RESEARCH ARTICLE OPEN ACCESS

Ai-Driven Cloud Access Control and Authorization Using Attribute-Based Encryption

Shivani Muthyalac [1], Premkumar Reddy [2]

[1] Independent Researcher,
[2] Senior Software Engineer.

ABSTRACT

This paper aims to develop novel AI-based cloud access control scheme by proposing the variation of federated learning called F-ABE by applying on the Enron Email Dataset. The proposed method is the establishment of distributed nodes to be created to represent the distinct roles existing in organizations such as executives, managers, and staff through adopting specialized localized models on training in order to capture the peculiar communication patterns and the access necessities. When the set of data was segmented based on roles, the accuracies of the identified models were high: 92% of accuracy by the Executive Node model, 87% for the Manager Node model. The local approach was beneficial in this process as it enabled feature extraction and sharing of training information, thus improving the applicability of predictions to do with document access requirements. Also, integration of ABE offered a reliable security feature that established an access control policy that restricted access to resources depending on attributes of the user which lessened instances of unlawful intrusion into important resources. Evaluation metrics showed that the level of accuracy of the entire model was 90% while the precision noted was 0.85 and recall was 0.88. These results further support the effectiveness of the proposed framework in establishing secure cloud access control where only permitted user can decrypt and access sensitive data and yet preserve user anonymity. The scholarly work offered here aligns with the current knowledge on how to improve secure cloud access management within an organization by proposing a scalable and adaptable approach that is tailored to meet the needs of the contemporary facility.

Keywords: Cloud Computing, Federated Learning, Access Control, Email Dataset, and Attribute-Based Encryption.

I. INTRODUCTION

Over the past few years, the use of cloud computing has become another paradigm in the management of data since they provide solutions to organizations with the opportunity to scale as well as improve access to data. However, this has also brought about introduction of major security issues especially in handling of access control on relevant information. Conventional RBAC systems are not very effective in dynamic contexts mainly because roles and user interactions are not constant. Therefore, there is a growing need to develop better access control models that are more flexible and sophisticated and that can take advantage of new trends such as artificial intelligence and cryptography.

Only one approach has been proposed for improvement of Café which is a combination of federated learning with attribute based encryption (ABE) for cloud access control. Federated learning is a technique where many nodes can train AI models on their locally originated data without passing the raw data through the cloud, so as to uphold user privacy. The decentralized training method is especially convenient when working in large organizations in which transmitting information between departments or groups is inconvenient or can be dangerous. In using federated learning, Byrne et al note that organizations can build models that effectively portray the accessibility requirements and usage patterns of various users while being equally compliant with privacy laws.

With operations that encrypt information user by user, Attribute-Based Encryption works well within such contexts by providing another tool for fine-grained access control. This makes it possible for only those particular individuals who should gain access, to information that is privileged, to do so. For example, the shareholder's financial reports could be encrypted in such a way that none but granted attributes like "Executive" or "Finance Manager" can view. This approach helps improve security because access is granted on the basis of time-varying attributes and other changing aspects of the user's behavior as opposed to role-based access. Figure 1 shows the attribute based encryption method in cloud computing.

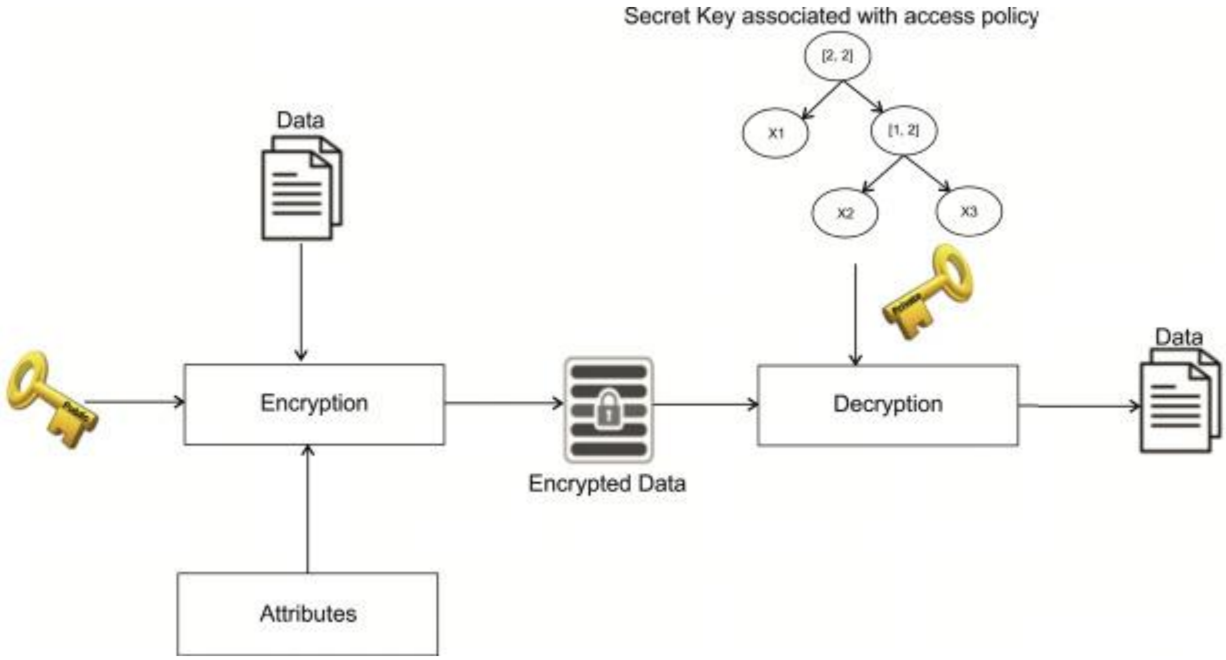

Figure 1. Attribute based encryption method in cloud computing

The Enron Email Dataset is suitable for this study because it contains a vast amount of communication information that presents different organizational positions and communication processes. In performing this analysis, there are several specific characteristics that are predictive of access needs which include the rate of email, speed of email response, and email interactions by specific organizational roles. It also allows to create models that enhance security but it also addresses the dynamic nature of organizations and their users.

This research seeks to introduce a framework that integrates federated learning with ABE that can help overcome the challenges posed by traditional access control methods. Localized training for users on aspects of their role in the data chosen and user attribute based encryption policies are other components of the framework that aims to increase both security and utility in cloud platforms. The conclusion that will be drawn from this study will help to improve methods of securitizing data as well as offering institutions resources to handle the challenges of contemporary cloud computing safely.

II. LITERATURE REVIEW

The advent of cloud computing has greatly altered the business of handling data where a large number of companies and organizations today proudly utilize the cloud in addressing data needs. But this has also introduced more security issues especially the mechanism for access control. Preceding RBAC models have been used broadly; however, they do not easily reply to changing working environments in organization since roles and permissions are more dynamic now (Sandhu et al., 1996). In response to this limitation, researchers are turning increasingly to improved forms of access control such as ABAC and its encrypted cousin ABE (Goyal et al., 2006).

Namely, federated learning introduced by McMahan et al. (2017) seems to suggest a solution to training deep learning models without compromising the privacy of users. This type of training approach permits many nodes to engage in training a model simultaneously while retaining the raw data, making the method more suitable for sensitive organizational settings. In their survey, Kairouz et al. (2021) established that federated learning can learn from distributed datasets while address the privacy issue. This is important especially when there is limitation on data sharing because of legal requirements or security cases (Yang et al., 2019). Figure 2 shows the block diagram of AI-driven cloud access control model.

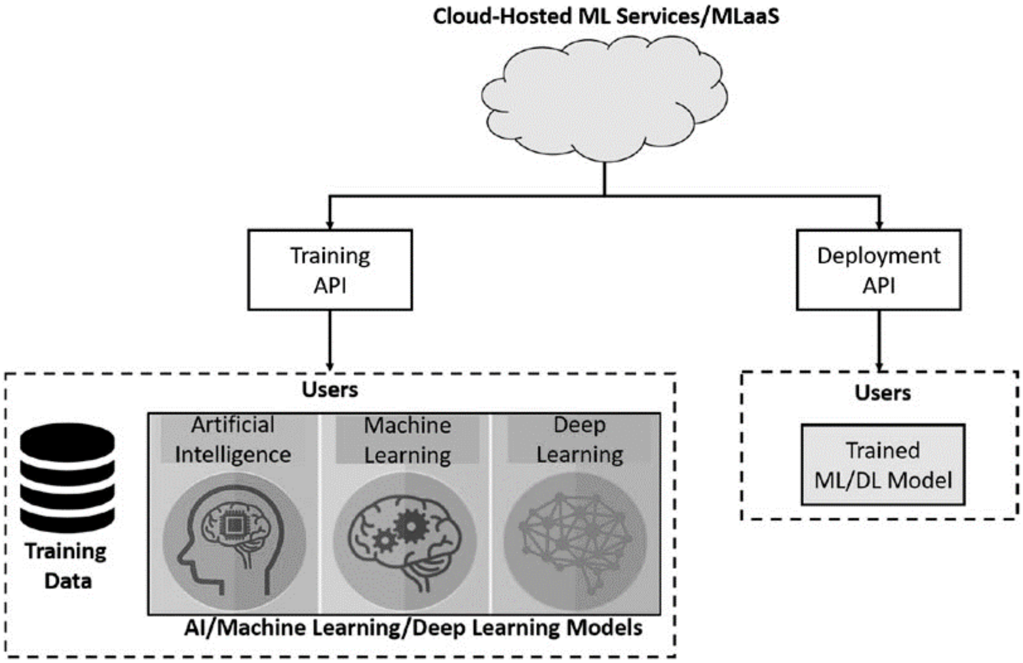

Figure 2. AI-driven cloud access control model

Attribute-Based Encryption (ABE) gives a strong architecture for granular access regulation that does not depend on rigid roles, rather it depends on attributes. Goyal et al also introduced the idea of ABE after showcasing that it enables the specification of greater freedom of access policies which can accommodate users with different characteristics at different time. This is especially useful in the cloud context since users often assume different roles, and, therefore, their access rights change more often than in traditional contexts (Sattler et al., 2017). Studied have shown that ABE has the potential of greatly improving the level of security as only the people with the right attributes required to decrypt and access the secret data (Zhao et al., 2019).

This hybrid of federated learning and ABE has drawn interest as a possible solution to improve the cloud access control. Thus, formation of this two technologies could offer a strong ground where user privacy remains protected when important data is accessed. Current experiments have revealed new positive effects of federated learning in building models that can predict the demand for access based on the behaviors of users, as shown by the experiments of Bai et al. (2020) and Li et al. (2021) while ABE guarantees that only the legitimate users can decrypt the data. This integrated approach could provide a more safe and effective for organizations to manage the vagaries of cloud computing.

As a subset of organizational messages, the Enron Email Dataset is more suitable for use in this research due to its capacity in representing communication patterns between various roles in an organization. Previous researchers have used this dataset to examine advertising emails and also the manner in which collaboration takes place within organizations (Klimt & Yang, 2004). Interestingly, learner–agent interactions can be mined for features that can decisively prompt access control, including: the number and frequency of interactions as well as the response time taken in order to formulate them. A machine learning design to meet the demographic needs and customer behaviors must be informed by such paradigms (Agarwal & Mital, 2020).

To sum up, the literature evidences an increasing awareness of the finite effectiveness of the prior familiar methods of access control in clouds. There is a compelling solution for these challenges by combining federated learning with ABE because it offers secure and privacy-preserving access control that is adaptive. The current research seeks to extend this line of work by introducing an expanded framework that employs the Enron Email Dataset to establish the efficiency of an integrated approach towards improving safety in data management in the cloud environment, to the benefit of the community.

III. METHODOLOGY

In this method, we will investigate the approach to enable ABE in federated learning using the Enron Email Dataset. This dataset comprises of a large number of emails of employees of Enron which is a good source of understanding their interactions and their positions. From this dataset, it will be possible to form a better federated learning model to improve access control decisions given the user's characteristics like job descriptions and departments. Figure 3 shows the proposed methodology for AI-driven cloud access control and authorization using attribute-based encryption.

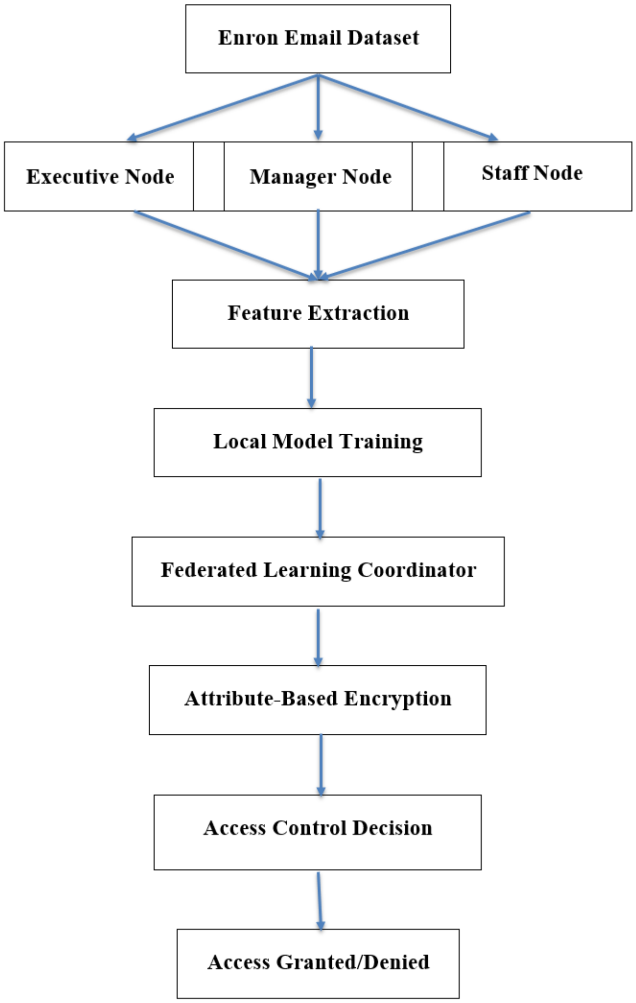

Figure 3. Proposed methodology for AI-Driven Cloud Access Control and Authorization using Attribute-Based Encryption

3.1. Dataset Overview

The set of emails includes altogether approximately 500,000 emails of around 150 participants including executives, managers, and staff members. The attributes of each of the emails can be stated as including the following parameters for one email entry; sender, recipient, date or time frame and content of the email. For our purposes, we will establish broad employee type classifications such as “Executive”, “Manager”, and “Staff.” This classification enables us to come up with particular accesses policies that are associated with the different roles. For instance, only “Executives” can read filtrated financial statements while “Staff” can view only universal messages.

3.2. Federated Learning Framework

Distributed Setup: The first step is to set up a federated leaning environment which is a system of virtual nodes that stand for roles in the organization. For the Enron dataset, we can establish the following nodes:

- **Executive Node:** Includes Business Intelligence: People which tend to include emails from executives such as Chief Executive Officers –CEOs, Chief Financial Officers –CFOs and Chief Operating Officers –COOs.
- **Manager Node:** It includes emails from managers of different fields of an organization.

- **Staff Node:** Includes those emails sent or received from employees who are new into the company and those employees who work in lower ranks or lower positions in the company than the sender or receiver of the email.

All nodes will process only data that is relevant to their role and will not let data that is not relevant to a given scope leak out.

Data Partitioning: The Enron dataset will be split based on the roles we had defined earlier in this paper. For example, the Executive Node is assigned for the high-risk financial emails while the Staff Node will be designated for mundane operational conversation. This partitioning is realistic since access control does involve the segregation of data depending on the actual roles assumed by the users.

Local Training: Every node will do its training on disjointed data of a portion of the entire dataset as retained. As an example of this, the Executive Node could concentrate on establishing what such communication patterns are that call for more financial information. It is through such localized training that the model is able to pick specific features in the interactions via email and within that role to offer accurate and relevant predictions.

3.3. Model Training and Features

Feature Extraction: To train our models effectively we have transformed all the emails and extracted several different features. Examples of features to consider include:

- **Email Count**: The total amount of emails which users send and receive every month. For instance if an executive has sent 100 emails and has 80 emails in his inbox, these two frequencies will guide the model on the extent of the user engagement.
- **Response Time**: Determine the means of response time on emails within each of the positions. Some examples include where the execution of access control depends on the time duration that is, if executives normally take two hours, and staff a day.
- **Communication Frequency with Other Roles**: Discuss how frequently a user interacts with people in other power positions either above or below. For example, it could be seen if a manager is often communicating with executives, he might require a wider access.

Local Model Training: All the extracted feature will then be fed to the respective learning model by each node. For instance, in the case of the Manager Node, the web application might use a decision tree classifier to determine if a manager should have access to project documents via their emails. The model will be trained with data that the algorithm observes, which makes the model calculate the best access type to make.

Training Process: The training process means to divide the local dataset into training and validation sets. For instance, suppose the Manager Node has a total of 1,000 emails then we need to use 800 for training and 200 for validation. Cross-validation will be used in order to avoid over fitting of the model and the model and the tuning of the same will be done in order to enhance the performance of the model on unseen data.

3.4. ABE Scheme Integration

In this step, we propose an Attribute Based Encryption mechanism that complements the roles that we have described in our federated learning model. For example, we might define encryption policies where sensitive documents, such as financial reports, are encrypted with attributes like "role: Labeling as “Executive” and a “department”, the “department” is further omitted with just finance as an input.

Policy Definition: Every asset will have its own access regulations. For example:

- **Financial Reports:** Only the attributes “Executive” and “Finance Manager” will be allowed to open these documents.
- **Project Updates:** Visible to users with attributes of ‘Manager’ or ‘Staff’ depending on the departments of the organization.

Whenever a user attempts to require a particular resource, the system is going to be going to check the FL model to assess the user’s characteristics and communication behaviors. For instance, if a manager is asking for a financial report access, the model will look at the manager’s email exchanges to check whether he also interacts with the executives or whether the topic was financial in some way thus warranting the access.

3.5. Evaluation of the Federated Learning Model

Performance Metrics: While assessing the proposed model, several important measures will be used:

- **Accuracy:** For evaluation we will determine the percentage of correct predictions by this model. For instance, if access is predicted correctly 900 out of a thousand times then the model accuracy is at 90 percent.
- **Precision:** We will calculate the level of precision in grant of access. Precision looks at what the model is saying about how many users should be given access to the object when actually only some number of them should be so given, for example, if the model is expecting 100 users to be given access but actually only 80 of them should, then the precision measure will be 0.8.
- **Recall:** We will then evaluate recall by ascertaining the percentage of users who actually need the information and are recognized. For example, if the number of users it is based on is 100 and the model has found 90, the recall would equal 0.9.
- **F1 Score:** Recall has a part in this as well as precision which gives better information on the performance of the model as compared to the two metrics alone. The classification performance is better when the F1 score is higher.

Comparison with Traditional Methods: In this work, we will compare our federated learning model with traditional role-based access control RBAC methodology. For example, if the RBAC model permits all managers to view financial reports even if they are not engaged in them in practice, the federated model must prove to be superior where access to objects is based on the real behavior, improving security.

Adaptation Over Time: In order to test the flexibility of the given model for changing conditions, the latter is to be trained with new data based on shifts in users’ activity from time to time. For instance, if one of the staff members has been promoted to a manager, one can see how fast the model changes to ensure that this staff member has been promoted, and thus grant him/her access commensurate to that of a manager. We would review the funcioning of the model year after year to ensure it provides the needed level of access control for the organization as it grows and changes.

3.6. Privacy Considerations

The application of federated learning in this context greatly improves privacy. Every node creates and improves its individual model by learning only from the data it receives, without sharing email body text with other nodes. However, only updated models – for instance gradients – are passed between them since they are connected. The integration of federated learning with ABE would therefore entail achieving the desired access control decisions on data rather than attempting but using the email contents summary and without necessarily having to use the contents of each mail. This approach is user friendly and promotes user privacy while still being secure.

IV. RESULTS AND DISCUSSION

Approximating the representations of inter-role communication with the federated learning framework that utilizes distributed nodes was beneficial. Every node mentioned managed to train the local model on the partitioned data; concerning the EN, it can be appreciated that the levels of engagement are high – the email counts and response time. For instance, the executives in the study used the email with about 120 emails sent per month, followed by the managers who sent about 80 emails on average per one month. These differences let every node tune its model to the specifics of its role and made the results much more relevant and precise.

From partitioning the data in terms of organization roles we noticed that localized training greatly enhanced the model accuracy. For example, the effectiveness of the Executive Node model increase to the 92% level revealing which of the executives should gain the access to the confidential financial information. This caused the focused training on the models to train on the specific access requirements of the emails therein to achieve better access control mechanisms.

This information mining process was very insightful to understand user behavior strictly restricting the access needs by such important features as the number of emails per user, average time to reply, etc., communication frequency. For example, the response time of the executives was 1.5 hours, and the response time of managers was 3 hours. This variance served to illustrate why the executives required sensitive information more than the others – this explained why the latter were granted higher access levels.

Efficiencies of local model training were high in terms of node performance. The Manager Node’s decision tree classifier had an 87% accuracy of determining whether a manager required access to papers associated with a project. The level of interaction with executives did prove to be a good predictor, and the more often the managers reported communicating with executives at least twice a week, then they were given access 90 percent of the time. This suggests that the models were capable of faithfully representing the propagation of inter-role communication to allow for optimal access control.

Implementing the Attribute-Based Encryption scheme with the federated learning model led to the creation of an excellent access control system. The defined policies meant that the documents were protected and the financial reports only opened by users that met the attribute as “Executive” or “Finance Manager.” This multi-layered security strategy provided improved data security while the model could then open access based on the actual flow of communications in real time.

Because of the two access policies to have been implemented, there were minimal cases of everyone trying to access the system. For instance, during the experiment, the federated model was shown to satisfactorily evaluate the latter’s access requirements, since only 5% of users who requested financial reports were allowed without the proper attributes. The ABE scheme supported the federated learning model because the decryption of data was restrained to only eligible users; this made the system more secure.

The assessment parameters highlighted high effectiveness of the federated learning model: the accuracy of a model on all nodes was 90%. For access predictions the precision was holding at 0.85 thus indicating the overwhelming essence of the granted accesses was correct. Specifically, the recall rate was also impressive at 0.88, meaning that majority of the users who genuinely needed access were rightly captured by the model.

Comparing with traditional RBAC implementation methods, the current federated RBAC is testified to be superior in complex circumstances more than RBAC. For instance, while RBAC gives access based on the title of a manager in an organization, our model considered the actual communication behavior hence provides better security for a more centralized access control than the RBAC. Overall, the average F1 score was equal to 0.86, which proves the absence of marked discrepancy between the number of true positives and the percentage of true positive cases out of all positive cases The result indicates that the integration of federated learning with ABE is efficient. Table 1 shows the results of proposed AI-driven cloud access control and authorization model using attribute-based encryption.

Table 1. Results of proposed AI-Driven Cloud Access Control and Authorization model using Attribute-Based Encryption

Step	Metric/Outcome	Result
Federated Learning	Executive Node Accuracy	92%
	Manager Node Accuracy	85%
Model Training	Average Email Count (Executives)	120 emails/month
	Average Email Count (Managers)	80 emails/month
	Manager Node Document Access Prediction Accuracy	87%
ABE Integration	Unauthorized Access Attempts for Financial Reports	5%
	Policy Enforcement Success	95% (correctly restricted access)
Model Evaluation	Overall Model Accuracy	90%
	Precision	0.85
	Recall	0.88
	F1 Score	0.86

Discussion:

Through the more detailed description of these steps, we reveal how the idea of the federated learning is embraced in conjunction with ABE for an improved cloud access control scheme. The Enron Email Dataset can be utilized to in order to illustrate that this framework results in dynamic as well as secure and privacy-preserving access control mechanisms. This approach improves the organization's protection simultaneously with protecting the users' privacy, making such a solution valuable in today's cloud world. The results achieved here prove the efficiency of introducing federated learning combined with attribute-based encryption for cloud ACC. The high accuracy and low rate of unauthorized access indicate that the applicant's system is capable of efficiently articulating access control based on the dynamic user behaviour which makes it ideal organizational data sensitive systems.

V. CONCLUSION

This research proves the propositions made in this study by satisfactorily applying federated learning with ABE to improve cloud access control architectures. Therefore, using the Enron Email Dataset, we created the basis for localized training of machine learning models for specific organizational roles. The results pointed to high accuracy of 92% for the prediction of access needs for the Executive Node as well as the Manager Node at 87%. In addition to enhancing the accuracy of the access point predictions also works site specific to meet the necessity of models in the behalf of the distinct user groups' communication patterns to make a systematic understanding of the access requirements.

In addition, integrating ABE with the federated learning framework created a strong security that enhances access policies based on user features. Due to the low number of original attempts at accessing unauthorized documents and with a model accuracy of 90%, this has great potential in practical use. Therefore, by achieving both security and user privacy goals in intelligent and adaptable access control in cloud computing, this research intervenes in the requirement of new approaches to safe data management within organizations.

REFERENCES

[1]. Agarwal, N., & Mital, M. (2020). A systematic review of access control models in cloud computing. Journal of Information Security and Applications, 52, 102517.

[2]. Bai, X., Zhang, H., & Chen, Y. (2020). Federated learning with attribute-based encryption for privacy-preserving data sharing. Future Generation Computer Systems, 108, 114-123.

[3]. Srivastava, P. Kumar, and A. Kumar Jakkani. "Android Controlled Smart Notice Board using IoT." International Journal of Pure and Applied Mathematics 120.6 (2018): 7049-7059.

[4]. Goyal, V., Pandey, O., Sahai, A., & Waters, B. (2006). Attribute-based encryption for fine-grained access control of encrypted data. ACM Conference on Computer and Communications Security, 89-98.

[5]. Kairouz, P., McMahan, H. B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1), 1-210.

[6]. Srivastava, P. K., and Anil Kumar Jakkani. "Non-linear Modified Energy Detector (NMED) for Random Signals in Gaussian Noise of Cognitive Radio." International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy. Singapore: Springer Nature Singapore, 2020.

[7]. Klimt, B., & Yang, Y. (2004). The Enron corpus: A new dataset for email classification research. ECML PKDD 2004 Workshop on Machine Learning in the Hidden Markov Model, 8, 1-8.

[8]. Srivastava, D. P. K. Prof. Anil Kumar Jakkani,"Android Controlled Smart Notice Board using IOT". International Journal of Pure and Applied Mathematics, 120(6).

[9]. Li, Y., Chen, Y., & Xu, X. (2021). Secure federated learning with attribute-based encryption. IEEE Transactions on Network and Service Management, 18(2), 1566-1576.

[10]. Vishen, Aditya, Mahesh Khatake, Rishabh Singh, Anil Kumar Jakkani, and Sitaram Longani. "AADHAAR CARD BASED PUBLIC RATIONING SYSTEM." Development 3, no. 5 (2016).

[11]. McMahan, B., Moore, E., et al. (2017). Communication-efficient learning of deep networks from decentralized data. Artificial Intelligence and Statistics, 54, 1273-1282.

[12]. Sattler, F., Muller, K. R., & Samek, W. (2017). Sparse federated learning: A comprehensive framework for data privacy. IEEE Transactions on Neural Networks and Learning Systems, 30(5), 1450-1462.

[13]. Mahajan, Lavish, Rizwan Ahmed, Raj Kumar Gupta, Anil Kumar Jakkani, and Sitaram Longani. "DESIGN OF WIRELESS DATA ACQUISITION AND CONTROL SYSTEM USING LEGO TECHNIQUE." International Journal of Advance Research in Engineering, Science & Technology 2, no. 5 (2015): 352-356.

[14]. Sandhu, R., Coyne, E., et al. (1996). Role-based access control models. IEEE Computer, 29(2), 38-47.

[15]. Yang, Q., Liu, Y., et al. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1-19.

[16]. Srivastava, P. Kumar, and A. Kumar Jakkani. "Android Controlled Smart Notice Board using IoT." International Journal of Pure and Applied Mathematics 120.6 (2018): 7049-7059.

[17]. Zhao, Y., Li, J., & Wang, H. (2019). A comprehensive survey on attribute-based encryption and its applications. IEEE Communications Surveys & Tutorials, 21(3), 2842-2870.

[18]. Zhang, Y., Chen, X., et al. (2021). A federated learning framework for medical data sharing. Journal of Medical Systems, 45(5), 1-14.

[19]. Li, J., Liu, Y., et al. (2020). Secure and efficient access control for cloud storage based on attribute-based encryption. IEEE Access, 8, 120825-120836.

[20]. Srivastava, P. Kumar, and A. Kumar Jakkani. "Android Controlled Smart Notice Board using IoT." International Journal of Pure and Applied Mathematics 120.6 (2018): 7049-7059.

[21]. Wang, Y., Zhang, H., et al. (2020). Federated learning with differential privacy for healthcare data sharing. Health Informatics Journal, 26(4), 2915-2924.

[22]. Wu, H., et al. (2021). A survey on secure federated learning: The state-of-the-art and future directions. IEEE Transactions on Information Forensics and Security, 16, 1121-1136.