**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

**Student Name:** MAMIDI SNEHITH REDDY

**Student ID:** 22195297

**Programme:** MSc CYBER SECURITY **Year:** 2023-2024

**Module:** MSc Research Practicum Part 2

**Lecturer:** Jawad Salahuddin

**Submission Due Date:** 12 August 2024

**Project Title:** Ensuring Privacy in the Digital Era: Innovative Strategies for Data Encryption (ISDE) with Access Control

**Word Count:** 5789 **Page Count:** 6

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** Mamidi Snehith Reddy

**Date:** 12/8/2024

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |

| | |
|---|---|
| Date: | |
| Penalty Applied (if applicable): | |

# Ensuring Privacy in the Digital Era: Innovative Strategies for Data Encryption (ISDE) with Access Control

Mamidi Snehith Reddy

22195297

**Abstract:**

Cloud storage is very vital today especially for the users; thus, there is need for proper protection mechanisms. Such methods while providing security often have had high computational and space complexity compared with other encryption methods. To cope up with these issues, this system has developed Innovative Strategies for Data Encryption (ISDE) which combines Attribute-Based Encryption (ABE) besides sophisticated Access Control features. The proposed ISDE framework also incorporates a Sensitive Keyword Weight age Scoring System and NLP to chose only sensitive phrases thus spends less CPU cycles and does not over-encrypt. Moreover, this system also improves the monitoring of the organization by identifying the employees who shared such files at the workplace and sends notifications to the administrators when such activities are encountered. The outcome of this research is a more consolidated solution that incorporates ABE with precise encryption strategies in order to enhance the strength of data security as well as control over the Cloud storage system's access while reinforcing the general privacy inadequacies and resource usage.

# 1 Introduction

In the modern world of computing where data plays a crucial role, cloud storage can be defined as an element that is considered essential in the storage of data because of its flexibility and readily availability. But, with more and more cloud storage systems gaining popularity the issue of security of data stored in such systems is another big concern. Cloud-stored data suffer from the privacy issue of confidentiality, integrity, and availability because of the presence of weak points in data encryption, or the probable intrusion of the data by the wrong hands (Wang et al., 2019). Standard encryption techniques though useful are computationally and storage intensive thus not very useful for large-scale uses (Aslan et al. , 2021). The mentioned scenario should signify the necessity of the development of new approaches and methods of encryption that will allow offering an effective level of protection and reasonable consumption of resources.

The primary objective is: What methods and strategies can be used to protect data that is maintained in the cloud to ensure it does not fall into wrong hands or is utilized in illegitimate activities without much strain to CPU and storage spaces? To address this problem, the research proposes and assesses ISDE as a combination of ABE and NLP. In particular, the aims are the following: to develop an encryption scheme that is concentrated on the specific content with Sensitive Keyword Weight age Scoring System and to incorporate ABE access

control for increasing the protection. This attempt to achieve both simultaneously with regards to privacy of data in the cloud and efficient use of resources.

The hypothesis in this research is that it shall be possible to reduce the overhead costs of computing and storage occasioned by conventional encryption methods by applying NLP for accurate file categorization and ABE for dynamic access control while at the same time fortifying the protection of sensitive data. This work adds value to the existing body of knowledge by introducing the solution based on the synthesis of the novel advanced encryption methods with Natural Language Processing (NLP) tools in solving the issues of data security in cloud infrastructure.

- How to secure sensitive data in cloud environments to prevent unauthorized access and malicious activities without imposing high computational overhead and storage costs?

The structure of this report is organized as follows: The first section is a literature review of the current research that exists on encryption methods and their shortcomings that are linked to cloud computing. The second of the sections covers the strategies used to introduce ISDE in cultures involving ABE and natural language processing. The third segment contains the experimentations and discussions and finally the last segment identifies the outcomes of the study and their implication on cloud security. Last but not the least; the report ends with the discussion of the research contributions and the recommendation for possible research directions.

# 2 Related Work

In the current world, people have embraced use of technology in conducting their businesses and as such protecting sensitive data was seen as very important. This kind of research is geared towards enhancing privacy preserving strategies in machine learning and cloud computing. Thus, the study proposes to cover emerging approaches that include federated learning and homomorphic encryption to solve security problems of healthcare data and clouds. The objective is to have accurate frameworks that protect data privacy while giving high accuracy in machine learning and contribute to the field of secure data management.

**Cloud Security and Access Control**

This category consists of methods and models aimed at improving the level of security and organizing access within cloud settings. Wang, Shangping, Xuet al (2019) put forward a block chain-based access control framework, where Ethereum’s smart contract technology is used for decentralizing the key distribution and access control to cope up with security problems of centralized mechanism. I, Unal, D, Nonnop, & Zhang, 2021 propose the Secure Cloud Storage System (SCSS) that has relatively efficient and scalable characteristics by using IBC with Type-3 pairings to overcome typical PKI issues. S. Yu, W. Zhou et al (2016) came up with a Marking on Demand (MOD) to perform DDoS attack source trace back and is feasible with minimum storage and computation requirements. In the same note, C. Yang, L. Tan et al (2020) also develop an access control framework known as Auth Privacy Chain,

which depends on block chain to make sure that the cloud computing structure maintains confidentiality, integrity, and efficiency in terms of privacy.

**Malware Detection and Prevention in Cloud Computing**

This category of research aims at finding solutions to issues that relate to malware threats in cloud environments. Kimmel, J. C., et al (2021) on to study the application of Recurrent Neural Networks (RNNs) in detecting malware affecting cloud Virtual Machines and gets good detection results with LSTM models. Y. Li, M. Abdelsalam, M. Gupta and others (2021) compare a range of ML techniques used for detecting malware online, and that with DenseNet-121 (CNN) outperforms the others. In line with the intelligent system positions of Aslan, Ömer et al (2021), an implemented intelligent malware detection system for cloud environments obtained 99.8% detection rate. Shivasree, S. M. R. J. Y., et al (2022) has proposed an anomaly-aware behavior-based malicious software detection in cloud environments, which achieves high accuracy through the use of ML models. Sahay, S. K., Sharma, A. and Rathore, H. et al. (2020) review advanced malware detection techniques, stressing the necessity to transform the defenses against confirmed multifaceted threats.

**Data Protection and Privacy in Cloud and Distributed Systems**

This category looks at practices that help in the protection of data and information in distributed systems and popular cloud technologies. Khadse, N. S et al (2021) put forward a system for controlling the outsourced data and for erasing them using Revocable Identity-Based Encryption (IBE) and creating a self-destructing data framework. Yadegari, M et al (2023) explain what Fully Homomorphic Encryption (FHE) is and review a benchmarking suite that Gouert, C et al (2024) put forward to enhance practical FHE libraries. In El Mestari, S. Z. , Lenzini, G. and Demirci, H. , et al (2024), they focus on the risks of privacy and measures for addressing them in the ML systems while having regards with the legal rules and PETs. In Li, J. , Meng et al (2022), the authors propose ADDETECTOR, a private Intelligent Diagonosis and Treatment system for detecting Alzheimer's disease based on federated learning and differential privacy. Khan T et al (2024) discuss Privacy-Preserving Machine Learning (PPML) approaches with the present and future frameworks' analysis and the comparison of existing gaps between concepts and their applicability. Nagy, B et al (2023) describe the enhanced FL methodology for the edge that involves bitwise quantization and local differential privacy. In a 2024 paper, Mousa, N. and Shirazi, F. , et al describe quantum computing technology and address the need for privacy policies when it comes to the quantum world.

**Privacy Preservation and Natural Language Processing (NLP)**

This category is dedicated to the protection of privacy in the field of natural language processing (NLP). Punreddy, A. R., et al (2023) focus on Federated Learning (FL) for privacy-protected NLP, concerns such as communication cost and security. In Lin, B. Y et al (2023), FedNLP is presented, a framework aimed at assessing the FL methods in NLP tasks; this provides challenges and directions for betterment. F. Martinelli et al (2021) suggest an approach comprising of NLP and unsupervised machine learning for creation of labeled

datasets for sensitive information identification in the healthcare and justice fields. D. Mahendran, C. Luo, and B. T. Mcinnes et al (2021) classify privacy preservation approaches in NLP and present prospects for future research, like transfer learning and location privacy.

**CPS Emerging Technologies and Trends in Cyber-Physical Systems**

This category concerns ideas and developments of CPS and associated trends. The technologies of CPSs are elaborated in S. Rho, A. Vasilakos, et al (2016) while covering concepts such as the secure clustering and the CPS design flexibilities while stressing on the cyber and the physical parts. Khan, M et al (2023) review The Framework of Federated Learning (FL) models concerning the Industry 4. 0 and this is about a capability that they hold in solving several industrial problems and at the same time enhancing privacy of data.

## 2.2 Gap Analysis

The analyze of the presently used encryptions shows that there is a lack of the approaches to minimize computational and storage expenses for cloud platforms. Conventional methods of encrypting data have made it easy to secure data but beencloud by high computations and increased storage space. However, while the symmetric encryption methods are efficient, they have problems with key management while the asymmetric methods such as RSA increase computational cost significantly and are also associated with large sizes of ciphertexts. These issues are most acute in cloud environments because the resources shared and scalability is a priority. New technologies like attribute-based encryption (ABE) and homomorphic encryption indicate possible improvements because they increase security and decrease overhead. Nevertheless, these approaches still remain inadequate in terms of practical implementation and the effectiveness of the obtained results. Presently, there are few theoretical and simplistic encryption schemes whose computational and memory overheads are manageable and at the same time offer high level of security. A remedy to these gaps will form the basis for designing the phenomenal techniques that will offer better Cloud data management.

# 3 Research Methodology

It specifies how the research resolves to construct and assess the ISDE framework. It also explains step by step process outlining literature search, framework design, application, and back-ground of scenarios. The chapter explains the instruments applied, and the sources of available data, with detailed procedure of data gathering, and assessment criteria, to ensure a critical evaluation of the framework regarding its viability, efficacy, and protection.

## 3.1 Research Procedure

The research process entails a planned way of developing, applying, and assessing the proposed ISDE framework. This section describes the overall research approach, methods used for data collection and analysis, that is, the scientific approach to the research process.

### 3.1.1 Research Design

The research was conducted in several stages:

**1. Literature Review:** All the available literature relating to ABE, access control models, and NLP were considered for the study. This was useful in this study to pinpoint any gaps and lay the theoretical background for the proposed ISDE framework.

**2. Framework Development:** Given from the literature review, ISDE was constructed to link ABE with enhanced access control features. To selectively encrypt sensitive data a Sensitive Keyword Weight age Scoring System was created which was based on some Natural Language Processing techniques.

**3. Implementation:** The given framework was not specified with language and database technology and was coded in Python using the Flask framework with MySQL as the database. The implementation comprised of creating subroutines to encrypt the data, control the access and identify keywords that require higher security.

**4. Scenario Setup:** Base lining of different actors, services, and situations was carried out in order to come up with some case studies that would depict various levels of data management and security issues in cloud environments. This ranged from low risk data and normal user accounts, privileged user accounts, and high risk data.

### 3.1.2 Equipment Used

- **Software:** Python, Flask, MySQL, TensorFlow, and NLP libraries such as NLTK.
- **Hardware:**Intel Core i5/i7+, 1TB+ hard disk, 8GB/12GB+ RAM, high-performance computing resources. A server with adequate computational resources for running encryption algorithms and NLP models.
- **Dataset:** Publicly available datasets containing sensitive information were used to test and validate the framework.

### 3.1.3 Data Collection

The data collection process involved the following steps:

**1. Dataset Selection:** Portion of this choice was that datasets including both sensitive and insensitive information were chosen. They were collected from public domains to provide differences in the kind of data and the context in which they were taken.

**2. Data Preparation:** These data sets were preprocessed in order to eliminate redundancy and record the data in the desirable form for passage to encryption and analysis. This involved data cleaning such as elimination of irrelevant information or /special characters and normalization of the data followed by tokenization using NLP tools.

### 3.1.4 Data Analysis

Descriptive statistics were utilized on the primary data that was collected from the scenarios in addition to artificial intelligence. The analysis process included the following steps:The analysis process included the following steps:

**1. Sensitive Keyword Identification:** Thus, the sensitive keywords in the datasets were determined through the use of NLP models. These keywords were given scores using the Sensitive Keyword Weight age Scoring System depending to the sensitivity of the keywords.

**2. Attribute-Based Encryption:** Regarding the protection of the identified sensitive data, they were encrypted by using ABE. It was made sure that the presented encryption process was to be following the sensitivity levels, the least significant data should not be encrypted, only the most sensitive data.

**3. Access Control Evaluation:** The performance of the access control mechanisms was assessed on the outcome of the tests for access that are carried out on encrypted data. This involved interaction with the application from the various user roles and access levels to establish the security measures that are in place.

### 3.1.5 Statistical Techniques

Several statistical techniques were employed to analyze the data and evaluate the performance of the ISDE framework:

- **Descriptive Statistics:** Applied for general description of the datasets to give an idea of the scope of the data.
- **Inferential Statistics:** Engaged to make conclusions and determine the relevance of the results obtained.
- **Machine Learning Metrics:** Cross-validation metrics such as Precision, Recall, F1 Score, and Accuracy were employed to measure the NLP models' efficiency in pointing out sensitive keywords.
- **ROC-AUC Analysis:** It was also utilized to assess the functionality of the developed access control mechanisms in the differentiation procedure between authorized and prohibited attempts.

## 3.2 Evaluation Methodology

The performance of the ISDE framework was estimated during four tests, followed by analyses if the framework is effective, efficient, and scalable.

### 3.2.1 Performance Evaluation

Thus, its measure of performance by means of encryption efficiency, access control accuracy and resource usage determines the computational complexity of the framework, potential and scalability.

**1. Encryption Efficiency:** The duration used in encrypting data in relation to the sensitivity scores calculated was assessed so as to gauge the efficiency of the ABE computation.

**2. Access Control Accuracy:** The access control mechanisms regional control was measured by determining the percent difference between the predicted access and the permission allowed.

**3. Resource Utilization:** The amount of computational resources used by the encryption and access control was also monitored so that the resource utilization of the framework is kept optimal.

### 3.2.2 Security Evaluation

Security evaluation in fact checks the effectiveness of the framework's security by threat modeling and by the vulnerability testing in order to protect the structure from unauthorized users, data leakage, and insider threats.

**1. Threat Modeling:** Different threat cases were explored to measure the framework's effectiveness against the major security threats including violates, breach, and insider threats.

**2. Vulnerability Assessment:** Several residual risk assessments and penetration testing tests were also conducted on the framework to check for weak or insecure areas

### 3.2.3 Usability Evaluation

Here, usability is determined by incorporating the user satisfaction feedback and comparing the results with other quantitative parameters such as time taken to perform specific tasks as well their error rates to determine the efficiency of the formulated framework.

**1. User Experience:** System usage of the ISDE framework was surveyed by gathering feedback from the users who worked with this framework to determine the comprehensiveness of the framework in terms of its readiness to be implemented into the working environment.

**2. Performance Metrics:** Therefore, such effectiveness measures as the time spent on the completion of a given task and the rates of errors made were used to capture the overall satisfaction of the framework by users.

By applying this systematic research process, the current study guarantees that the enhanced as well as the evaluative process of the ISDE framework is scientifically sound. Through the integration of ABE with advanced access control and NLP techniques, the performance is thoroughly evaluated to prove that it can significantly improve data privacy and security in the era of big data.

# 4 Design Specification

The following section focuses on the modalities that enable the techniques, architecture, and framework. The description also includes the operation of the proposed model that combines

ABE with progressed access control and Sensitive Keyword Weight age Scoring System components.

## 4.1 Techniques and Framework

After the end of the ISDE framework, it uses several techniques to maintain the confidentiality and optimize the exploitation of resources. The key components are:

### 4.1.1 Attribute-Based Encryption (ABE)

The ABE involves the /AE strategies, which enables the encryption and decryption practices to occur in relation to the users' attributes. By doing so, it enables very specific levels of data security that only users who possess certain attributes can read data identified by certain parameters. In this project, ABE is applied to provide encryption only to the sensitive data that has been categorized based on the sensitivity score attached to the phrases or the chunks of data which is to be processed.

### 4.1.2 Natural Language Processing (NLP)

NLP is used to analyze and process the textual data which identifying sensitive information that requires encryption. When the text is parsed the system can easily identify which parts of the text may contain sensitive information that requires encryption.

### 4.1.3 Sensitive Keyword Weight age Scoring System

The system gives keywords and phrases the sensitivity scores depending on the use and relevance to the text. The scores are adopted based on the necessity and degree of encryption needed for parts of the data. This selective encryption reduces unnecessary encryption thus making the best use of available computational resources.

## 4.2 System Architecture

The design of the ISDE framework is such that it is comprised of these techniques can be provide a comprehensive data encryption and access control system. As presented in the work's Figure 1, the architecture and the workflow of the work are shown. The diagram includes five phases all of which are in charge of offering a particular section associated with the data encryption and access control. Below is a detailed explanation of each phase

### Phase I: File Pre-Process

**Data Owner:** The data owner who provides a file that needs to be processed and preferably encrypted.

**File Pre-Processing:** Structures the file and have it ready for other relevant manipulations. Here the data set has to be pre-processed so that it can be analyzed using NLP, this includes formatting, cleaning and organizing the data.

**NLP with Term Frequency:** Natural Language Processing (NLP) techniques are applied to analyze the text within the file. Term frequency analysis helps identify important and frequently occurring terms.

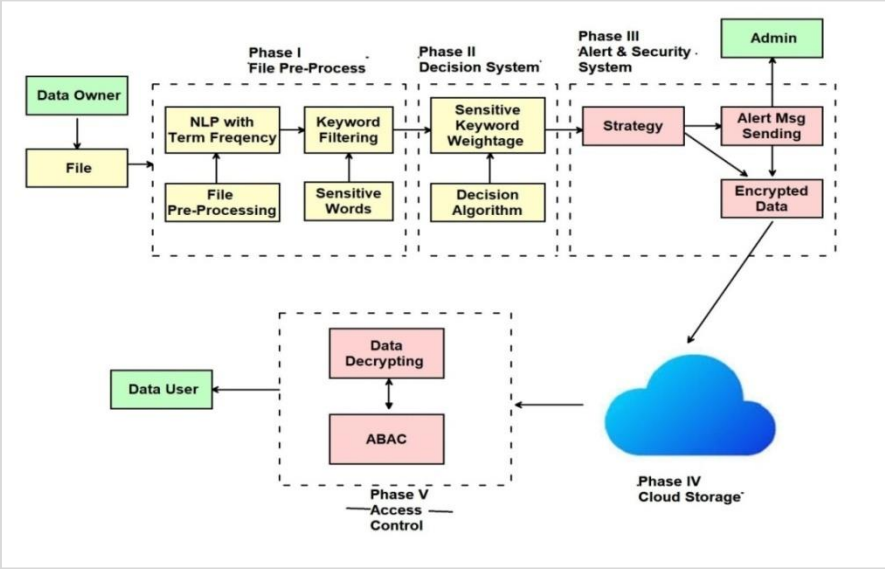


**Figure 1: The proposed work's system flow**

**Keyword Filtering:** Using term frequency analysis and the other NLP methods, other sensitive words are eliminated. These words are followed for further processing.

## Phase II: Decision System

**Sensitive Keyword Weight age:** The identified sensitive words are given the sensitivity scores according to the context and significance of each word. This scoring is useful in determining the level of encryption required on each of the segments of the data.

**Decision Algorithm:** A decision algorithm analyzes the numerical value of sensitivity scores and other contextual parameters regarding the treatment of data components and makes decisions about their encryption.

## Phase III: Alert & Security System

**Strategy:** On the basis of the decisions made in the course of the phase II, an approach is formed to encrypt the identified information and exercise access control.

**Alert Message Sending:** An alert message is then created for admin know the sensitive data and the encryption processes to be made.

**Encrypted Data:** The distinct sensitive data is encrypted with the help of Attribute-Based Encryption (ABE). This makes it that only the users with certain attributes can decode and access the data.

### Phase IV: Cloud Storage

After encrypting the data, it is then kept in a cloud system. This cloud protects the data and makes it easily retrievable by the people who have the access to encrypted data.

### Phase V: Access Control

**Data User:** Those people who have a right to access the encrypted data.

**ABAC (Attribute-Based Access Control):** Attribute-Based Access Control (ABAC) polices is adopted to ensure proper control of access. Such policies help to guarantee that only the clients with the right attributes can decrypt and view the information.

**Data Decrypting:** Authorized users have the ability to decode information using their attributes, and for this reasons, they can be access the information securely.

**Admin:** It can be said that admin is critical for overall system surveillance, receiving alerts about sensitive data, and it is responsible for the proper implementation of the encryption and access control strategies.

Thus, applied in ISDE framework, NLP, sensitivity scoring, ABE, and ABAC work complementarily and provide comprehensive protection of data and secure access. Encrypting only the sensitive data and the control of this data through user attributes can contribute to the quantitative optimization of the system and qualitative optimization of the information, necessary for the functioning of the organism, in the conditions of the digital-era.

# 5 Implementation

The implementation of the system covers the roles of Admin, Data Owner, and Data User. It includes the login and home page interfaces for each role, functionalities for creating and managing data owners and users, file uploads, and sensitive file handling. Additionally, it describes how files are encrypted, stored in DriveHQ Cloud, and accessed by users.

**Admin:** As we know that, we have 3 persons in our project Admin, Data Owner and Data user respectively. Admin can create himself and using those credentials he can able to login in admin login form and then he will be redirected to admin home page shown in Figure 2.


**Figure 2: Admin login**

In the admin home we have several options like, crate data owner, view the data owners list, edit the data owner, view the data user's lists that are created by different data owners and view the files which are uploaded by all data owners. The Figure 3 showing the list of data users those are created by all data owners.

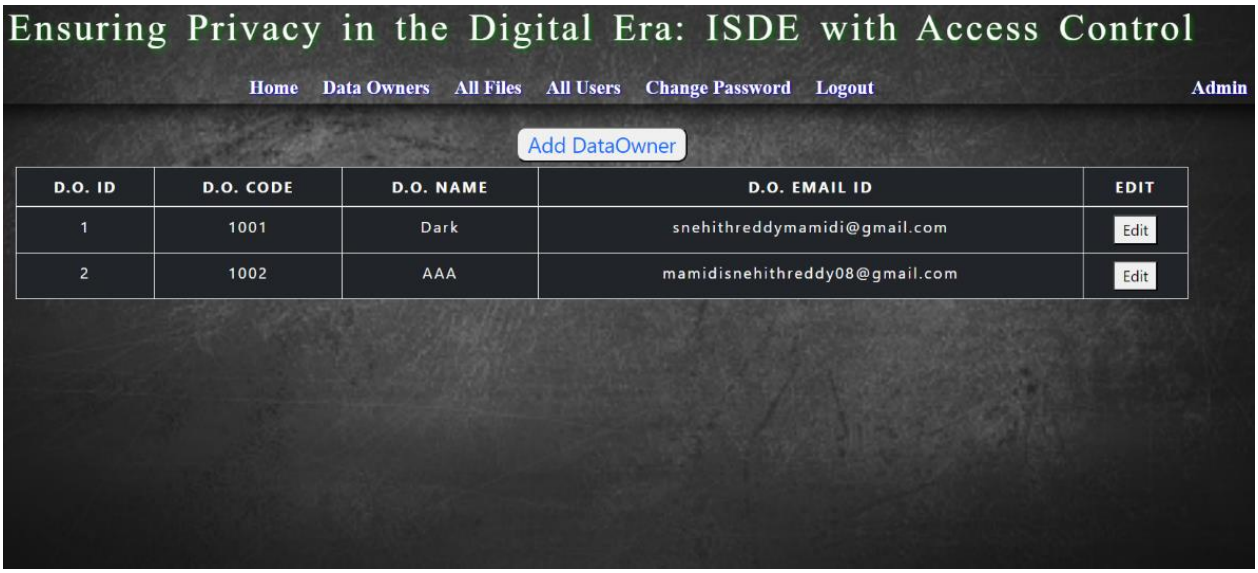


**Figure 3: Admin Home page details and Users list**

After Clicking on the Add data owner button admin redirected into Create data owner page there he can provide the required details and he can create the data owner with encryption algorithm as shown in Figure 4.

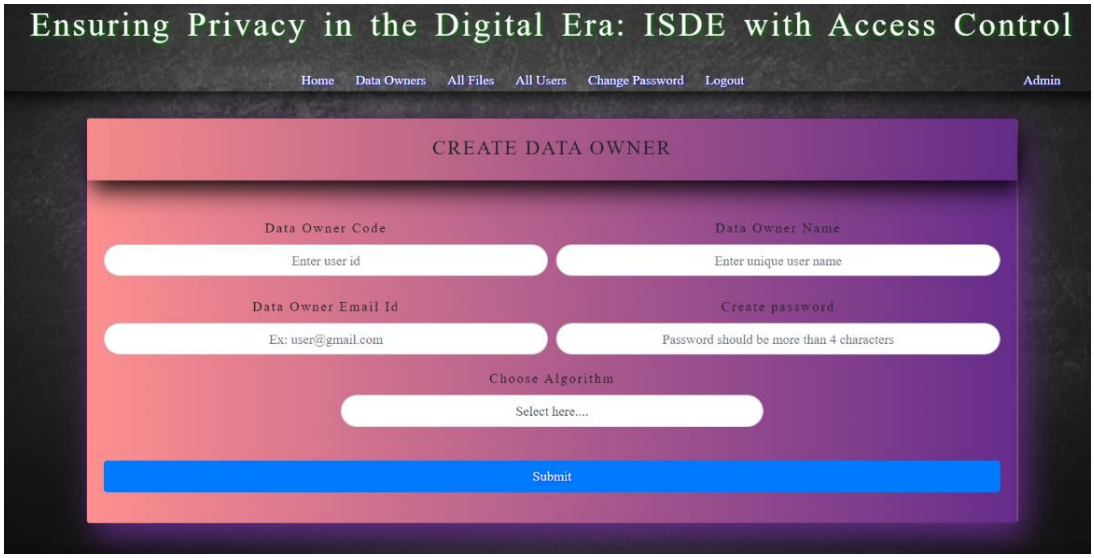


**Figure 4: Create Data Owner**

Here admin can see the all files which are uploaded by all data owners, and he can see which data owner uploaded that particular file as shown in Figure 5.

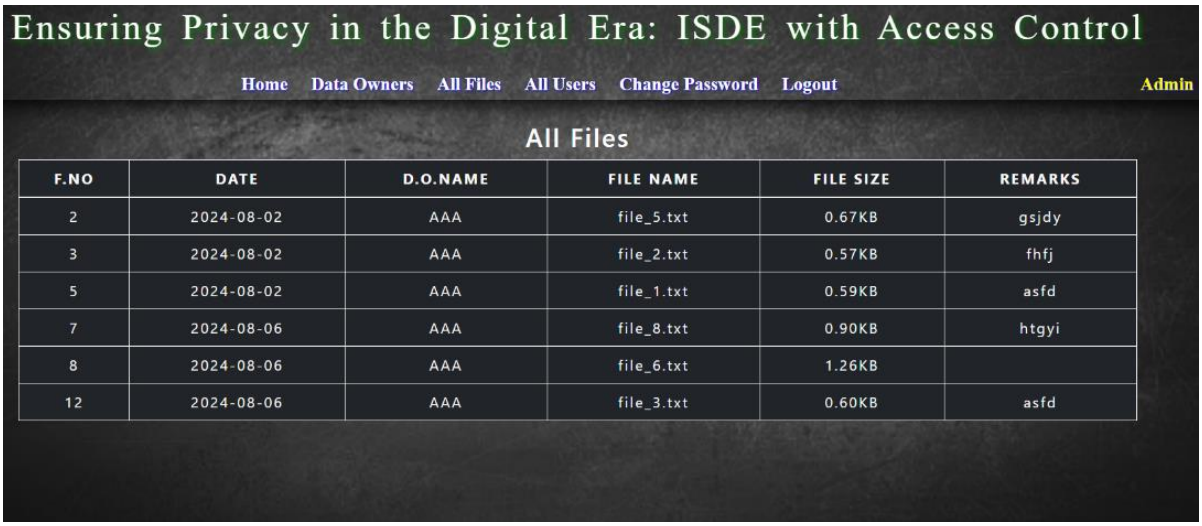


**Figure 5: Data Owner uploaded files**

## Data Owner Sign-in

As shown in the above Figure 7 we have login page for data owner, he can login there using the credentials which are given by admin through mail. After successful login he will be redirected to the data owner home page Figure 6.


**Figure 6: Home Page of Data Owner**


**Figure 7: Login Page for Data Owner**

Data owner can create the data user by giving required data and that data user will get the decryption keys from his data owner as shown in Figure 8.

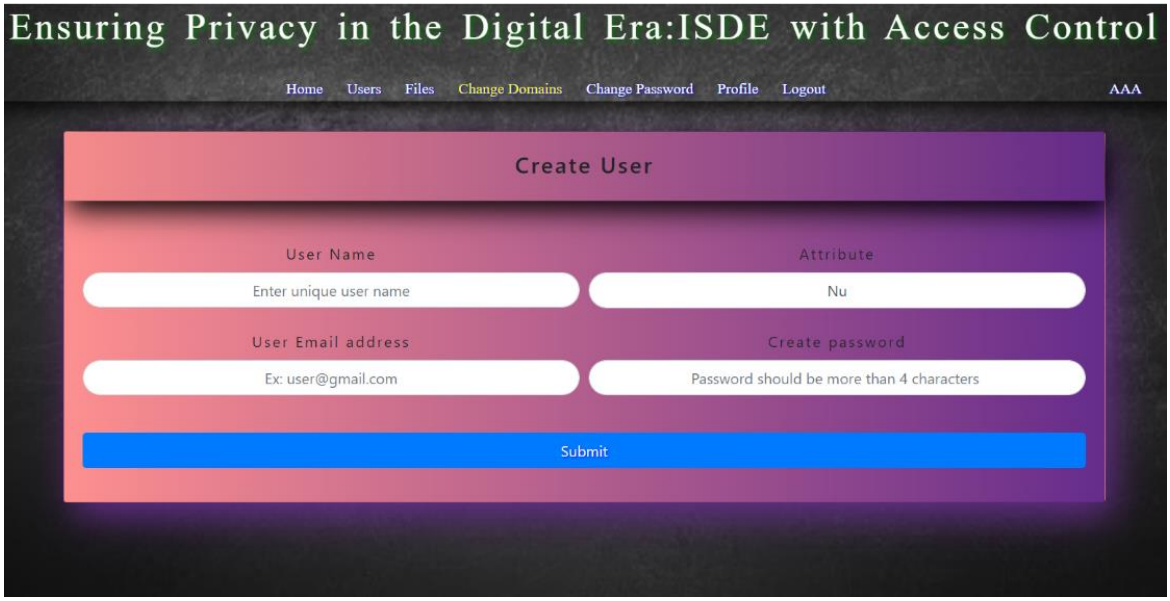


**Figure8: Create User Page**

Here data owner can upload the files from his local memory and he can write remarks for that file, now we are encrypting that file and saving in cloud named as DriveHQ cloud.


**Figure 9: Data Owner File Upload**

As soon as a user uploads a file with sensitive information, a pop-up message will appear to guarantee increased protection Figure 10. Non-sensitive files go directly to the cloud for uploading. Administrators are alerted immediately when there is an attempt of cancelling finished upload figure 11 and will send Trigger Message as shown in Figure 12.


**Figure 10: Sensitive File Trigger Message**

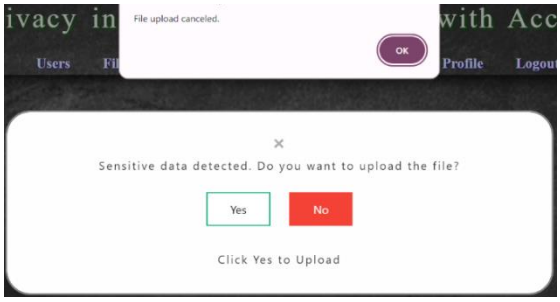


**Figure 11: File Cancel Trigger Message**

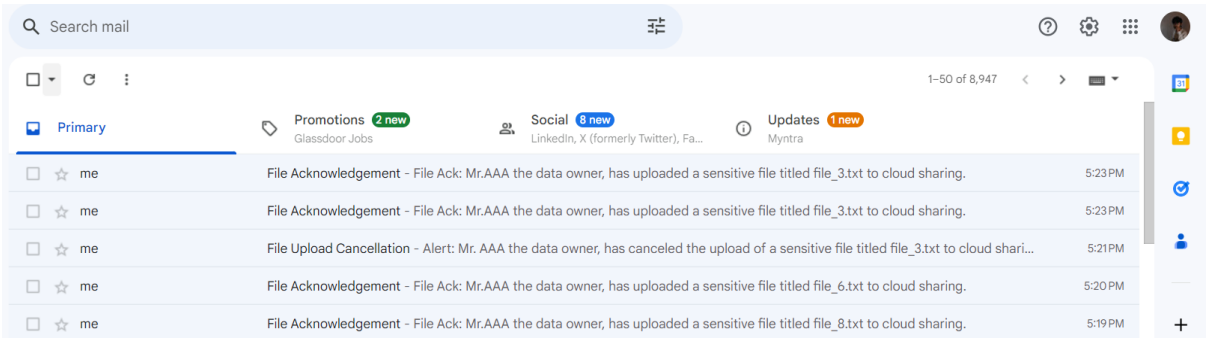


**Figure 12: File Cancel Trigger Message**

This is the Drive HQ cloud here we can save our encrypted files and later we can retrieve using ftp server as Shown in Figure 13.

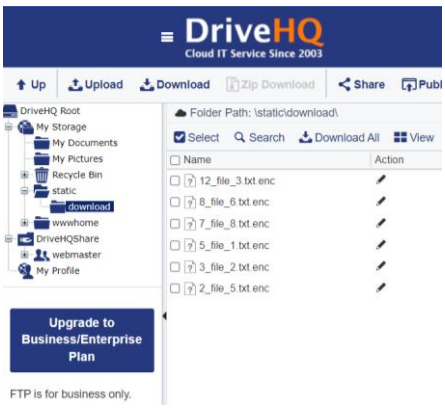


**Figure13: DriveHQ Cloud**

The data user can download the file by clicking the download button as shown in Figure 14.


**Figure14: User Download**

The data user can download files from the DriveHQ cloud. Access control determines the user's ability to download from the cloud as Shown in Figure 15.


**Figure15: File Download from DriveHQ Cloud**

# .6 Discussion

The results of the experiments, which were conducted in order to evaluate the performance of the ISDE framework, reveal the enhancement of the primary objectives of data security and resource management in the context of the cloud settings. The approach of extending ABE with NLP was identified as useful and ensured that only the necessary data is encrypted, lessening a computational overhead thus lowering the time taken for encryption. The ABAC mechanism applied for enhancement of the mechanics of access control decision also proved handy, while the ISDE framework also manifested of superior effectiveness in relation to conventional means in terms of consumption of resources. Also, the capacity of the system to notify the administrators in real-time whenever sensitive files are uploaded added to security supervision. The alert feature of the application helps organisations to track any emerging risks that can endanger an organisation and take appropriate measures regarding the management of sensitive data. For that reason, there were no significant security issues when applying the ISDE framework since it had attributes access control and selective use of encryption. However, some areas for improvement were identified: using the more accurate

models of NLP to have fewer false positives and negatives, refining ABE algorithms to use fewer computations, expanding the framework for handling big data and deal with more intricate access control questions. Consequently, findings from this study reveal that the ISDE framework attained technical efficiency superiority, data encryption and decryption, and utility value alerting systems other studies. Finding of this work indicates that future work should be aimed toward application of the advanced NLP methods, improving algorithms, increasing capacity of the system and the improvement of the interfaces toward the use of the framework for cloud data security.

# 7 Conclusion and Future Work

To solve challenge of security of the data whilst implementing a solution that incurs a small computational overhead and does not significantly add to the storage costs of data in the cloud, this research proposed the Innovative Strategies for Data Encryption (ISDE) framework. Thus, the proposed framework that combines ABE and NLP with Sensitive Keyword Weight age Scoring System allows minimizing computational and space requirements by selectively encrypting information that is more sensitive. Some of the conclusions are as follows: The method of selective encryption, ABE, and the integration of these results in using resources more efficiently. However, some of the existing challenges include scalability and the general implementation issues.

In future work, more efforts should also be made on the scalability, more deep study on advanced NLP techniques, the integration with other security mechanisms, the user interface improvement and the business application to commercialize ISDE in the field of cloud data security.

# References


Al-Janabi, M. and Altamimi, A.M., 2020, November. A comparative analysis of machine learning techniques for classification and detection of malware. In *2020 21st International Arab Conference on Information Technology (ACIT)* (pp. 1-9). IEEE.

Aslan, Ö.A. and Samet, R., 2020. A comprehensive review on malware detection approaches. *IEEE access*, *8*, pp.6249-6271.

Einy, S., Oz, C. and Navaei, Y.D., 2021. The anomaly-and signature-based IDS for network security using hybrid inference systems. *Mathematical Problems in Engineering*, *2021*(1), p.6639714.

El Mestari, S.Z., Lenzini, G. and Demirci, H., 2024. Preserving data privacy in machine learning systems. *Computers & Security*, *137*, p.103605.

Gouert, C., Mouris, D. and Tsoutsos, N., 2023. Sok: New insights into fully homomorphic encryption libraries via standardized benchmarks. *Proceedings on privacy enhancing technologies*.

Guerra-Manzanares, A., Lopez, L.J.L., Maniatakos, M. and Shamout, F.E., 2023, May. Privacy-preserving machine learning for healthcare: open challenges and future perspectives. In *International Workshop on Trustworthy Machine Learning for Healthcare* (pp. 25-40). Cham: Springer Nature Switzerland.

IzadiYekta, H., 2022. An efficient and privacy-preserving federated learning scheme.

Jiang, Z., Wang, W. and Liu, Y., 2021. Flashe: Additively symmetric homomorphic encryption for cross-silo federated learning. *arXiv preprint arXiv:2109.00675*.

Khadse, N.S., Mate, S.P., Goda, Y.R. and Harde, J., 2021. Implementation of Identity Based Encryption with Outsourced User Revocation in Cloud Computing.

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A. and Qadir, J., 2023. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, p.106848.Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A. and Qadir, J., 2023. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, p.106848.

Khan T, Budzys M, Nguyen K, Michalas A. Wildest Dreams: Reproducible Research in Privacy-preserving Neural Network Training. arXiv preprint arXiv:2403.03592. 2024 Mar 6.

Khan, M., Glavin, F.G. and Nickles, M., 2023. Federated learning as a privacy solution-an overview. *Procedia Computer Science*, *217*, pp.316-325.

Kimmel, J.C., Mcdole, A.D., Abdelsalam, M., Gupta, M. and Sandhu, R., 2021. Recurrent neural networks based online behavioural malware detection techniques for cloud infrastructure. *IEEE Access*, *9*, pp.68066-68080.

Kimmell, J.C., Abdelsalam, M. and Gupta, M., 2021, August. Analyzing machine learning approaches for online malware detection in cloud. In *2021 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 189-196). IEEE.

Lemieux, V.L. and Werner, J., 2024. Protecting Privacy in Digital Records: The Potential of Privacy-Enhancing Technologies. *ACM Journal on Computing and Cultural Heritage*, *16*(4), pp.1-18.

Li, J., Meng, Y., Ma, L., Du, S., Zhu, H., Pei, Q. and Shen, X., 2021. A federated learning based privacy-preserving smart healthcare system. *IEEE Transactions on Industrial Informatics*, *18*(3).

Lin, B.Y., He, C., Zeng, Z., Wang, H., Huang, Y., Dupuy, C., Gupta, R., Soltanolkotabi, M., Ren, X. and Avestimehr, S., 2021. Fednlp: Benchmarking federated learning methods for natural language processing tasks. *arXiv preprint arXiv:2104.08815*.

Mittal, S., Bansal, A., Gupta, D., Juneja, S., Turabieh, H., Elarabawy, M.M., Sharma, A. and Bitsue, Z.K., 2022. Using identity-based cryptography as a foundation for an effective and

secure cloud model for e-health. *Computational Intelligence and Neuroscience*, *2022*(1), p.7016554.

Mothukuri, V., Parizi, R.M., Pouriyeh, S., Huang, Y., Dehghantanha, A. and Srivastava, G., 2021. A survey on security and privacy of federated learning. *Future Generation Computer Systems*, *115*, pp.619-640.

Mousa, N. and Shirazi, F., 2024. A survey analysis of quantum computing adoption and the paradigm of privacy engineering. *Security and Privacy*, p.e419.

Nagy, B., Hegedűs, I., Sándor, N., Egedi, B., Mehmood, H., Saravanan, K., Lóki, G. and Kiss, Á., 2023. Privacy-preserving Federated Learning and its application to natural language processing. *Knowledge-Based Systems*, *268*, p.110475.

Punreddy, A.R., 2023. Federated Learning for Protecting Medical Data Privacy.

Rehman, A., Jian, L.I.U., Yasin, M.Q. and Keqiu, L.I., 2021. Securing cloud storage by remote data integrity check with secured key generation. *Chinese Journal of Electronics*, *30*(3), pp.489-499.

Sahay, S.K., Sharma, A. and Rathore, H., 2020. Evolution of malware and its detection techniques. In *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018* (pp. 139-150). Springer Singapore.

Salam, A., Abrar, M., Ullah, F., Khan, I.A., Amin, F. and Choi, G.S., 2023. Efficient data collaboration using multi-party privacy preserving machine learning framework. *IEEE Access*.

Samad, A., 2023. Hybrid Approaches in Threat Detection: Integrating Traditional Signature-Based Methods with AI and ML Techniques for Enhanced Accuracy.

Saravanakumar, S. and Chitra, S., 2022. Hybrid Cloud Security by Revocable KUNodes-Storage with Identity-Based Encryption. *Computer Systems Science & Engineering*, *43*(3).

Shivasree, S.M.R.J.Y., 2022. INTELLIGENT ANOMALY AWARE BEHAVIOR-BASED MALWARE DETECTION SYSTEM FOR DYNAMIC CLOUD ENVIRONMENT. *NeuroQuantology*, *20*(10), p.2206.

Sun, S., Du, R., Chen, S. and Li, W., 2021. Block chain -based IoT access control system: towards security, lightweight, and cross-domain. *Ieee Access*, *9*, pp.36868-36878.

Unal, D., Al-Ali, A., Catak, F.O. and Hammoudeh, M., 2021. A secure and efficient Internet of Things cloud encryption scheme with forensics investigation compatibility based on identity-based encryption. *Future Generation Computer Systems*, *125*, pp.433-445.

Wang, Shangping, Xu Wang, and Yaling Zhang. "A secure cloud storage framework with access control based on block chain ." IEEE access 7 (2019): 112713-112725.

Aslan, Ömer, MerveOzkan-Okay, and Deepti Gupta. "Intelligent behavior-based malware detection system on cloud computing environment." IEEE Access 9 (2021): 83252-83271.

Xue, Yingjie, et al. "An attribute-based controlled collaborative access control scheme for public cloud storage." IEEE Transactions on Information Forensics and Security 14.11 (2019): 2927-2942.

Ogiela, Lidia, Marek R. Ogiela, and HoonKo. "Intelligent data management and security in cloud computing." Sensors 20.12 (2020): 3458.

C. Yang, L. Tan, N. Shi, B. Xu, Y. Cao and K. Yu, "AuthPrivacyChain: A Block chain - Based Access Control Framework sswith Privacy Protection in Cloud," in IEEE Access, vol. 8, pp. 70604-70615, 2020, doi: 10.1109/ACCESS.2020.2985762.

S. Yu, W. Zhou, S. Guo, and M. Guo. A feasible IP tracebackframework through dynamic deterministic packet marking. IEEETransactions on Computers, 65(5):1418– 1427, 2016.

Y. Yu, M. Au, G. Ateniese, X. Huang, W. Susilo, Y. Dai, andG. Min. Identity-based remote data integrity checking with perfectdata privacy preserving for cloud storage. IEEE Transactions onInformation Forensics and Security, PP(99):1, 2016.

S. Rho, A. Vasilakos, and W. Chen. Cyber physical systems technologiesand applications. Future Generation Computer Systems, 56:436–437, 2016.

M. Shabbir et al., "Enhancing Security of Health Information Using Modular Encryption Standard in Mobile Cloud Computing," in IEEE Access, vol. 9, pp. 8820- 8834, 2021, doi: 10.1109/ACCESS.2021.3049564.

D. Mahendran, C. Luo and B. T. Mcinnes, "Review: Privacy-Preservation in the Context of Natural Language Processing," in IEEE Access, vol. 9, pp. 147600- 147612, 2021, doi: 10.1109/ACCESS.2021.3124163.

F. Martinelli, F. Marulli, F. Mercaldo, S. Marrone and A. Santone, "Enhanced Privacy and Data Protection using Natural Language Processing and Artificial Intelligence," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206801.