to handle all the data with proper security mechanism for the purpose of use, access, store and mobility in distributed cloud environments. In cloud storage, public data does not essential any security actions but private and sensitive data needs appropriate security measurements to preserve them nonviolently. Sensitivity based data classification efficiently solves the security issues arising during the cloud data mobility for data governance application. This classification model avoids the under security and over security problems for cloud data storage services. It reduces the unnecessary overhead and processing time. Sensitive data classification model is developed by using the fuzzy knowledge base (FKB) system. The FKB system provides the flexible reasoning to solve the inaccuracies and uncertainties problem of the real world. This data classification model considers the cloud security threats and parameters in accordance with the cloud data governance system. It makes the sensitive data classification model more effective and efficient in terms of security aspects.

1.2 Problem Statement

Huge amount of data is generated because of the digital and paperless society. These data convey the organization information which plays the key role for the decision making that directly impacts the organizational goal. Rights to information add on some security threats

during the mobility of data. Data governance helps to protect the data against potential threats and guide sensitivity based mobility throughout the life cycle of data. Effectiveness of data governance is the foremost problem to enable automatic data centric security solutions based on the sensitivity level of data during the mobility. Traditional security solutions were not sufficient to prevent the data threat during the data mobility in the distributed cloud environment. Data governance deems to implement its own data governance policy for ensuring the data centric security solution before executing the data agility. Automatic discovery, detection and classification of the data with proper security action is the major problem for any organization in the era of big data. To build a highly trusted data classification and mobility model for data governance will be a viable solution for the given problem.

This thesis works maintaining the under and over security issues in the organization data. This will solve the secure data mobility in the cloud computing environment by integrating the service available in cloud computing and organization policy by the use of technology.

1.3 Objective

The overall objective of this research work is to enhance the applicability of cloud data governance by developing the secure data classification system for cloud data mobility.

- To develop the Fuzzy Rule Base System (FRBS) for sensitive data classification with security management
- To measures the performance of cloud data mobility based on the data centric security solution

1.4 Scope of the Work

The public and private organization implementing cloud computing for restructuring the data governance application. To justify secure data mobility in the cloud environment from the perspective of data governance is a key research area of this work. Defining the security parameter based on potential threat available in the data, that helps to measure metadata

based risk impact level of data. The major concern of this thesis is to provide secure data mobility by implementing integrated methods for data classification and securing in the cloud environment. That will be the sensitivity based data classification and security mechanism. This research work will not cover the data generation, data access of the data because different organizations have their own security requirements and policy for governing their data into a cloud.

1.5 Structure of Report

The thesis is organized as follows.

Chapter 1: This chapter gives a brief introduction of the security requirements scenario for the organization data on the basis of cloud data governance security parameters.

Chapter 2: In this chapter, a background studies and literature review on cloud data governance and its security parameter, different classification method for sensitive data classification and fuzzy rule base classification approach for data classification and mobility model.

Chapter 3: In this chapter, explains the methodology used in the thesis work. The fuzzy classification system, model development, datasets, implementation framework and experimental setup as well as validation methodology.

Chapter 4: In this chapter, covers experimental results, analysis and validation with discussion.

Chapter 5: In this chapter provides the research contribution on this work as well as explains the limitation and further research direction.

CHAPTER 2: BACKGROUND STUDIES AND RELATED WORK

This chapter provides the study of theoretical, methodological and technical contributions in a particular topic that have been followed in the scope of this thesis work. This section also demonstrates the research gap and understanding of some of the existing research work and their results with analysis. These activities provide the starting point of this thesis work or help to define clear problem statements in a given particular research domain. The following subsection provides a detailed discussion on cloud data governance and their security issues, some existing methods related to sensitivity based data classification and fuzzy rule base system for data classification approach.

2.1 Background Studies

2.1.1 Cloud Data Governance and Security Issues

Cloud data governance security checklist helps to measure the computing security in the cloud computing. Cloud vulnerability is the major security issue that causes the malicious threat on personal and confidential data. The implementation of cloud data governance assists to reduce the risk associated with cloud computing and helps to manage data with correct procedure. Cloud service providers and consumers faced massive threats and attacks especially in IaaS service layers. This layer deals with the storage, server and networking services. The security issues of IaaS layers is a serious problem for data breaches and data losses.

[13] Describes the significance of implementation of data governance in any organization. The author emphasizes the absence of confidentiality, integrity and availability causes loss of control of data. They also explained, who want to move their data in the cloud they should implement cloud data governance rules. These rules are set on the basis of data owner requirements. National Institute of Standard and technology (NIST) also advised companies for the implementation of cloud data governance to move to cloud platforms.

The researchers [5], [6], [7] have suggested that a security checklist is used to measure the security level of cloud service providers (CSPs) and cloud computing service (CCSs). These checklists are useful for good choice of CSPs and CCS and move their data into the cloud computing environment.

In [9] their work, Cloud computing is changing the way organizations are functioning nowadays. A Cloud based system changes the services seamlessly without expending many resources in setting up new systems. Although there are many advantages of Cloud computing, issues related with Security and Privacy are some of the major challenges, which needs to be addressed for the successful deployment of a Cloud based System. Security has become a key limitation in the development of a cloud based data Governance System. Therefore, the security of the system should be assessed regularly to ensure reliability and confidentiality of the System [10].

In [11] their work, data governance focused on the use of metadata. It can, for example, be used to provide an audit trail for data collection. This enables enforcement managers to show that they are appropriately managing data. Important personal data, for example, must be kept confidential, with access restricted to certain people. Freedom of information law, on the other hand, may demand that unique information services be provided with a preservation schedule and publishing scheme. Any metadata specifications have data components that are expressly designed to keep track of a document's audit trail [12].

Based on the above study [10] there are 27 security threats and 8 security parameters are identified in IaaS layers. The confidentiality, integrity and availability are three parameters that are mostly focused for cloud data governance applications. The CIA triad is widely used in references for developing security guidelines, security checklist for the various treatment of organizational data [3].

2.1.2 Cloud Data Mobility

Nowadays organizations have moved their data in a cloud environment for the storage, computing and treatment purposes. Cloud computing provides the availability of data when and where needed. Cloud data mobility causes a malicious threat of data that requires the

proper data protection strategy or policy. The appropriate security management is essential to reduce the under and over security problem for cloud data mobility.

In [6] their work, performed data classification and automatic security management of data based on the sensitivity. The mobility of data with appropriate security mechanisms reduces the security threat on data. However, many research efforts have been dedicated to cloud data mobility issues. In this [8] paper discusses full cloud mobility management from two aspects i.e. data mobility management and user mobility management based on the previously proposed data centric security framework. That also minimizes the threat of data by the use of different data protection techniques. Traditional security mechanisms were not sufficient to handle cloud data mobility to maintain the sensitivity level of data. In this paper implement the data classification and security model to reduce the potential threat available in the data.

The above literature analyzes the problem and effectiveness of cloud data mobility. The mobility of data in cloud computing is measured by evaluating in terms of data transmission response time, throughput and delay. We can conclude that the secure data transmission method causes redundant network bandwidth and computing overhead due to the extra transmission of data. When adding an extra security algorithm that increases the response time and delay time. That causes the degradation of throughput so, the same level of security management for all data is not applicable before executing the data mobility in a cloud database. The choice of appropriate data governance security management method is the major research gap for cloud data mobility that will provide the minimum degradation of the cloud system with appropriate security actions. Some of the literature said sensitivity based data classification approach contributes to the overall performance of the system that is clearly demonstrated in [8]. To investigate the appropriate security management model for cloud data mobility is the key research scope of this thesis. This method will be able to move organization data by considering the security parameter of cloud data governance i.e. confidentiality, integrity and availability of data while before

executing over cloud storage. That also delivers the maximum performance improvement by avoiding the redundant encryption process for public dataset.

2.1.3 Sensitive Data Classification with Security Algorithm

Sensitivity based data classification method provides the data centric security solution for maintaining the appropriate security mechanism that minimizes the security overhead on the data. There are many techniques used for sensitive data management that have been demonstrated in this section.

In [6] also stressed that to achieve the secured information exchange mechanism in cloud environments, AES techniques are implemented which guarantee the user's information security located on cloud servers. A data classification cloud model was proposed to attenuate the shortcoming of data confidentiality issues based on the distributed cloud environment. To measure classification of data as two main streams i.e. Sensitive and Non-sensitive data was classified using k-Nearest Neighbor (KNN) classifiers and Rivets, Shamir and Adelman (RSA) algorithms were used to successfully encrypt data labeled as sensitive data.

In [8] there work was solved the data classification and security issues based risk impact level (RIL) of data. It deals with big data classification and security issues in cloud environments. Hadoop Map Reduce framework was integrated in cloud environment to solve the security issues available during the data mobility in distributed environment. This method can significantly improve the data security based on the availability of data. This also reduces the potential threat observed in the data by managing the Risk based security algorithm. In this method data are classified on the two classes i.e. sensitivity and public.

In research [15] and [16] clearly explained the need of security management during the data storage, mobility and exchange for to safeguard the integrity, confidentiality and availability of data as they are located in a wide range of hardware specifications. Security algorithms which employ cryptographic techniques were used to encrypt and decrypt the data. Security algorithms were implemented on the basis of their sensitivity level during

storage and transmission. On the one hand, the implementation of cryptographic techniques boosts data security and improves computing security. While, on the other hand, these implementations of security algorithms come at the cost of high resource utilization in terms of time, memory, and CPU usability. Three encryption techniques, viz. AES, DES and RSA were implemented with a comparison based on time of encryption and decryption, thus providing the effectiveness of these algorithms.

In [8] their work, at first, the data classification was performed using the MAV in cloud computing environments. Then, we studied the comparative analysis of security algorithms based on the encryption and decryption time, file size and total execution time. The RSA algorithm is used for the encryption of highly sensitive data due to its complexity in deciphering by intruders [8]. As security is our primary concern, even before the exchange, the use of RSA algorithm is justifiable even though we have to tradeoff file size and query execution time. Since our system is implemented in a cloud environment there is no limitation of storage and processing capacity [15], [16]. Thus, implementing the RSA algorithm won't have a significant impact on cloud-based e-Government System. The data exchange between two databases is achieved on the basis of user privilege, which further boosts the security management and mobility based on the sensitivity level of data. Hence, to achieve a highly secured cloud government system needs to consider data classification and appropriate security management before storing and mobility of data in distributed environments.

In the above literature, scholars have used machine learning and multi-criteria decision-making methods. In a study, [8] the Metadata Attributes Values (MAV) is used to classify data based on the Risk Impact Level (RIL) during data mobility in the cloud computing environment. In this paper, we propose cloud data storage and mobility methods by employing the appropriate security mechanism. This model classifies the data based on the security requirements of the data owner. The appropriate security requirement is calculated by using the MAV function [8] on the basis of the risk impact level of organizational data. To solve the problem we need to design and develop an integrated methodology for sensitive data classification with appropriate security solutions in the cloud computing environment.

This methodology aims to fulfill secure cloud data mobility for the e-governance application. This model has been quantitatively and technically analyzed by calculating the data transmission throughput during the mobility of data. This will perform secure data mobility in the cloud environment by integrating the cloud computing and organizational policy.

2.1.4 Rule Base Fuzzy System

RBFS uses an approximate reasoning mechanism that is able to express the ambiguity and subjectively present in human reasoning. Fuzzy system is developed to consist of fuzzy sets, fuzzy logic and corresponding mathematical relations. The all parameters of the fuzzy system are described in the section below [17].

Fuzzy logic: Fuzzy logic modal for the development of nonlinear systems. It provides a transparent representation of the system by the linguistic interpretation as a rule. Fuzzy logic is suitable for such systems that cannot be precisely described by mathematical models. It is used for modeling with the problem of uncertainty and imprecision. It is used to close interpretation of human thinking and knowledge [17].

Fuzzy sets: Fuzzy sets are the classical sets with characteristic function values between 0 and 1. The membership value is limited between 0 to 1. The fuzzy set is represented by given equation.

$$A = \{(A(x)/x) | x \in X\} \quad (2.1)$$

The core concept of the fuzzy set theory is a membership function. The membership function described in the form of $A_x: A \rightarrow [0,1]$ this is asociated with the fuzzy set. The extensively used membership functions are generally represeted in the trapezoidal, triangular and S or guession [13].

Fuzzy rules: Fuzzy rules describe the Implication between fuzzy propositions. It is represented by the if-then rules as follows:

If {antecedent proposition} then {consequents proposition}

$$R_i = if\ x\ is\ A_i\ than\ y\ is\ B_i \quad (2.2)$$

The term x is the antecedent linguistic variable and A_i is the linguistic term and y and B_i are the consequent linguistic variables and linguistic term. In general, the relations induced by fuzzy rules are derived from fuzzy conjunction, fuzzy disjunction and fuzzy implication [14].

Fuzzy reasoning: To decide the conclusion from fuzzy rules we need the inference algorithm that is called the fuzzy reasoning. We need to calculate the fuzzy relation for each rule i.e. denoted by $R_i = (X \times Y) \rightarrow [0,1]$. The relation of input and output variables of the rule is calculated by the minimum (conjunction) operator $(\wedge)$ from the fuzzy rule. $R_i = A_i \times B_i$, that is represented in the given form $\mu_{Ri}(x,y) = \mu_{Ai}(x) \wedge \mu B_i(y)$, this can be computed in individual rule [18].

For the entire model, a fuzzy relation is computed by the use of a disjunction or maximum operator that is computed by a given equation.

$$R = \sum_{i=1}^{K} R_i, \mu_{Ri}(x,y) = Max[\min({}_i\mu_{Ai}(x) \wedge \mu B_i(y))] \qquad (2.3)$$

Fuzzy inference mechanism: Fuzzy inference system known as fuzzy model. It is a computing framework that includes fuzzy sets, if-then rule and fuzzy reasoning concept. Two basic category of fuzzy model are:

i. Mamdani
ii. Takagi -Sugeno-Kang (TSK)

Fuzzy model consists of a fuzzy rule base, which induces the fuzzy rules and uses the membership function and then inference mechanism to decide the conclusion from fuzzy if - then rule [23].

Defuzzification: Defuzzification is the process of converting a Fuzzifier output into a single crisp value with respect to a fuzzy set. The Defuzzifier value in FLC (Fuzzy Logic Controller) represents the action to be taken in controlling the process.In a defuzzification, where the crisp value is assigned to the aggregated result of the system. This value should be selected to represent the best result of the systems [23].

i. Mean of maxima (MoM)
ii. Center of Gravity (CoG)
iii. Center of sums method (COS)

2.2 Related Work

In the literature, different methods and techniques are used for the data classification for security management in the cloud computing environments. In our proposed method, we used fuzzy logic for sensitive data classification for data governance application. Thus, we demonstrate the related work regarding fuzzy logic for data classification.

In [17] their work, their practice emphasizes the importance of data protection and privacy in effectively conducting day-to-day and long-term business activities. The use of fuzzy logic-based data classification to decide the necessary data protection standards based on different government and business policies is an innovative method for achieving data security. Organizations may use data classification to help them make better decisions. Data from the organization must be assessed. This paper investigates fuzzy logic-based categorization. The use of a hierarchical fuzzy logic (HFL) classification system for categorizing organizational data based on governmental and company data policies is also discussed in this article. The use of an HFL to increase data classification is shown in this article. HFL decreases the necessary file size. This paper shows how an HFL system can be used to increase data protection, privacy, and user access control by determining data and user access control based on the sensitivity of the data, government and corporate data security standards, and user level security controls .

In [21] their work, emphasis on data classification in an organization is a basic necessity for proper information protection and management according to [12]. The repercussions of failing to properly enforce a data classification system in an enterprise can be serious and expensive, both financially and in terms of the company's image. This approach also focuses on a proper data classification based on fuzzy logic.

In [26] their work, Fuzzy based data transformation approaches are used in their work to enforce privacy-preserving clustering in a centralized database setting. In case one, a fuzzy

data transformation approach is proposed, and various tests are carried out by changing the fuzzy membership functions from Z-shaped fuzzy membership function, Triangular fuzzy membership function, and Gaussian fuzzy membership function to transform the original dataset [18]. A hybrid method is proposed as a variation of the fuzzy data transformation approach defined in case one and Random Rotation Perturbation in case two (RRP).

In [32] their work, demonstrate the cloud data classification model by using the fuzzy logic approaches. This model has three input variables and single output variables in a fuzzy system. These input and output variables are the linguistic terms or variables. Which represents the degree of confidentiality, Integrity and availability. The output variables denotes the degree of sensitivity or class of data i.e. Confidential, Sensitive and Public. CIA triad is used to classify data according to owner security requirements.

2.3 Research Gap

In this section we perform the comparative analysis of the related works and finalize the research gap and problem associated with the information security management system. From the review of relevant literature as discussed above, we can conclude that data security in the distributed environment is the major issue. Data governance tries to protect organization data against potential threats and monitor the data by employing the risk-based security activities throughout the life cycle of data [8]. The effectiveness of data governance is the foremost problem to enable automatic data-centric security solutions due to the rapid technological advancement [10]. Traditional security solutions were not sufficient to prevent the data threat during the data mobility in the distributed cloud environment [8]. Cloud data governance applications seem to implement their own data governance policy for ensuring the data-centric security solution before executing the data agility. Automatic discovery, detection, and classification of the data with proper security actions are the major research gap in this domain [8], [9] and [10]. Building a highly secured data classification and mobility model can fill this research gap. To develop the appropriate data security model for cloud data governance with minimum data transmission throughput.

CHAPTER 3: RESEARCH METHODOLOGY

For successful work, it is necessary to formulate a research methodology. It provides planning, design, implementation, validation and documentation will be carried out for this thesis work. In this thesis, FRBS is used for the secure data classification with mobility in a distributed cloud environment. The fuzzy system is implemented in the field of cloud data governance that helps to manage security requirements of the data based on the Sensitivity level. All the attributes are scaled by using the organization policy and security checklist cloud data governance [10]. The organization expert team can only define the security requirements of the dataset under the consideration of organization requirements.

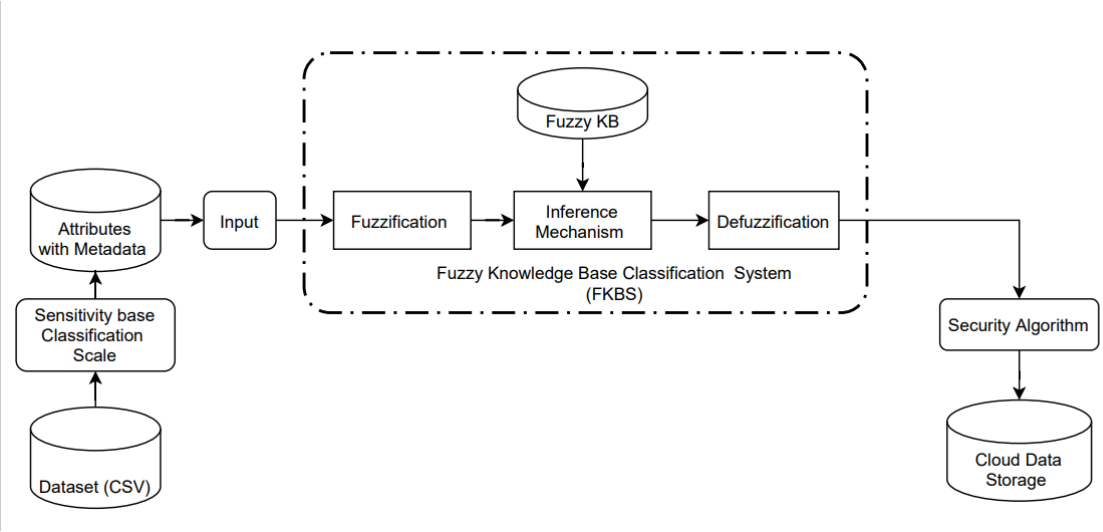

Figure 3.1: Proposed Methodology

Finally, data has been automatically classified and stored in a cloud database according to a fuzzy rule base system based on appropriate security management [17]. Fuzzy based classification systems are able to deliver highly accurate and interpretable models for imprecise and uncertain real world problems. The cloud data governance security

checklists are important for measuring the computing security in a cloud computing environment. It helps to manage and organize organization data with correct procedure.

3.1 Dataset

Dataset will be collected from the government organization which carries the confidential information of the citizen that should be protected during the different stages of data mobility. The National Reconstruction Authority (NRA) has provided a dataset for the purpose of research work which includes various information about the citizen. The description of the dataset is as follows: Attributes Id_number, name, district, Gp_np, ward number, house owner name, gender, Age, ctz_number, nationality, phone number, latitude and longitude.

We implemented a given dataset in RBFS for the classification while maintaining the security measurement according to the sensitivity level of data.

Table 3.1: Dataset Description

File size (GB)	No of Attributes (Column)	No of instances (Rows)	No of class
2	16	12000000	3

In our dataset, 6 attributes are the confidential class, 7 are the public and finally 3 are the sensitive class. We have 16 numbers of attributes with three different classes.

3.2 Data Preprocessing

Dataset should be preprocessed on the basis of input and output requirements of the proposed fuzzy model. The input features are used to describe the attributes of the dataset on the basis of security requirements using the security parameter of cloud data governance. Output classes represent the degree of sensitivity or security measurement that are described in the below sections.

3.2.1 Classification Scale

In this method classification is done for the protection of data based on the sensitive level of data. It is able to describe the strategic importance of data through security parameters of cloud data governance [19]. It considers the cloud data governance security checklist and guidelines for defining classification scale and measuring the security level of data. It could ensure the appropriate security mechanism for data without loss of organization information and business process [10]. DG is not properly implemented; there arises lack of integrity, confidentiality, loss of data and data breaches due to the threats available in the cloud computing environment.

3.2.2 Input and Output Variables

In this method classification is done for the protection of data based on the sensitive level of data. It is able to describe the strategic importance of data through security parameters of cloud data governance [5]. It considers the cloud data governance security checklist and guidelines for defining classification scale and measuring the security level of data. It ensure the appropriate security mechanism for data without loss of organization information and business process [10]. DG is not properly implemented; there arises lack of integrity, confidentiality, loss of data and data breaches due to the threats available in the cloud computing environment.

Secure adoption of cloud computing enforces the organization moving their computing infrastructure in a cloud. Infrastructure security is a major concern for securing the information of the organization in cloud computing environments. The infrastructure threat includes manmade and natural disasters and deliberate threats. Deliberate threats are defined based on the cloud data governance security checklist. This security checklist helps to measure the computing security in IaaS layer [10]. Attackers can attack the IaaS layer of the cloud computing; this attack could affect the other PaaS and SaaS layers. From [9], we take 27 security threats and three security parameters for cloud data governance. These security parameters have been finalized under the consideration of cloud data governance

security checklist for IaaS layer [10]. The details of the security parameters are as follows which is used for the measuring the security requirements of data:

Confidentiality: it ensures the protection against the disclosure of information without any legitimate process. Loss of confidence of public information can be minimized by preventing unauthorized access to sensitive data. Classification ensures the confidentiality of data based on the sensitivity level. Finally we achieved the confidential data mobility in a distributed cloud environment with a high confidence level.

Integrity: Safeguarding the information from improper modification ensures the information authenticity. It permits the mobility, update, modify and delete the sensitive data only authorized users. Secure data mobility by employing the encryption techniques based on the sensitivity level of data classified data that ensures the integrity of the data.

Availability: timely and reliable access and use of information called availability. Sender cloud and receiver cloud shared their user credentials for transmitting data from sender to receiver. It concerns the effects of publishing, interactive and transaction processing and service delivery of the organizational data. It guarantees the data is available and accessible anytime from anywhere.

These security parameters are used as an input parameter in our FKBCS for sensitive data classification. That can be identified by the risk analysis of the organization, rating of risk analysis helps to decide the sensitivity level of data. Inaccurate risk analysis causes the misclassification of our system. The scaling of the input parameter makes the data classification more flexible and accurate.

CDGSP= (Confidentiality, Impact), {(Integrity, Impact), {(Availability, Impact)}

CIA is used as reference for developing the classification scale for FRBCS, because it covers the security threat of cloud data governance. This parameter carries the impact of security level and needs of security measurement according to threat. This knowledge is imposed on attributes of our dataset. Each classification scale is described by the impact level that has been mentioned in tabular form.

Table 3.2: Scale for Confidentiality

Scale	Needs	Impact
High	Information having a significant impact on the organization if it were to be communicated outside the persons specifically named	3
Medium	Information intended to remain within the organization. Communication outside the organization can harm the business.	2
Low	Information that can be made public without feared impact on the entity or the organization.	1

Table 3.3: Scale for Integrity

Scale	Needs	Impact
High	Intolerable loss of integrity. Any alteration of data would have a high impact on the company	3
Medium	Any alteration of data would have a significant impact on the company	2
Low	Tolerated loss of integrity. Any alteration of data will have a low impact on the company	1

Table 3.4: Scale for Availability

Scale	Needs	Impact
High	Very low tolerance for unavailability. If this need is not met, the company has a high impact	3
Medium	Tolerance for average unavailability. If this need is not met, the company has a significant impact	2
Low	High tolerance for unavailability. If this need is not met, the company has a low impact	1

Table 3.5: Impact of Description

Scale	Description with Need
High	A concern with confidentiality, integrity or availability could have a limited negative effect on the activities of the company, the assets of the company or for individuals
Medium	An affection of confidentiality, integrity or availability could have a serious negative effect on the activities of the company, the assets of the company or for individuals
Low	A concern with confidentiality, integrity or availability could have a catastrophic adverse effect on the business and the assets of the business or for individuals

Classification output

Information and data are the critical assets for any organization that needs to safeguard lawfully. In this work, Information Security Management System (ISMS) is implemented for identifying and defining a criteria for assets management and protection. Data classification based on the sensitivity level is one major criteria for security management of the organization data [12]. This classification strategy ensures the secure data storing, collecting and processing over the distributed environment. That also helps to mobility of data with appropriate level of security protection according to level of sensitivity. Data classification based on the sensitivity level is essential practice for public and private organizations. There are various standards and systematic approaches available for data classification. The essential requirements for data classification needed detailed description about the classification procedure, approach and guidance [19].

Different classification labels fill the different security requirements for data based on the sensitivity. These requirements overcome the under and over security condition of organization data. Different classes of data can treat appropriate security solutions by employing the security algorithm in a particular class of the data. In this work, solve the security issues in cloud data mobility by classifying the data according to priority, needs and degree of protection based on criticality of data. This proposed classification system is classified into three: public, sensitive and confidential class [8]. Classification label describes the subjective judgment of the organization assets or data that causes the inconsistencies classification. This subjective judgment is undertaken by using the various international and national standard policies. ISO/IEC 27005 (2018) is used for the labeling of data in terms of sensitivity [27].

Labeling is the well managed consequences of the data classification. Addition of metadata in our dataset also describes the labeling of data. We can add numeric value for our dataset for the classification purpose that identifies and defines an appropriate class of data based on the sensitivity level.

Public: it contains the general information of the public which can easily be viewed and accessed by any user without any restrictions of data. These data are open for public access

and are not protected. This data does not need any security mechanisms and benchmarks for the protection.

Sensitive: this class describes the moderate level of risk of data. This data is not explicitly restricted to the people but within the organization can be easily accessible for data processing. It requires a reasonable level of security action for control of the organizational profile as well as citizen information.

Confidential: This contains the important information of the citizen that view and accessed by only authorized person of the organization. This data are averted by international and national standard, agreement and legislative laws. This classes of data needs the high level of security for prevent loss of organization profile as well as citizen information.

3.2.3 Scaled Input

This scaled input is used as a model input which processes in the range between 0-100. These values are not able to be defined exactly so fuzzy logic is used to process as a linguistic term which is explained table 6.

Table 3.6: Scaled Input for dataset

Attributes	Input variables	Linguistic processing Subset
PA_Number	C	H (0-35), M (25-65) L (55-100)
	I	H (0-35), M (25-65) L (55-100)
	A	H (0-35), M (25-65) L (55-100)
BENIFICARY_NAME	C	H (0-35), M (25-65) L (55-100)
	I	H (0-35), M (25-65) L (55-100)
	A	H (0-35), M (25-65) L (55-100)
	C	H (0-35), M (25-65) L (55-100)

PHONE_NUMBER	I	H (0-35), M (25-65) L (55-100)
	A	H (0-35), M (25-65) L (55-100)

3.3 Fuzzy Rule Base Classification System

Classification plays a significant effort for data protection and makes it possible to characterize the strategic importance of the dataset. The classification ensures the level of data protection without impeding the flow of information and business process. The cloud data governance security parameter confidentiality, Integrity and availability are used for defining a classification criteria. The output of the fuzzy rule classification system represents the degree of sensitivity i.e. confidential, sensitive and public [17], [32]. The input variables are used for defining the features in their antecedent and output class is their consequent part. A fuzzy classification can be expressed by:

$R_K = IF\ X_1 is\ A_1\ AND\ IF\ X_2 is\ A_2\ AND\ \ldots\ldots AND\ X_m is\ A_m\ THAN\ Class = c_i$ (3.1)

Where, R_k is the rule identifier, $X_1 \ldots, X_m$ are the input linguistic variables, $A_1 \ldots . A_m$ features value of linguistic variables and $c_i \in C$ is the output classes.

3.3.1 Fuzzy Knowledge Base

Fuzzy knowledge Base (FKB) stores a fuzzy IF-THAN rule which is provided by experts. It consists of a fuzzy database and a fuzzy rule base. The figure 3.2 describes the fuzzy knowledge base for the purpose of sensitive data classification.

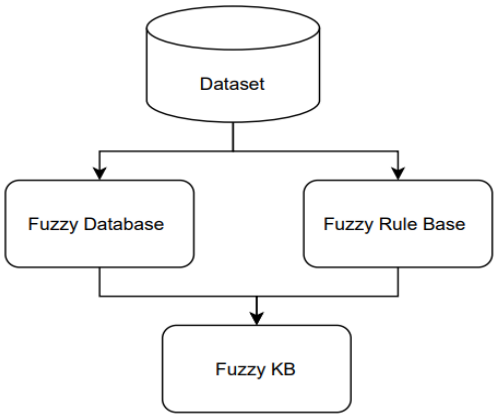

Figure 3.2: Generation of FKB by Fuzzification

3.3.2 Fuzzy Database

It is the definition of attributes also known as features in terms of fuzzy sets. It is the process of defining various numbers of fuzzy sets for each attribute of the datasets. Fuzzy database contains the definition of fuzzy sets related to the linguistic variables that are used in the fuzzy rule base.

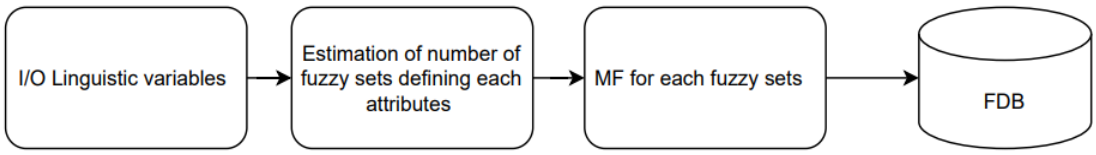

Figure 3.3: Fuzzy Database Definition Process

The fuzzy database can be developed by the use of following steps which are described in the below sections.

3.3.2.1 Linguistic Variables

In our FKB system, each attribute has a three number of input variables that represent confidentiality, integrity and availability. These three input variables represent the degree of CIA which is imposed by the data owner or expert knowledge. Degree of CIA has been established according to the cloud data governance security checklist [8], [32], [10]. Each number of input variables have their subsets for the purpose of defining the membership function of the linguistic variables. This given table shows the linguistic variables for the input and output of our system. Each attribute of our data set has deals by using the linguistic variables of table 7 and output class levels can treat with the table 8. These variables were decided based on the policy of different professional bodies which are Cloud Security Alliance (CSA), COBIT, NIST and ENISA [9]. These input and output variables represent the classification scaling and are used for the purpose of defining the linguistic variables of the system [10].

Table 3.7: Linguistic term for input variables

Input Linguistic Variables	**Subset**		
Input 1: Confidentiality (C)	Low	Medium	High
Input 2: Integrity (I)	Low	Medium	High
Input 3: Availability (A)	Low	Medium	High

The only one output variable represents the degree of sensitivity or levels of classification that are confidential class, Sensitive Class and Public class. Output linguistics variables and their subsets are used for identifying the classification levels of our system. Table 8 shows the output linguistic variables and their subsets.

Table3.8: Linguistics Term for Output Variables

Output Linguistic Variables	**Subset**		
Output: Sensitivity (S)	Low	Medium	High

Classification level (Degree of sensitivity)	Public	Sensitive	Confidential

3.2.2.2 Membership Function

After defining the linguistic variables for each attribute with their subsets. We define a membership function for each subsets of the input and output variables. Linguistic variables are used for estimating the number of fuzzy sets. Membership function (MF) represents the linguistic variables in terms of fuzzy sets or values by partitioning the given variables. In this work we use a trapezoidal membership function for input variables and triangular for output variables.

There are different classes of fuzzy membership functions, but triangular and trapezoidal fuzzy membership functions are commonly applied in real world problems. The triangular membership is given by equation (3.2) [17].

$$f(x,a,b,c) = max\left(\min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right),0\right) \quad (3.2)$$

Here, a, b and c are use params the membership function for the each output variables were computed by given equation.

Trapezoidal membership function will be used for the input variables in our model. Each subset of input variables will be use a membership function so, we need to define the n membership function for each subset of input attributes. Given equation (3.3) explains the purpose membership function [16].

$$f(x,a,b,c,d) = max\left(min\left(\frac{x-a}{b-a},1,\frac{d-x}{d-b}\right),0\right) \quad (3.3)$$

Here, a, b, c and d are params, the membership values are computed for each input variables and their subsets.

Definition of fuzzy database is defining the fuzzy sets that modeled the linguistic variables. In this thesis work, $X_1, \ldots\ldots, X_n$ are the linguistic variables for the input or rule antecedent which is describe by the security parameter and their subset. Y Indicates the linguistic variables for output or rule consequent which is described by sensitivity level

and their subset. $C(X_i) = \{v_i^1, \ldots\ldots\ldots, v_i^{k_i}\}$ Represents the subset for linguistic variablesX_i. $\mu_A(x)$ Represents the degree of membership value x to fuzzy sets. The fuzzy database for our work is define by equation (3.4) [17].

$$(D_b) = \{(x^1, y_1), \ldots\ldots\ldots (x^d, y_d)\} \quad (3.4)$$

Where, x^i is a real values for inputs variables and y_i is the real values for the fuzzy output variables.

3.3.3 Fuzzy Rule Base

Fuzzy set theory for the computational representation and processing of imprecise and uncertain information. Fuzzy system work in the domain of imprecision and uncertainty that cannot directly process by available computing approaches. FRBS is use approximate reasoning mechanism that can able to express the ambiguity and subjectively present in the human reasoning.

Syntax for Fuzzy Rule Generation

It capture the imprecise knowledge from human expert and express their knowledge in terms of directives and strategies. The fuzzy rules are represents given form:

IF (ANTECEDENT) **THEN** (CONSEQUENT)

This fuzzy rules can established a relations among the variables of antecedent and consequents clauses.

Antecedent: It contains the input variables with linguistic term that represents by fuzzy sets or degree of membership. It always indicates a fuzzy proposition or conjunction/disjunction.

Proposition: It represents the membership value of fuzzy sets. It is calculated by the use of logical connectives. Logical connectives is used to define the fuzzy proposition i.e. disjunction and conjunctions. The following proposition are used for this thesis work which are describes in the below.

Atomic: atomic proposition are used to convert the input and output linguistic variables into fuzzy set value.

Compound: Compound proposition is made by conjunction and disjunction of atomic proposition. Fuzzy rule base considered the conditional proposition which is express in above syntax. To generate the fuzzy rule by the use of compound proposition in both the antecedents clauses and consequents clauses. Proposition for antecedent part express in following form:

IF C=LOW AND I=MEDIUM AND A=HIGH

Proposition express the knowledge of security parameter with appropriate scaling and its membership degree.

Conditional: Fuzzy rule uses a conditional proposition but both clauses of fuzzy rule can be formed by using a compound proposition. That includes the finite number of atomic proposition which is connected by conjunction or disjunction.

IF C=LOW AND I=MEDIUM AND A=HIGH THEN CONFIDENTIAL

Assumes P and Q are the proposition for antecedent and consequent part of the fuzzy rule. That is expressed in the form: **IF P Then Q**

Proposition P can be represents the input variables of the system which is confidentiality (C), Integrity (I) and availability (A). It is represented in the give form:

C is L **AND** I is M **AND** A is H

Proposition for consequents represents output variables or our system which is class level of dataset. Which can be represented in given form:

P is L **AND** S is M **AND** C is H

L, M and H are the fuzzy sets for input variables Confidentiality, Integrity and Availability similarly output variables Public, Sensitive and Confidential.

The rule can be execute the relation between variables of input and output which is express by P and Q. This is characterized by degree of membership pair (p, q) ∈ (P x Q) [18]. This degree of membership can established the relationship among the input and output fuzzy sets by using the membership function.

$$R(p,q) = f\big(A(p), B(q)\big), \forall (p,q) \in P \times Q \qquad (3.5)$$

The function can express in the given form:

$$f: [0,1]^2 \rightarrow [0,1] \qquad (3.6)$$

Table 3.9: Decision Matrix for Fuzzy Rule Generation [10], [32]

Input linguistic variables	**Output Class (sensitivity)**		
	Confidential	**Sensitive**	**Public**
Degree of Confidentiality (C)	High	Medium	Low
Fuzzy Conjunction/Disjunction	OR/AND	OR/AND	OR/AND
Degree of Integrity (I)	High	Medium	Low
Fuzzy Conjunction/Disjunction	OR/AND	OR/AND	OR/AND
Degree of Availability (A)	High	Medium	Low

After completing all above steps of the fuzzy knowledge base we defined the fuzzy rule base i.e. called inference rule. In a fuzzy system linguistic terms of input and output variables are used to generate the rule base for the reasoning purpose. These rules are represented in terms of matrix decisions; see in table (10).

3.3.4 Fuzzy Inference Engine

It is a reasoning system based on the concepts of fuzzy set theory. It is used for extracting accurate conclusion from approximate data [17]. There are many types of inference engine

used in fuzzy system. In this work, we choose Mamdani minimum inference engine that are used AND by MIN and OR by Ma.

The relations is induced by the processing of conjunction function, disjunction and implication function. In our work we use a fuzzy conjunction that is denoted by minimum (⋀) operator or algebraic product (·) or t-norm or definition of Mamdani [21].

$$f_m\big(A(p), B(q) = MinA(p) \wedge B(q)\big), \forall_{x,y} \in P \tag{3.7}$$

For the consequent purpose we use the disjunction function. It denotes the s-norm or t-conorms. That is express by given function.

$$f_M(A(p), B(q) = \max(f\big(A(p), B(q)\big), \forall (p, q) \in Q \tag{3.8}$$

Fuzzy implication: This implication function is characterized by classic logic and intuitionist logic. It can be used when multiple variables of antecedents and consequent part of the fuzzy rule. This can be expressed following form:

$$P(x_1, x_2, \dots\dots, x_n) = A_1(x_1)\ t\ A_2(x_2)\ t\ \dots\dots t\ A_n\ (x_n)$$

$$Q(y_1, y_2, \dots\dots, y_n) = B_1(y_1)\ t\ B_2(y_2)\ t\ \dots\dots t\ B_n\ (y_m \tag{3.9}$$

Rule can be induced or defined by given form:

$$R(x_1, x_2, \dots\dots, x_n, y_1, y_2, \dots\dots, y_m = f(P(x_1, x_2, \dots\dots, x_n), Q(y_1, y_2, \dots\dots, y_n)) \tag{3.10}$$

The rule is formed by the use of the connection OR instead of AND. We use a t-norm for the calculation of relation among input and output variables of P and Q.

3.3.5 Defuzzification

It transform fuzzy set into crisp set with accurate results so, we need to choose the appropriate defuzzification mechanism in fuzzy system [19]. We have different defuzzification methods among them in this proposed model, we have choose the average method mean of maxima (MOM) for the defuzzification purpose. It defines the output decision on the amount of the tip as being the average of the abscissae of the maxima of

the set or resulting from the aggregation of the conclusions. The equation (3.11) describes the representation of MOM defuzzification approach [18].

$$x^* = \frac{\sum x_i \in M^{x_i}}{|M|} \tag{3.11}$$

Here, $\{M = \sum x_i | \mu_A(x_i)\}$ is equal to the height of the fuzzy set and |M| is the cardinality of the set.

This model are uses when the universe contains finite number of elements, in such a condition it gives the meaningful results.

3.4 Security Management

Automatic security management of organization data based on the sensitivity level is one of the major tasks of this thesis work. In previous section data were classified based on the sensitivity level of data into three i.e. public, sensitive and confidential classes that need different security measurements [10], [32]. Cloud computing provides many features like data storage and processing remotely via network. Cloud data mobility needs the appropriate security mechanism for protection of the data. Same level of security management is not sustainable solution for all the data. When classification task has been completed then, data has move to the cloud database with proper security actions. There are many security algorithm has been used for securing the data in the cloud computing.

To enhance further security on classified data, we take appropriate security action based on the sensitivity of the data. For the analysis purpose, we implemented different security algorithms, i.e. AES, DES, RSA and ECC in the distributed cloud databases [6], [15], [16], and [32]. We measured the encryption and decryption time along with the file size of data before and after classification. These security algorithms ensure the integrity during the mobility of data in cloud databases. In this thesis work, we choose the different asymmetric and symmetric algorithms for securing the data based on the sensitivity level. ECC was considered because of its asymmetric nature while AES were selected based on the symmetric nature of these algorithms.

Elliptical Curve cryptography (ECC)

ECC is a latest encryption method with stronger security. It is used to encrypt the data of confidential class. This uses the asymmetric secret keys for data encoding. In this system public key is generated by using the private key. The ECC method is expressed by the algebraic structure of elliptical curve. It uses a comparatively small size of secret key that needs higher computational resources. It also provides a faster then other asymmetric security algorithm. The performance of 256-bit ECC is equal to 3072-bit RSA key because it uses short keys [16]. It works on the less computational power and provides a secure and fast connection. This algorithm works based on the point's lies on the curve which is define by the public and private key.

Advanced Encryption Standard (AES)

The Advanced Encryption Standard (AES) algorithm is used for both security and speed. The installation of both hardware and software is much quicker. NIST recommends a new encryption standard to replace DES. It shows how to encrypt 128-bit data blocks in 10, 12, and 14 rounds based on key length 128, 192, 256 [. It contains a single S- box and identical algorithm applicable for decryption purpose. It can be seen on a variety of networks, including mobile devices. It has been extensively tested for a number of security applications. AES provides great flexibility for implementing based on parallel structure with effective resistance against attacks.

3.5 Implementation Framework

Proposed methodology has been implemented in the given architecture that is describe in the figure 3.4. This focus on how the model work, used technology and tool, software packages and library implementation environment evaluation method. This clearly describes the details implementation of purposed methodology.

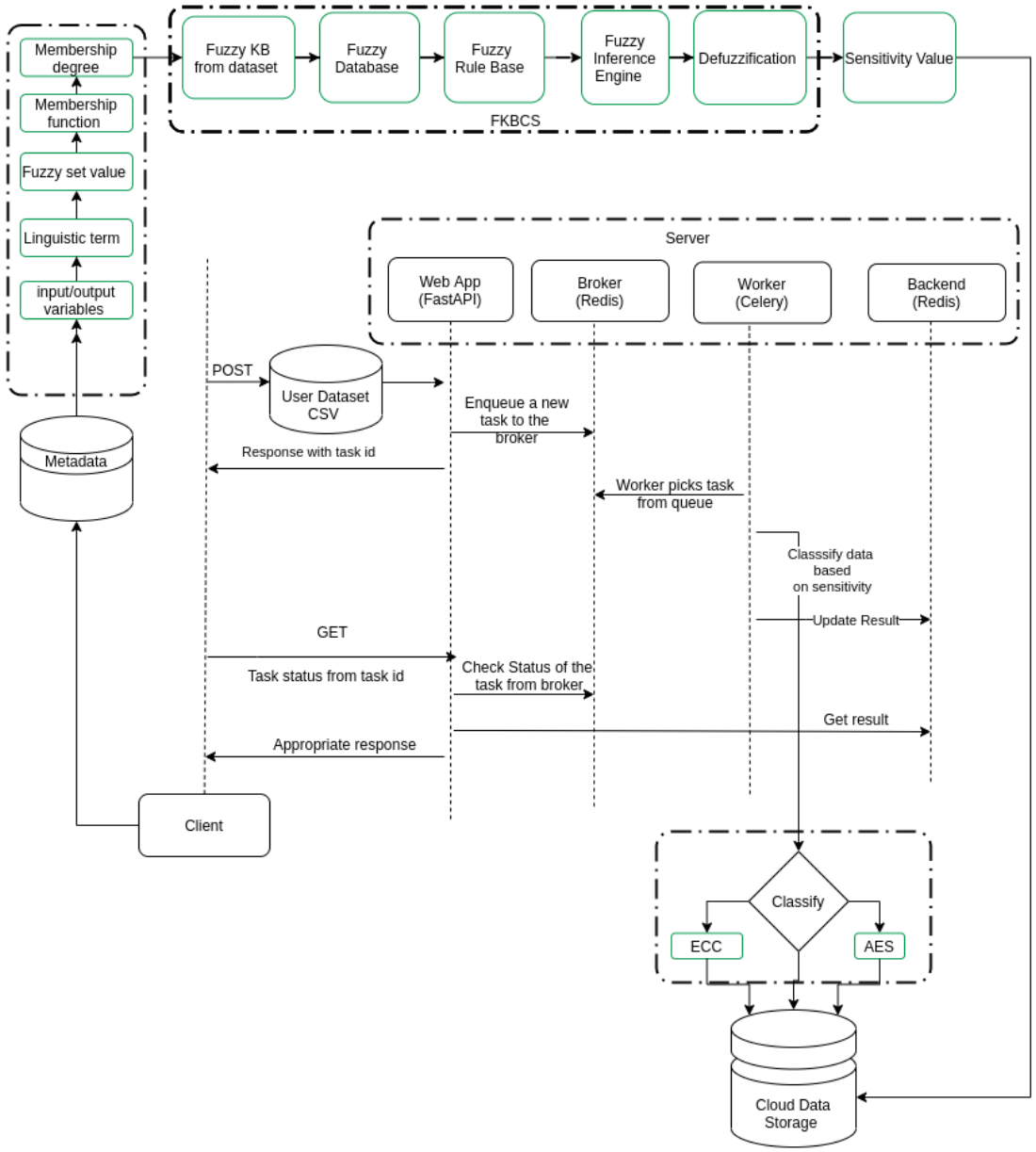

Figure 3.4: Implementation Architecture for Secure Data Classification and Mobility Model

3.5.1 Input for Fuzzy System

In this section we achieve the fuzzy input or fuzzy sets from each input and output linguistic variables. Fuzzy set is defined by the use of fuzzy set values which is defined by the human expert. Triangular and trapezoidal membership function are used representation of fuzzy

set values with respect to the membership value for the input and output linguistic variables and their subsets. Equation 3.1 and 3.2 used for representation of fuzzy values in terms of fuzzy sets. Figure 3.5 and 3.6 express the fuzzy database for input and output of the FKB data classification system. The x axis shows the universe of discourse these value is scaling by the human expert and y-axis shows the $\mu_x(x)$, which is degree of membership function. The range of the degree of membership function always [0, 1] which is expressed by equation 3.5.

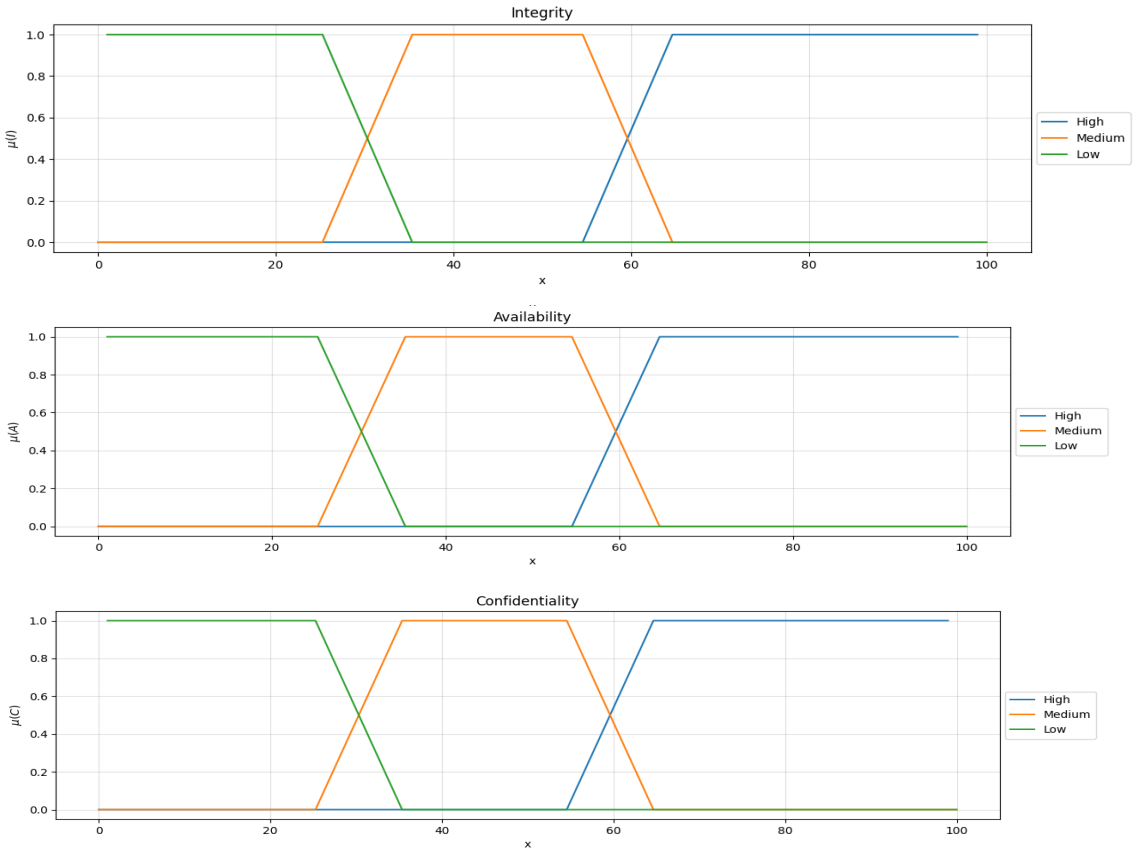

Figure 3.5: Fuzzy Database for Input Linguistic Term

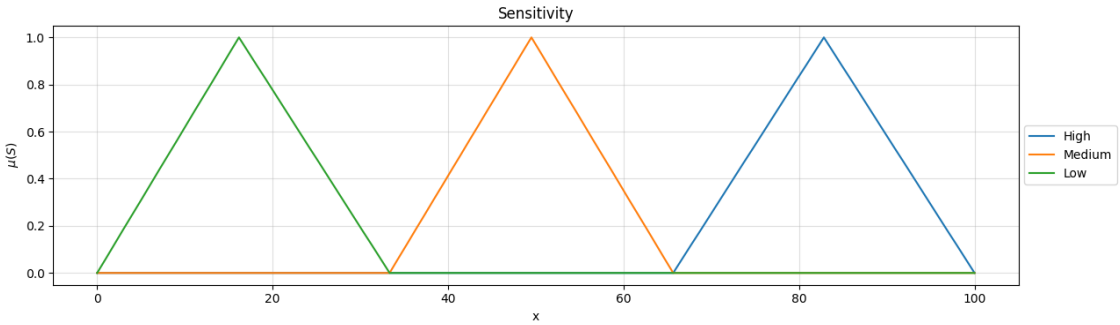

Figure 3.6: Fuzzy Database for output linguistic term

3.5.2 Rule Generated by (FS)

It is used for the purpose of representation and processing of imprecise and uncertain human knowledge. This knowledge is provided by the linguistic variables in terms of degree of membership. The fuzzy rule base executed if the forms of antecedents and consequents clauses that is express in the section 3.6.1. Antecedent and Consequent part of the fuzzy rule are formed by the use of atomic and compound proposition. Finally fuzzy rule has induced by using the conditional proposition.

The representation of the fuzzy rule base has mentioned in the figure (4.2). This represents the below rule in a graphical manner.

$R^{(1)} = \boldsymbol{IF}\ C = Low\ \boldsymbol{AND}\ I = Medium\ \boldsymbol{AND}\ A = high\ \boldsymbol{THEN}\ S = Confidential$

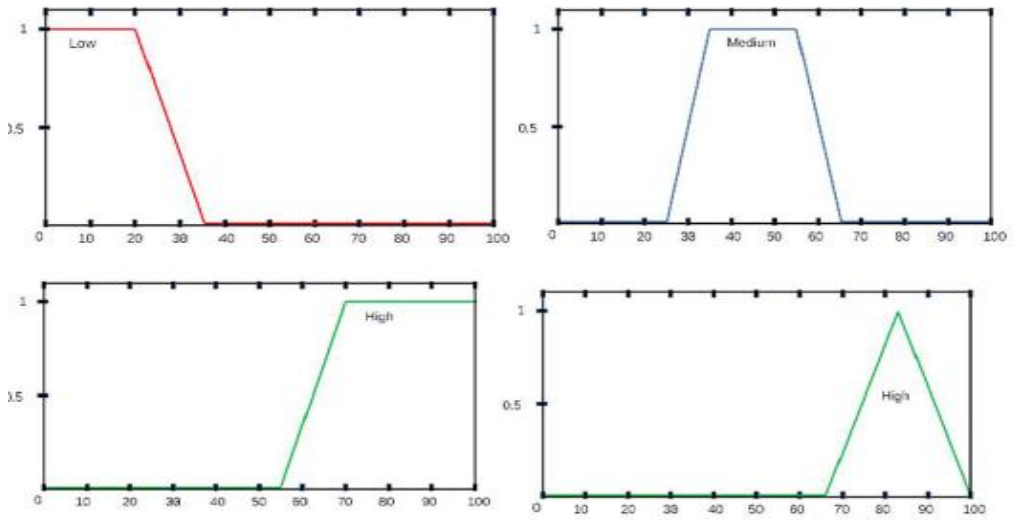

Figure 3.7: *Fuzzy rule base for (if C=L and I=M and A=H then S=confidential)*

Table3.10: Membership Degree for Fuzzy Sets

Input/output	Subsets		
Confidentiality (C)	L	M	H
	0.8	0.4	0.3
Integrity (I)	L	M	H
	0.6	0.7	0.5
Availability (A)	L	M	H
	0	0.8	0.7
Sensitivity (S)	L	M	H
	0.6	0.3	0.9

Using input and output membership function together within a fuzzy region we can calculate the degree of membership function using implication operator which is expressed in equation 3.6 and 3.7. Each attribute of the dataset generates $\mathbf{3^3}$ rules then finally provide the suitable rules by the use of degree of membership function. This calculation is done by the use of decision matrix of fuzzy rule generation which is presented in the table 9. To solve the issues of conflicting rules we assigned a degree of membership values in both antecedents and consequents part of the fuzzy rule base. In our rule antecedent's part includes the input variables and consequent part includes the output class of the system.

Using the fuzzy rule Mamdani inference mechanism plot the minimum degree of membership from all generated rule by the system then calculate the maximum degree of membership from the minimum value of all rules of the antecedents. This minimum value has plotted with the membership function of output. This plotted output is fuzzy output of this system that carry the sensitivity class of our datasets attributes. Using the equation 3.10 this fuzzy output is defuzzify. This value is stored in the cloud database schemas with predicted class or sensitivity level of data.

3.5.3 Implementation Environment

This method is implemented in the real cloud environment which is managed by AWS Cloud Service Provider (CSP). The configuration of the testing environment is mentioned below.

AWS EC2: It is a web services used for providing a secure and resizable computing capacity in the cloud computing environment. It allows for the user to obtain and configure a complete computing resources for variety of purposes. EC2 offers variety of choices storage, networking, processor and operating with fastest processor in a cloud environment. It provides a most powerful GPU instances for different application like Machine learning, graphics workload and windows workload [16]. There are number of instances with variety of configuration that fulfill the every business needs globally.

T2 micro instances: It is a general purpose AWS instance type it's able to achieve baseline level of CPU performance. The baseline performance is governed by CPU credits. It can use variety of applications such as low-latency, small and medium database, and development etc. It provides a full core performance if required. Following features are the t2 micro instances.

High frequency up to (3.3 GHz) Intel Xeon processor, 1 vCPU, 6 CPU Credits / hour, 1GB memory, low to moderate network performance and 30 GB SSD Amazon Elastic Block Storage (EBS) [13].

The performance of the AWS depends on vCPU, size of memory, network performance and stored data size or transferred data. The configuration of the cloud depends on their performance based on the user requirements. All data was managed by using the PostgreSQL database system.

3.5.4 Tools and Resources Used

This thesis work is formulated by the use of following tools and resources these are demonstrated in below:

1. **Python:** Python is a high-level programming language commonly used in machine learning research projects.

2. **Numpy:** Numpy is a library widely used for scientific computing in Python by providing a multidimensional array object, and functions and methods to process them.
3. **Pandas:** Pandas is a Python library for doing fast and quick data manipulation and analysis.
4. **Matplotlib:** Matplotlib is the most common plotting library for the Python language.
5. **PyCharm:** PyCharm is an Integrated Devlopment Environments (IDE) for python developers for web and data science application.
6. **Tyniec:** tyniec is used for manage the ECC and quickly create a key pairs (signing and verifying key).
7. **Asyncpg: asyncpg is library used for** database manage and interface specifically implementation for PostgreSQL server binary protocol in python environment.
8. **Uvicorn**: it is used for ASGI server implementation using uvloop and http tools for the host, gateway interface and open the port expose the API.
9. **multi-part:** it is used for assist to upload a file in a API
10. **Fast API:** it is use for to develop the high performance API with python framework.
11. **Redis:** it is an in-memory data structure used for database cache, background process and broker management.
12. **SQLAlchemy:** it is used for design an efficient and high performing database access in python environment.
13. **Docker:** it is a PaaS service used for containerized, OS-level virtualized and deploy. It support and run the code in any environments using a simple commands through a single API.

3.6 Measurement and Evaluation

We measures the performance evaluation of classification and sensitivity based mobility of data in terms of classification time, response time, delay time and throughput. The major

performance indicator of our model is execution time which is calculated by the use of software execution environment and performance of CPU and used memory [8].

Classification time: it measures the data classification time generated by purpose FKBCS.

Encryption time: it measures the data encryption and decryption time taken by security algorithm which are ECC and AES.

Execution time: it can be evaluated by the software use in execution environment, CPU performance and memory size used in the platform.

Delay time: it depends on the specific parameter of the network platform. Physical distance between the end user and location of the cloud storage (AWS region) as well as internal network capabilities and virtual traffic. These factors cannot be adjusted by the end user but only methods can choose for data transfer and invocation.

Throughput: it is the main cloud performance measurements parameter that can be calculated by measuring the time of given two factors i.e. execution time of all components and delay between them that can be calculated by using the following parameter.

3.7 Validation Method

In the scope of this thesis validation is the process of checking whether or not the system output is accurately classified or not in the appropriate class of the model based on the organizational requirements. The organizations provide their security requirements on the basis of cloud data governance security parameter i.e. Confidentiality, Integrity and Availability. In this process it is assumed that the attributes of sensitive or critical data has highest score of security parameter and less sensitive data has lowest score of the security parameters. For the validation purpose, there were not available any standard dataset that could be used in the domain of sensitive data classification, the expert in the field of IT (Database expert, system administrator, system developer, MIS expert and DG expert) those are working in the different level of government organization are chosen to evaluate the system with certain score. The generated output of the system is described in the previous section.

Validation data

The validation data given to the individual evaluator is as shown below. It included the attributes of dataset and their CIA values and sensitivity level with output class. The result provided by the expert is compare with output generated by the system. The result generated by the system is not given to the evaluator because they can freely evaluate and give their score to results.

The following table shows the information about the number of attributes of dataset and experts used for the validation purpose in this thesis work.

Table 3.11: Dataset used for Validation

No of attributes	**Input values**	**Output values**	**No of Expert**
14	C, I, A	S, C, P	10

The following sub-section provide the discussion on the validation data that has been used in this methodology and various considerations that have been enforced.

Experimental summary

The experiments were performed for 14 attributes of the data set. The table below provides the summary of all the experiments. The standard deviation presented in the table is calculated as

$$\sqrt{\sum_{i=0}^{n} \frac{(x-\bar{x})^2}{n-1}} \qquad (3.15)$$

And the standard error is given by:

Standard Error = Standard Deviation / Square Root of n

The value of n is 13.

CHAPTER 4: RESULTS ANALYSIS AND COMPARISION

Data security is the major concern during data mobility in the cloud environments. The main objectives of this thesis is to develop the data classification model by using the fuzzy rule base system. This research is focused on characterization of data based on the owner security requirements which is defined by using cloud data governance security parameters i.e. CIA triad and classified into three level classification i.e. public, sensitive and confidential. In addition, Fuzzy logic permits to response the unclear circumstances with input and output linguistic variables and their subsets which allows to generate the fuzzy rule base and fuzzy inference mechanism for sensitivity based data classification approach. Therefore, a fuzzy logic based classification model was devised and used previous work [8] for comparison and analysis. For the validation purpose, the expert score in terms of CIA triad was collected. These value is used to predict the output classes of data by using linguistic processing. It also measures the sensitivity based data mobility in distributed cloud environment in terms of securing time, response time and delay time.

In this work, the performance of proposed integrated method for data classification and security management model was tested and validated, thus supporting the secured data mobility in distributed cloud system with maximum throughput and minimum security overhead.

4.1 Experimental Environments

This method is implemented in the real cloud environment which is managed by AWS Cloud Service Provider (CSP). The computing environments describe Amazon T2 micro instances (3.3 GHz) Intel Xeon processor, 1 vCPU, 6 CPU Credits / hour, 1GB memory, low to moderate network performance and 30 GB SSD Amazon Elastic Block Storage (EBS) runs OS Linux 18.04LTS, PyCharm IDE 2020.2 x64.

The performance of the AWS depends on vCPU, size of memory, network performance and stored data size or transferred data. Proposed data classification and security method are tested on various file sizes roughly 0.25, 0.50, 1 and 2 GBs. The configuration of the

cloud depends on their performance based on the user requirements. All data was managed by using the PostgreSQL database system.

4.2 Experimental Results

The overall results of the system are described in this section. The experiment were setup to measure the system classification time, response time, delay time and throughput of the dataset according to the sensitivity level of data.

The experimental results below shows the performance of security algorithms on three classes of data, viz. public, sensitive and confidential, based on securing time and varying file sizes which impacts the mobility of the system in terms of response time, delay time and throughput.

4.2.1 Classification Time

The data classification time generated by purpose fuzzy knowledge base classification system has been measured in this section. Table 12 summarizes the total classification time for the purposed model according to different file size (0.25, 0.5, 1, 2 GB).

Table 4.1: Data classification Time

File size (GB)	**0.25**	**0.50**	**1**	**2**
Classification time (sec)	22.4	46.77	93.54	187.23

The dataset is classified on the basis of sensitivity level meeting the owner requirements while maintaining the confidentiality. The classification time for 0.25 GB data was 22.4 seconds while for 0.5 GB was 46.67 seconds, whereas for 1 GB data was 93.54 seconds and for 2 GB is 187.23 seconds. With the increase in file size, the data classification time also increases linearly, almost doubles with twice the increase in file size.

4.2.2 Performance of Security Algorithm

The performance of security algorithm were measured in terms of processing time and file size before and after encryption. The processing time represents the encryption and decryption time. File size was dependent on the sensitivity of data, as the encryption algorithm were selected on data sensitivity level, whose ramifications were the various file sizes after encryption.

Encryption and Decryption Time

We applied a security algorithm as per requirements of data for the purpose of storage and mobility over multiple cloud databases. RSA and ECC were implemented as an asymmetric algorithm while AES and DES were implemented as symmetric algorithm. We used RSA to generate 1048 bit public and private keys, ECC generates 256-bit, AES generates 64 and DES generates 128 bit private and public keys. The comparative analysis of these security algorithms were done in terms of encryption and decryption time along with file size.
The algorithm is selected on the basis of security strength, i.e. highly sensitive data uses more secured algorithm, thus more security overhead i.e. processing time and file size, was added. As mentioned in figure 2, the algorithms were compared based on the encryption and decryption time on different data sizes.

The AES algorithm performed the best while considering the encryption and decryption time when compared with DES algorithm, thus providing more secured data transmission than DES with minimum security overhead. The data of sensitive class required secured transmission with minimum security overhead. AES has been selected for the processing of sensitive data because of its symmetric nature the file size reduces while providing maximum security among the symmetric algorithms.

The RSA algorithm took the most time during encryption and decryption due to its need to generate a 1024 bit unique private and public key. For the comparison of asymmetric security algorithm, ECC provided more security strength due to the shorter key length, thus

reducing the processing overhead i.e. encryption time and file size. For the confidential data has processed by the use of ECC algorithm. Public class data has been processed without any security implementation that drastically minimize the security overhead than other classes. Thus, the selection of the algorithms in this thesis was focused on reducing the processing and transport latency of the cloud environment.

Table4.2: Comparison of Different Security Algorithm (Encryption and Decryption Time)

S.NO	Algorithm	File size (GB)	Encryption time (Sec)	Decryption time (Sec)
1	AES	0.50	7.65	4.78
	DES		14.35	5.26
	RSA		34.92	23.44
	ECC		28.55	14.32
2	AES	1	15.65	9.68
	DES		29.35	12. 45
	RSA		71.92	23.44
	ECC		63.10	26.33
2	AES	2	32.44	20.13
	DES		62.34	27.50
	RSA		140.54	50.45
	ECC		118.22	41.22

Original and Encrypted File Size

After encryption, the file size has increased due to the extra security overhead. The Table (14) provides the comparison of security algorithm based on the file size before and after encryption in the cloud computing environments. In this thesis, ECC and AES security algorithm were used for the purpose automatic security management based on the sensitivity level of data. The file size of 0.5 GB after encryption using AES algorithm was 0.87 GB, while for ECC algorithm the file size was 1.75 GB, almost double the file size when compared with former. Whereas, for 1 GB file size, the size after encryption for AES and ECC were 1.74 GB and 3.5 GB respectively. Similarly, for 2 GB file, the file size were 3.84 GB and 7 GB for AES and ECC respectively. Thus, when considering the file size, the performance of the AES algorithm was significantly better.

However, the ECC algorithm is considered better due to its asymmetric nature even though the file size after encryption is significantly large. Although the file size was humungous, the ECC algorithm was still preferred because of its complex encryption and decryption

that forbids unauthorized users which significantly increases the data security. So, ECC algorithm was suitable for confidential data and was selected in this thesis.

Table 4.3: Comparison of Security Algorithm Original File size Versus Encrypted file Size

Security Algorithm	Before Encryption File size in (GB)	After Encryption File Size in (GB)
AES	0.5	0.87
	1	1.74
	2	3.84
ECC	0.5	1.75
	1	3.5
	2	7

4.2.3 Mobility Measurements

The implementation of data centric security mechanism increase the security overhead. These security overhead affect the performance of cloud data mobility. We measured the securing time and transmitting time for the analysis purpose. The following section measure the details of the extra processing and transmitting overhead in terms of response time, delay time and throughput of the system.

Data Transmission Response Time

The data is transmitted when the user authentication is established between cloud systems. Authentication between two clouds has been done by integrating a public key with the use of private key. The secured data mobility in a cloud systems causes operational and processing security overhead. Therefore, sensitivity based security management are needed to achieve secure data mobility on the cloud. The cloud system does not provide the role base transmission of sensitive data. The security mechanism were started from source cloud system to destination cloud based on owner requirements by the use of fuzzy rule base classification system. The table 15 describes the desirable transmitting response time i.e. data securing and transmitting time for four partition of the dataset. The proposed secured and cloud baseline non-secured method transmits data files at 64Mb/sec transmitting rate. The user authentication process was established by using a common key that was provided

by the sender and receiver cloud. Secure data mobility in a cloud system decreases the response time, delay time and throughput due to the security overhead.

Table 4.4: Data Transmission Response Time

Method	**File size in (GB)**			
	0.25	0.50	1	2
Non -secured	31.1	63.9	114.8	285.77
Proposed Secured Method	35.37	74.45	156.85	333.12

The user authentication process was established by using a common key that was provided by the sender and receiver cloud. Secure data mobility in a cloud system decreases the response time, delay time and throughput due to the security overhead. The proposed model enhances the secured data mobility in cloud computing to transmitting data closer to performance of cloud base line or non-secured approach. The results of the both secured and non-secured method are presented in the figure 4.1 which denotes the total transmitting response time realized by our model.

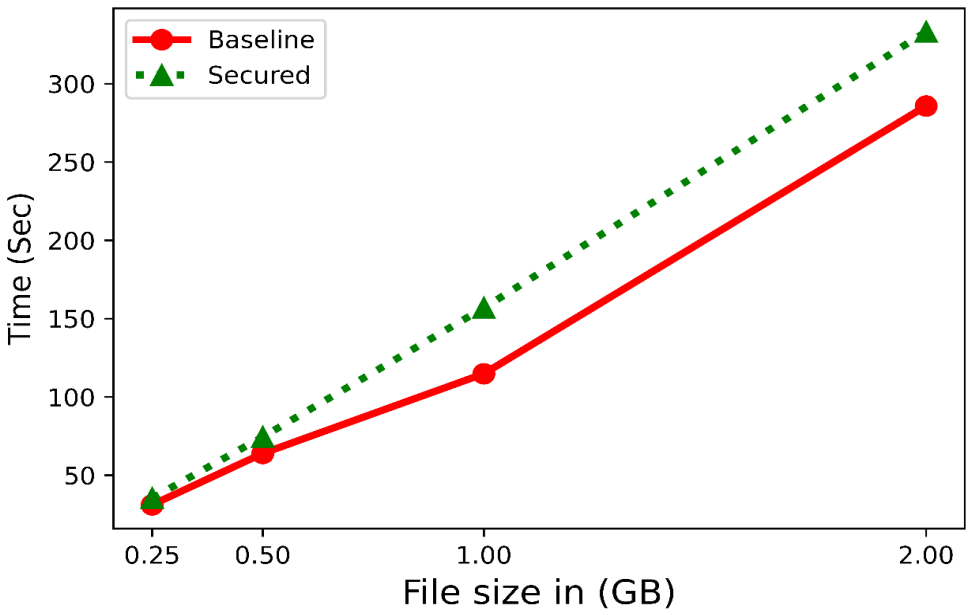

Figure 4.1 Data Transmission Response Time

Data Transmission Delay Time (DTDT)

The time taken to transmit a packet from the host to the transmission medium is called transmission delay. The bandwidth is directly proportional to the transmission delay that means bandwidth is increased, the transmission delay also increases. It is measured by evaluating the ratio of data transmission file in bit and available bandwidth of the system. Secure data mobility techniques causes extra processing overhead due to the extra transmission of data with encrypted version. This file size in increased after the implantation of security algorithm. This increased file size also increases the transmission delay time from source to destination cloud system. The secured mobility of data increases both response time and delay time.

Figure 4.2 shows the data transmission delay time in our proposed model and that is compared with the delay time of non-secured cloud baseline. However, this model increases the delay time and degrade the throughput of cloud system.

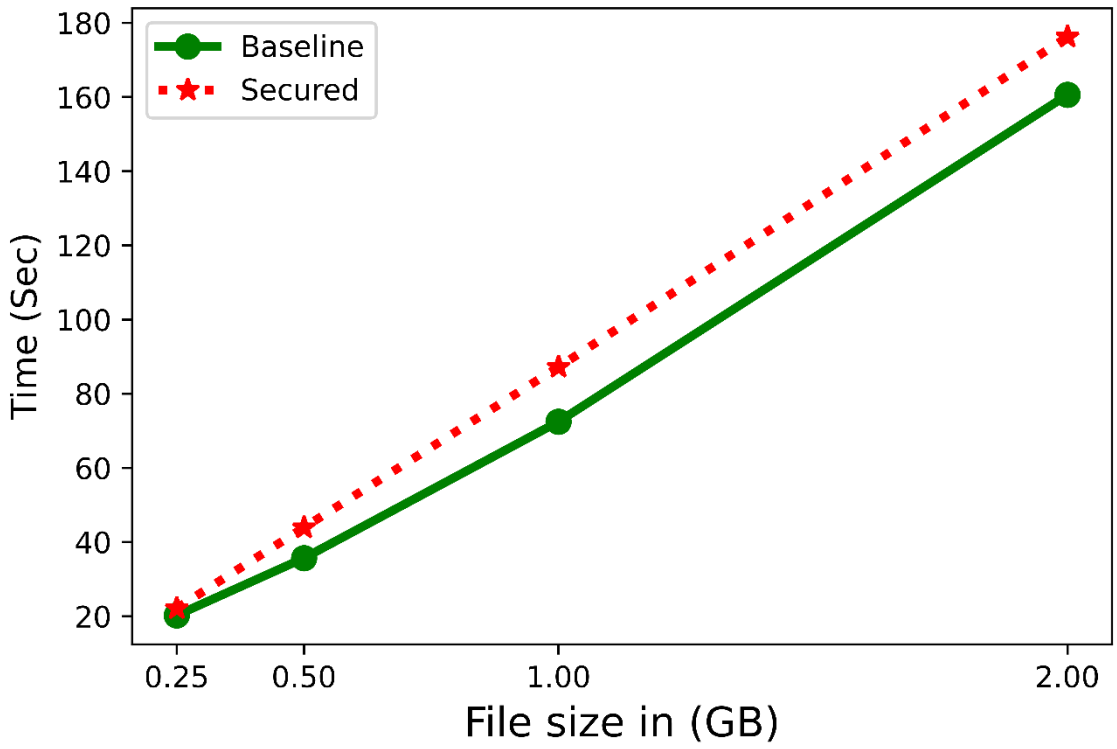

Figure 4.2 Data Transmission Delay Time

Data Transmission Throughput (DTT)

For the study of security implantation in cloud computing distress the data transmission throughput. The figure 4.6 represents the data transmission throughput is decreases when file size is increases due to the extra processing of security overhead but it provides the more secure and agile inter cloud data mobility. The throughput of the cloud system is decreased but data security was achieved during the mobility of data. This analysis assured secured method for data mobility cause shortage of throughput. Moreover, figure 4.6 suggests that the throughput is generally decreased when file size was expanded. This method enhances the securing and transmitting operations between senders and receiver cloud database. It automatically allows appropriate security management based on the criticality of data. Critical data were stored in the encrypted format key only known to the data owner. This method enhances the secure data mobility efficiency and decrease the both response time and delay time

Table 4.5: Data Transmission Throughput

Method	0.25	0.50	1	2
Secured [2]	12.061	10.81	10.22	9.141
Non secure	16.37	16.025	14.91	14.33
Secured	14.25	13.75	13	12.24

Average throughput

The average throughput of the cloud data mobility was measured with multiple observation in different dataset. The thirty number of observation were done and measured throughput of cloud data mobility. We calculated the mean and medium of the thirty observation for measuring the consistency and reliability of the proposed system. The figure 4.3 shows the average network throughput achieved by 30 number of observation. This provided the precise view of performance of cloud data mobility. This figure illustrates the overall change in mean and median of throughput as the number of observation increases up to 30. As the number of observation increases, the value of mean start to saturate. With the increasing number of observation particularly around N=15, the median value of throughput is higher than the mean of the total observation. This signifies that as the

number of observation increase, the majority of the throughput value is higher than the average of the observation. The distribution of result is positively skewed.

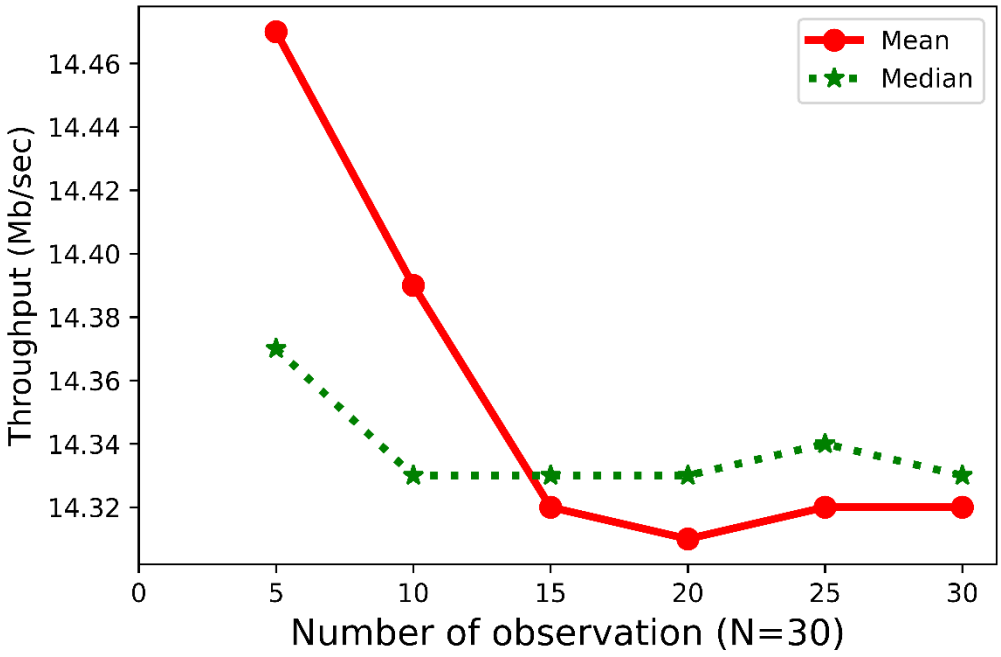

Figure 4.3: Average Throughput for Number of Observation

4.3 Validation Results

The validation experiment were performed in 14 attributes of the dataset. The experiment results measured and calculated the standard error by the use of equation 3.15. The evaluation of the system output is carried out by the IT professionals who have at least 7 years of experience. They all have managerial positions on their respective working area with different terms of reference (ToR). So, it can be said that the validation presented in this thesis work closely relates to the security requirements of attributes of any organizational dataset. The total number of the experts used in this is 10. The standard error for each attributes of the dataset are presented in the figure 4.4. The Standard deviation evaluates the amount of variability of each data set based on system results and expert score of individual attributes of the dataset. The standard error for the mean (SME) evaluates how far the sample mean of the data deviates from the population mean.

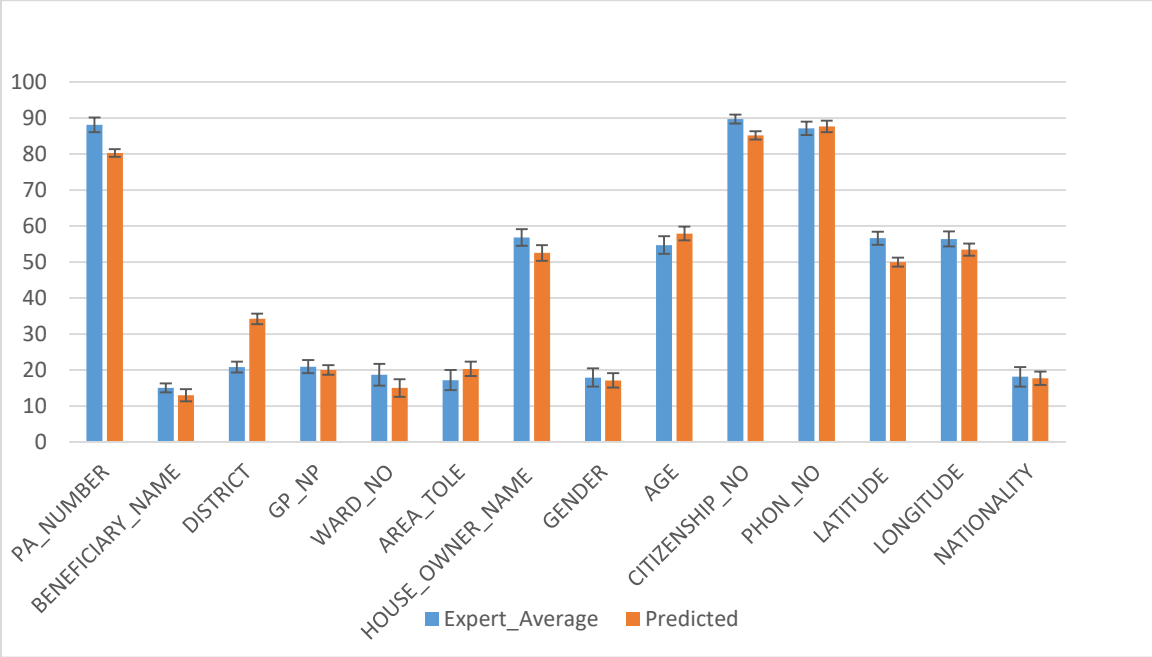

Figure 4.4: The Graph for the Standard Error of Each Attributes of Dataset (Expert Average Vs. predicted)

In this study, the standard error for the mean (SME) describes how precise the mean of the system score and expert score is estimated to the true mean of the output provided by the system and expert score for each attributes of the system. In our calculations, if the standard error is small then it represents the true mean. In the case, the standard error is high which shows the performance of system have remarkable irregularities of the overall system results and expert score. This below table 17 provides the average error for all the attributes of the dataset, i.e. 14 attributes, which are used for the evaluation of system results on the basis of expert score. From the observation of this table, most important point to be noted was the even distribution of the error across all the attributes of the dataset. This suggests that the performance of the system is consistent even though the average error percent is 19.12%. The scatter plot shows the correlative relationship among the system score versus expert score. The trend line is used to find the statically best fit to the system score versus expert score. This also helps to identify the unusual points of data that affected the calculations of the trade line. The regression line present in the graph can be used to predict the values of a dependent variable based upon the values of an independent variable.

Table 4.6: Overall Error of Each Attributes of the Dataset

Attributes	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**Average**
Error (%)	18.94	19.3	12.71	16.05	12.8	16.65	22.76	38.17	11.68	18.94	27.92	17.96	14.58	18.42	19.12

From the graph, it is obvious that the plot for all attributes description is evenly distributed. This graph shows the scatter plot for all the attributes description with system score on y axis and expert score on x axis. This figure 4.5 shows the data are grouped very closely with each other.

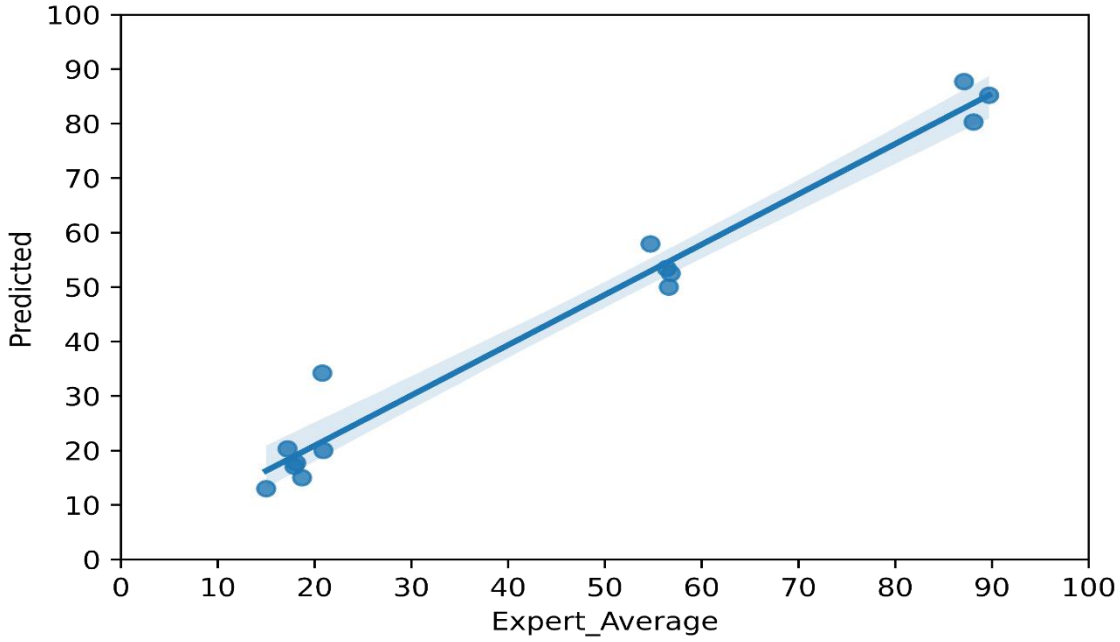

Figure 4.5: The Scatter Graph of Expert Score verses System score for all Attributes

4.4 Preformation Evaluation

This method focus on the data security management before executing the data in a cloud computing environments. The performance of the cloud system was measured using securing time, transmitting time, delay time and throughput, when secured method for storing, treatment and mobility of data was applied in the distributed environment. The secured method enforced additional security overhead which decreases the performance of cloud system.

The figure 4.6 represents the cloud system performance secured and non-secured method. The method presented in [8] had security overhead while data transmission but this proposed method had less security overhead when compared to the previous work. As AES and DES were implemented in this work which reduced the file size after encryption, thus reducing the overall security overhead of the system. Our secured data mobility model was able to manage automatic encryption based on the sensitivity level of data, which ensured the data centric security solution for any organizational data. This model stored highly critical data with encrypted format and the key is known only by the data owner. The integrity check would be adequate for non-critical data mobility in a public cloud. This requirements are able to enhance the mobility of data with more efficiently and minimize the total response time and delay time.

The throughput of the secured and non-secured method are described in a figure 4.6. This figure compares the throughput achieved in three different scenarios the red line represents the throughput achieved in the existing paper, the dotted green line represents the throughput achieved by our secured proposed model while the orange line represents the throughput achieved in cloud baseline. The results of proposed model in this thesis is compared with [8] and cloud baseline or non-secured method and our proposed model performed slightly better than model at [8]. Also, it improved the overall throughput achieved when compared to the existing work because of the implementation of better encryption algorithm thus achieving higher throughput. The encryption of data before being transmitted in distributed environment ensures the confidentiality of the data.

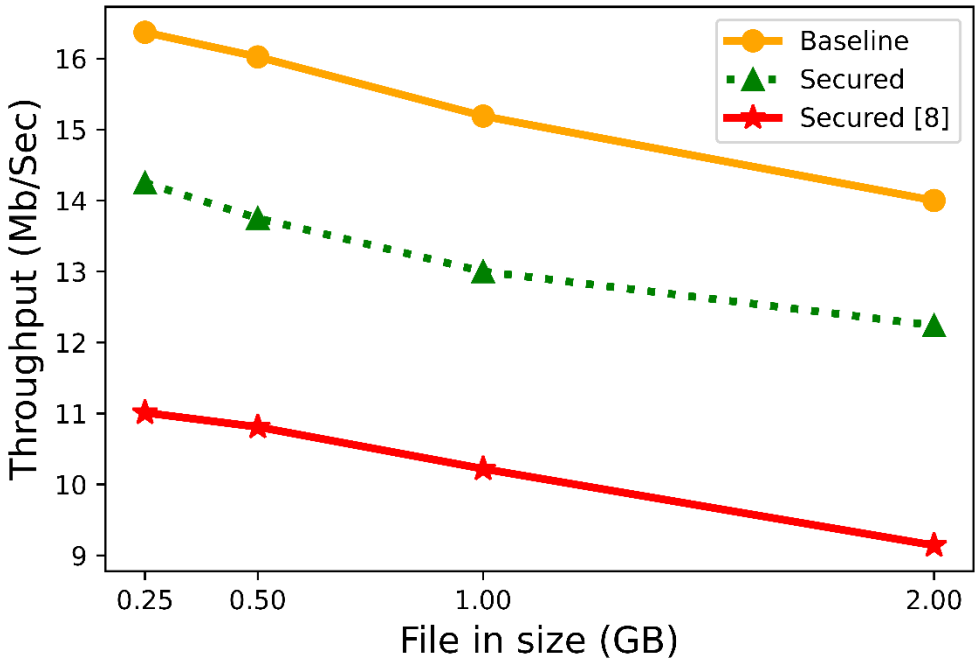

Figure 4.6: DTT (Comparison Secured Vs. Non-Secured)

The data centric security solutions addressed in our proposed method takes consideration on efficiency and security of data mobility in the cloud data storage. Thus, the protection of the data against security threat during data mobility ensures mitigation against the threat of loss of data. Therefore, the encryption of the data with the help of metadata helps protection against advance persistent threat (APTs) during the mobility.

The classification and securing time represents the security overhead that consumes the redundant network bandwidth and processing overhead. The model proposed in this thesis helps to minimize the sensitivity based processing overhead by reducing the total data transmission response time, delay time and data transmission throughput.

CHAPTER 5: CONCLUSION AND FUTURE RESEARCH

In this thesis work, we investigated and implemented the FKBCS for security mamangemet on government dataset according to sensitivity level of data. The input and output variables of this model describes the security parameter and sensitivity level of data.

5.1 Conclusion

We design and implement the two fold mechanism for data classification and security management for secured cloud data mobility that offers the data loss prevention (DLP). FRBCS is used to predict the classification label by processing the cloud data governance security parameter CIA as an input features of the system. These parameter are used to define the security requirements of attributes of dataset that ranges between 0-100. The security requirements of the dataset is provided by the data owner based on the sensitivity level of data. This proposed model identifies the security needs before transmitting the data, thus helping to reduce the under and over security overhead by considering the owner security requirements.

The efficiency achieved by this proposed model were demonstrated using the tabular and graphical representation method. The FRBCS delivers the maximum performance improvement by avoiding the redundant encryption process for public dataset. Data mobility was evaluate in terms of data transmission response time, throughput and delay. The experimental analysis was done in real cloud environments and the high performance of our model was achieved when compared to existing work.

5.2 Future Research

This research provides the significant opportunity for future enhancement in the field of secure cloud data governance. In this proposed model, the scope is limited to the design and implementation of FRBS for automatic data classification and security management based on the basis of sensitivity level of data by linguistic processing. In future research, we will develop the adaptive method for fuzzy rule generation by implementing the learning algorithm that needs more dataset. The input parameters can be added to develop more secure and efficient method for data classification and mobility in the cloud environment. The further research can be conducted by the implementation of big data classification and security in cloud computing environments using fuzzy logic classification system.

REFERENCES

[1] M. T. I. T. S. &. Security", "Data Classification and handling Policy".

[2] J. S. P. Manish Pokharel, "Issues of Interoperability in E-Govemance System and its impact in the Developing Countries: A Nepalese Case Study," *IEEE,* 2009.

[3] H. B. A. Bibi Zarine, "Making an Interoperability approach between ERP and Big Data context," *IEEE,* 2018.

[4] K. Priyank Singh Hada Central University of Rajasthan, "Security Engineering in G-Cloud: A Trend towards Secure e-Governance," *International Journal of Computer Applications (0975 – 8887),* 2012.

[5] D. T. Pradeep Tomar, "Integration of Cloud Computing and Big Data Technology for Smart Generation," 2018.

[6] K. Kaur, "A Secure Data Classification Model in Cloud Computing Using Machine learning algorithm," *Researchgate,* 2016.

[7] D. T. Larose, "Discovering Knowledge in Data: An Introduction to Data mining," 2005.

[8] A. G. S. N. a. I. K. I. Hababeh, "An Integrated Methodology for Big Data Classification and Security for Improving Cloud Systems Data Mobility," 2019.

[9] E. B. &. K. Hameed, "A systematic literature review of data governance and cloud data governace," 2018.

[10] N. A. Kamariah Abu Saed1, "Data Governance Cloud Security Checklist at Infrastructure as a Service (IaaS)," *(IJACSA) International Journal of Advanced Computer Science and Applications,* 2018.

[11] S. S. M. A. MUJEEB1, "Data Governance in the Cloud," *International Jurnal of scientific engineering and technology rearch ,* 2017.

[12] D. haynes, Metadata for information management and retrival, facet publishing , 2018.

[13] M. M. B. S. K. V. Spoorthy1, "A Survey on Data Storage and security in cloud computing," *International Journal of Computer Science and Mobile Computing,* 2014.

[14] A. G. S. N. I. K. Ismail Hababeh, "An Integrated Methodology for Big Data Classification and Security for Improving Cloud Systems Data Mobility," *IEEE,* 2018.

[15] M. A. T. J. L. M. Y. Z. Zardari, "Data Classification Based on Confidentiality in Virtual Cloud Environment," 2014.

[16] D. S. S. kiran, "ENHANCE DATA SECURITY IN CLOUD COMPUTING USING MACHINE LEARNING AND HYBRID CRYPTOGRAPHY TECHNIQUES," *International Journal of Advanced Research in Computer Science.*

[17] D. H. M. Mohammadian, "A hierarchical fuzzy logic systems frame work for data security," 2017.

[18] A. S. Ehsan Pourjavad, "The Application of Mamdani Fuzzy Inference System in Evaluating Green Supply Chain Management Performance," *springer,* 2017.

[19] N. A. A. W. R. N. H. z. H. Kamariah Abu Saed, "Data Governance Cloud Security Assessment at Data Center," *IEEE,* 2018.

[20] M. A.-R. a. E. Benkhelifa, "A Conceptual Framework for Cloud Data Governance-Driven Decision Making," *IEEE,* 2017.

[21] D. H. Masoud Mohammadian, "Data classification process for security and privacy based on a fuzzy logic classifier," *Int. J. Electronic Finance,* 2009.

[22] T. N. T. M. Hisao Ishibuchi, "Performance Evaluation of Fuzzy Classifier Systems for Multidimensional Pattern Classification Problems," *IEEE,* 1999.

[23] C. C. A. Talon, "Selection of appropriate defuzzification methods: application to the assessment of dam performance," *HAL,* 2017.

[24] H. Budapest, "Improvement possibilities of the maximum defuzzification methods," *IEEE,* 2019.

[25] D. P. M. &. A. Sachdeva, "A Study of Encryption Algorithms AES, DES and RSA for," 2013.

[26] M. A. R. Mrs. TamilSelvi. S., "Information Confidentiality Using Fuzzy Based Data Transformation Method," *International Journal of Computer Techniques,* 2017.

[27] EU, "Risk Management and Assurance Models," *ISO27005:2011 risk assessment metrics,* 2011.

[28] I. D. N. A. B. Mykhailo Klymash, "The “Data Embassies” Concept as a Secure Communication Core for e-Gov mplementing in Emerging States," *IEEE,* 2019.

[29] M. V. M. A. T. K. a. F. F. Moghaddam, "A reliable data protection model based on re-encryption concepts in cloud environments," *IEEE,* 2016.

[30] A. O. A. A. R. T. M. A. J. B. Awotunde, "Evaluation of Four Encryption Algorithms for Viability, Reliability and Performance Estimation," *NIGERIAN JOURNAL OF TECHNOLOGICAL DEVELOPMENT,* 2016.

[31] X. A. Mingxiang He, "Information Security Risk Assessment Based on Analytic Hierarchy Process," *Indonesian Journal of Electrical Engineering and Computer Science,* 2016.

[32] A. Z. A. H. Oussama Arki, "A Cloud Data Classification Model Using Fuzzy logic," *IEEE,* 2020.

[33] E. B. K. H. Majid Al-Ruithe*, "A Conceptual Framework for Designing Data Governance," *science direct,* 2016.