Science Journal of Education
2023; 11(3): 110-116
http://www.sciencepublishinggroup.com/j/sjedu
doi: 10.11648/j.sjedu.20231103.14
ISSN: 2329-0900 (Print); ISSN: 2329-0897 (Online)

Authentication Based on Attribute Encryption with Machine Learning

Weiran Tang, Ga Xiang*, Yawei Ren

School of Information Management, Beijing Information Science and Technology University, Beijing, China

Email address:
xiaotang11100@163.com (Weiran Tang), gxianga@126.com (Ga Xiang), ryw@bistu.edu.cn (Yawei Ren)
*Corresponding author

To cite this article:
Weiran Tang, Ga Xiang, Yawei Ren. Authentication Based on Attribute Encryption with Machine Learning. *Science Journal of Education*. Vol. 11, No. 3, 2023, pp. 110-116. doi: 10.11648/j.sjedu.20231103.14

Received: May 12, 2023; **Accepted**: June 8, 2023; **Published**: June 14, 2023

Abstract: In recent years, even under the access control based on attribute encryption, the information resources in the network will inevitably be illegally obtained. There will still be the phenomenon of attackers posing as legitimate users to pass identity authentication. Therefore, it is necessary to take measures for behavior detection before identity authentication to establish trust between users and the network. After preprocessing the data set of user behavior information, the machine learning model is built to predict whether the user behavior is abnormal or not. Logical regression model, KNN model, and decision tree model are mainly built. After analyzing the prediction results, authentication is required. Therefore, an algorithm based on CP-ABE is proposed. First of all, establish a ciphertext access structure. Secondly, extract the user's valid identity element. Finally, verify the identity and decrypt. From the experimental results, the accuracy of the above three machine learning models is more than 75%. But the f1-score of the decision tree model is up to 93%, which is the highest, indicating that the decision tree model is largely suitable for dealing with the problem of behavior detection. In addition, the CP-ABE algorithm can determine whether the user has the right to access the information according to the user's identity effectively and quickly. The solution prevents and controls users with abnormal behavior and failed identity information verification effectively. It combines machine learning algorithm and algorithm based on attribute encryption successfully and makes certain contributions to the research of problems in this field.

Keywords: Machine Learning, CP-ABE, Behavior Detection, Authentication

1. Introduction

Identity authentication is applied to almost all systems and is an important means to achieve access control. Its main role is to restrict the user's access to certain information items or to restrict the use of certain control functions according to the user's identity and a defined group to which it belongs. [1] Compared with traditional access control, the main purpose of identity authentication is to prevent illegal users from entering the system. At present, some scholars have improved the identity authentication scheme. For example, Liu et al. proposed a threshold identity authentication scheme based on biometric identification [2], which combines BGN semi-homomorphic encryption algorithm and Shamir secret sharing to solve the problem of excessively high user rights in single identity authentication. In order to solve the problem of cross-chain identity authentication in cross-chain technology, Wang et al. proposed a cross-chain identity authentication scheme based on IBE of relay chain [3], which uses digital identity as the global identifier of cross-chain network to complete the identity authentication of cross-chain transactions.

Identity authentication can also be achieved through attribute encryption. Attribute-based encryption (ABE) [4] was proposed by Sahai et al. in 2005, whose purpose is to improve the fault-tolerant performance of identity encryption system based on biological information. Compared with the IBE system represents user's identity in a simple way, the attribute-based encryption system has more expressive power. It represents user's identity through the attribute set, which is composed of one or more attribute elements. In this way, it can extend the expression of user identity from unique identifier to a combination of multiple attribute elements. By embedding the access structure into the key or ciphertext, Key-Policy Attribute-Based Encryption (KP-ABE) [5] or

Ciphertext-Policy Attribute-Based Encryption (CP-ABE) [6] is formed. In KP-ABE, the attributes are embedded in the ciphertext, and only when the attribute set of the ciphertext satisfies the key policy can it be decrypted. In CP-ABE, the attributes are embedded in the key, and only when the user's attribute set satisfies the ciphertext policy can it be decrypted. It not only protects the ciphertext information, but also limits the user's decryption authority. In attribute encryption, Sahai et al. introduced the access control structure of Shamir secret sharing [4]. They build an access tree and set several threshold values on all non-leaf nodes, which can judge the input attribute elements. If the match is successful, the secret value can be obtained and the decryption will be continued until user get the secret value on root node.

Machine learning technology has made countless contributions to human beings nowadays. It has gradually penetrated into various industries, such as traditional prediction, image recognition, natural language processing and so on. [7] Through using of machine learning algorithms and training a large number of real data, a more accurate model is constructed so that it can be used to make appropriate decisions. Guo et al. proposed a method of app traffic identification under ShadowSocksR (SSR) proxy with machine learning [8]. They extract multiple features of traffic and building a machine learning model, which detects whether the traffic behavior of a software is generated by SSR agent. Chen et al. proposed a credit evaluation model based on dynamic machine learning [9], which solved the problem of category imbalance and high-dimensional features in data. It can also update the model by detecting users' credit behavior dynamically. At the same time, various encryption technologies resolve some security problems pointedly in view of the need of machine learning. For example, the idea of homomorphic encryption is to encrypt plaintext directly, the result of operation on ciphertext is equivalent to the result of re-encryption on plaintext by the same operation. [10]

Of course, attribute-based encryption schemes can also be combined with machine learning. Thanks to the advantages of attribute-based encryption, which is applicable to the situation that the user is not fixed in the distributed environment [11], machine learning and attribute-based encryption gradually permeate each other. Many scholars have proposed several encryption schemes according to the characteristics of machine learning algorithms. For example, Kurniawan et al. proposed a system to protect the machine learning engine in the Internet of Things. [12] It skillfully inserted CP-ABE and embedded the attribute-based encryption module. By cooperating with the model loader and the machine learning engine in a specific order, the supervised learning model can be integrated, which effectively resolves the security problem of model damage caused by malicious training. Similarly, focusing on the openness and heterogeneity of Edge Intelligence (EI), Zhou et al. proposed a ciphertext-policy attribute-based proxy re-encryption scheme based on CP-ABE and re-encryption technology [13], which enhanced the flexibility of accessing when sharing the EI model, whose advantage is that user can delegate the access rights to other people while ensuring the data security.

However, most of the above work focuses on strengthening machine learning algorithms, while there are still risks in identity authentication. Even under the access control based on attribute encryption, attackers can still impersonate legitimate users to pass identity authentication. Then an authentication based on attribute encryption with machine learning is proposed to focus on how to use machine learning algorithm to strengthen identity authentication.

The machine learning algorithms used in the solution include logical regression algorithm, KNN algorithm and decision tree algorithm. The CP-ABE algorithm is used as encryption algorithm. First, generate the data set of user behavior reasonably and randomly. Then, the data set is preprocessed so that the computer can recognize the data set. Taking the processed data set as the input of three machine learning algorithms, extract features and build machine learning models according to the characteristics of each algorithm. The advantages and disadvantages of the model is judged by comparing f1-score, and the model with the best comprehensive performance is selected for user behavior detection to screen out a number of users with normal behavior. Finally, use the CP-ABE algorithm for authentication, and take the user's identity element as the input of the algorithm. If the decryption is successful, the authentication will pass, otherwise the authentication will fail.

2. Basic Algorithm

2.1. K-Nearest-Neighbor Algorithm

K-Nearest Neighbor (KNN) [14] is a basic classification and regression algorithm, which belongs to supervised learning. The core idea of the algorithm is to select k samples with the closest distance from each sample in the test set by calculating the distance between each sample in the test set and all samples in the training set. The category with the largest proportion of the k samples is the category of the tested sample. The distance reflects the similarity between two samples. Euclidean distance is the most commonly used, which is recorded as ρ. Mark the vectors of two sample points as $(x_1, x_2, \dots, x_n)$ and $(y_1, y_2, \dots, y_n)$, then the Euclidean distance of the two samples is:

$$\rho = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

The value of k can be set arbitrarily, but the selection of k will have an impact on the model built by KNN algorithm. If k is too large, the prediction model will become simple, which is vulnerable to the impact of sample imbalance, resulting in under fitting. If k is too small, the prediction model will become complex, which is vulnerable to the influence of outliers, resulting in over fitting.

As shown in Figure 1, assuming that the red triangle sample belongs to abnormal behavior, the blue square sample belongs to normal behavior, and the green round sample is the sample to be tested. In the experiment, the two indicators of login time and the number of failed logins are selected as the eigenvector

of each sample. Then, calculate the Euclidean distance between the sample to be tested and all the samples in the training set. The category of the sample to be tested can be obtained by setting the k value. For example, when k=3, the sample to be tested will be classified as abnormal behavior. When k=5, the sample to be tested will be classified as normal behavior.

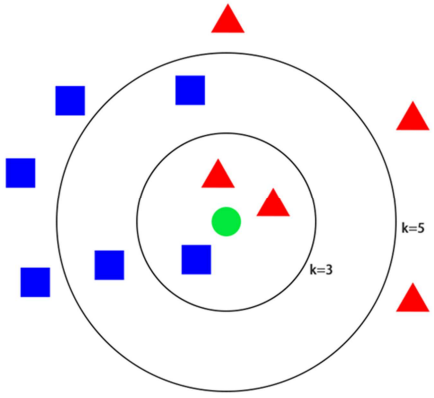

Figure 1. *Example of KNN.*

2.2. Logistic Regression Algorithm

Logistic regression algorithm [15] is a classification algorithm, which is often used to solve binary classification problems and belongs to supervised learning. Logistic regression is based on linear regression [16], and the linear regression model is defined as:

$$f(x) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b \tag{2}$$

b is a constant term, $w_1, w_2, \ldots, w_n$ is the weight, let $w = [b, w_1, w_2, \ldots, w_n]^T$, $x = [1, x_1, x_2, \ldots, x_n]^T$, then $f(x) = w^Tx$.

Next, the Sigmoid function [17] will be introduced, whose expression is:

$$g(x) = \frac{1}{1+e^{-x}} \tag{3}$$

Figure 2 below shows the sigmoid function.

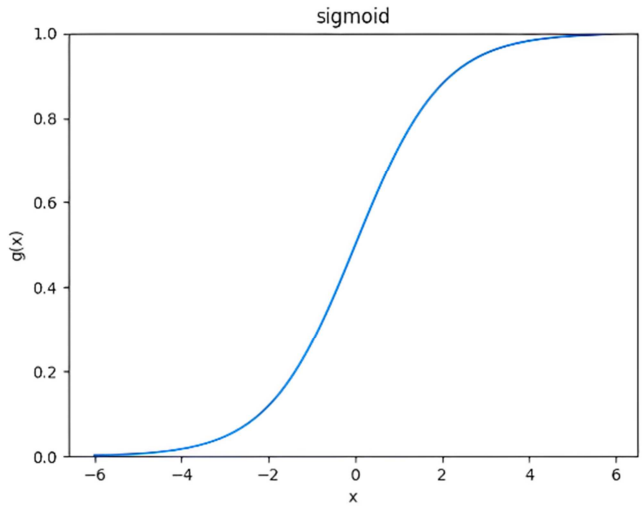

Figure 2. *Sigmoid Function.*

Assuming that the dependent variable $f(x)$ follows the Bernoulli distribution, use the sigmoid mapping function to process the output of the linear regression, and map $(-\infty, +\infty)$ to $(0,1)$. However, the results are not probability values in mathematical sense, so they cannot be directly used as probability values. The classification rule of the algorithm is to judge whether it belongs to the specified category by comparing the results of logistic regression and the set threshold (The default threshold is 0.5.).

The experiment is carried out in this way. Firstly, the data set of user login behavior is digitized so that the computer can recognize it. Then select the login time and the number of failed logins to conduct linear regression processing to obtain the parameters of the linear regression function. Then the output of linear regression is mapped to sigmoid function, and the positive and negative examples will be automatically determined when establishing the logical regression model. When forecasting, input the above two indicators to get the sigmoid function value. If the function value is greater than the threshold value, it is judged as a positive example, otherwise it is judged as a negative example.

In order to measure the difference between the predicted results of logistic regression and the real results, the loss can be calculated by using the log-likelihood function. The function expression is as follows:

$$cost(g(w^Tx), y) = \begin{cases} -log\big(g(w^Tx)\big) \ if \ y = 1 \\ -log\big(1 - g(w^Tx)\big) \ if \ y = 0 \end{cases} \tag{4}$$

y ∈ {0,1}, 0 and 1 represent the category of the sample.

It can be seen from the Figure 3 that when y=1, the greater the logistic regression result, the smaller the loss. When y=0, the smaller the result of logistic regression, the smaller the loss.

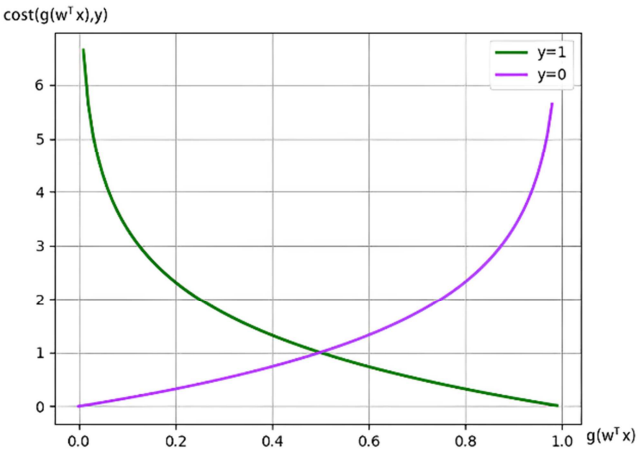

Figure 3. *Log-Likelihood Function.*

To get the total loss of the model, the logarithmic likelihood function can be expressed as follows:

$$cost(g(w^Tx), y) = \sum_{i=1}^{n}\{-y_i \log[g(w^Tx)] - (1 - y_i)\log[1 - g(w^Tx)]\} \tag{5}$$

In addition, the gradient descent algorithm [18] can also be

used to optimize the model which can adjust the weight parameters, so that the logistic regression results that should belong to the specified category are closer to 1, and those that do not belong to this category are closer to 0, so as to reduce losses.

2.3. Decision Tree Algorithm

Decision tree algorithm [19] is often used to deal with classification problems, which belongs to supervised learning. The model constructed by this algorithm is interpretable. The algorithm will divide each feature to generate visual classification rules, and the order of features determines whether the model constructed by the decision tree algorithm can be classified efficiently. The decision tree is divided according to three criteria, namely information gain, information gain rate and Gini coefficient, which generate ID3 decision tree, C4.5 decision tree and CART decision tree respectively. Since the principles of the three criteria are similar, information gain is used as an example to illustrate the working principle of the decision tree.

Entropy [20] is an index used to measure the uncertainty of events. Let X be a discrete random variable with finite values, and its probability distribution is:

$$P(X = x_i) = p_i, i = 1,2, \dots, n \tag{6}$$

then the entropy of random variable X is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{7}$$

Information gain is defined as the amount of information a feature provides for the entire classification model. The greater the information gain, the more important the corresponding feature is, and the more information it provides. The calculation method is the difference between the information entropy of the sample space and the conditional entropy of the sample space under given characteristic conditions. Record that the information entropy of data set D is $H(D)$, and the conditional entropy of data set D under the given conditions of feature A is $H(D|A)$, then the information gain of feature A on training data set D can be expressed as:

$$g(D, A) = H(D) - H(D|A) \tag{8}$$

The generation of a decision tree is a recursive process. The feature with the largest information gain value is selected as the root node of the decision tree. The training set is classified according to this feature. If the split subsets belong to one category, a leaf node can be branched from the root node. If the split subset is of high impurity, the feature with the largest information gain value is selected from the remaining features as the internal node of the decision tree, and so on until the training set is divided.

Figure 4 is an example. After calculation, it can be judged that the information gain value of login time is the largest. According to this feature, the training set is divided into two sets, one of which is abnormal behavior and its purity is 0. Therefore, the category of abnormal behavior is directly used as a leaf node. Another set has both normal behavior and abnormal behavior, so it still needs to be divided according to other features. Then select the feature of the number of failed logins to divide the set, and the impurity of the set obtained is 0, so the categories of these two sets are taken as leaf nodes. So far, the training set has been divided and the decision tree has been established. This model can be used to judge the category of samples in the test set.

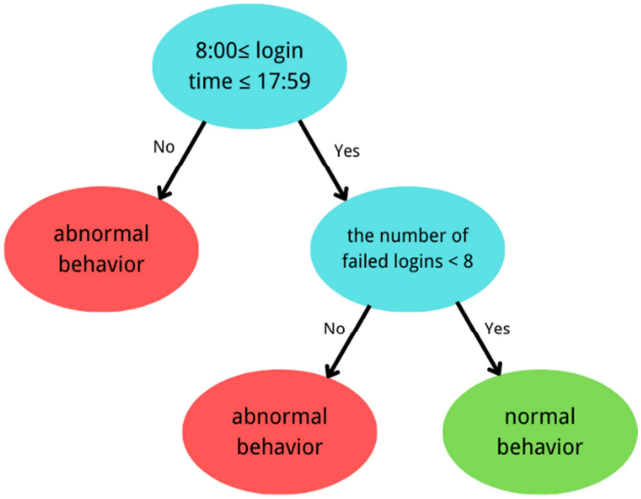

Figure 4. *Example of Decision Tree.*

2.4. Ciphertext Policy Attribute Based Encryption

The algorithm of ciphertext policy attribute-based encryption (CP-ABE) [6] mainly has four steps, namely setup, encryption, key generation and decrypt.

First, bilinear mapping is expressed as $e: G_0 * G_0 \rightarrow G_1$. In the setup, choosing a prime order bilinear group G_0, whose generator is g. Select two random elements α, β from Z_p, which is the additive cyclic group of the integer field. By calculating $h = g^{\beta}$ and $e(g,g)^{\alpha}$, the public key (PK) can be obtained. The encryption reflects the core of CP-ABE algorithm - embedding the access structure into the ciphertext, setting a secret value s at the root node, and getting the ciphertext $ct = Me(g,g)^{\alpha s}$ (M is plaintext.). Then, using the structural characteristics of the access tree, the secret values are dispersed through random polynomials until secret values are distributed to the leaf nodes. Calculating $C = h^s$ at the same time. The key generation part shows the characteristics of CP-ABE algorithm indirectly. It converts user attributes into secret key (SK). Meanwhile, it randomly selects an element r from the group Z_p and calculates $D = g^{(\alpha+r)/\beta}$. Decryption is a recursive process. After deduction, whether it is a leaf node or an intermediate node, the decryption result is $e(g,g)^{rq_x(0)}$. Therefore, inputting the ciphertext and key into the decryption algorithm. If the user's attributes meet the conditions, the expression of the secret value s of the root node is $A = e(g,g)^{rs}$. Finally, after the following calculations, the plaintext M can be expressed as follows:

$$\frac{ct}{e(C,D)/A} = \frac{Me(g,g)^{\alpha s}}{e(h^s, g^{(\alpha+r)/\beta})/e(g,g)^{rs}} = M \tag{9}$$

As shown in Figure 5, assuming that the secret value s used to encrypt plaintext M is 4, it will be placed at the root node of the access tree and generate a random polynomial $f(x) = 3x + 4$ (the highest degree of the random polynomial is one less than the threshold value.). By substituting the index values of the sub nodes into the polynomials, each sub node can obtain the secret value respectively, so that the secret value s can be dispersed. (As shown in Figure 5, the secret values scattered by the root node are 7, 10, and 13.) Then taking the newly generated secret value as the constant term of the random polynomial of the sub-node (As shown in Figure 5, $f(x) = 7$ and $f(x) = x^2 + 2x + 13$.), and continuing to follow the above operation until the leaf node obtains its own secret value. (The number code of the root node in Figure 5 represents the user attributes. For details, please refer to Table 1.

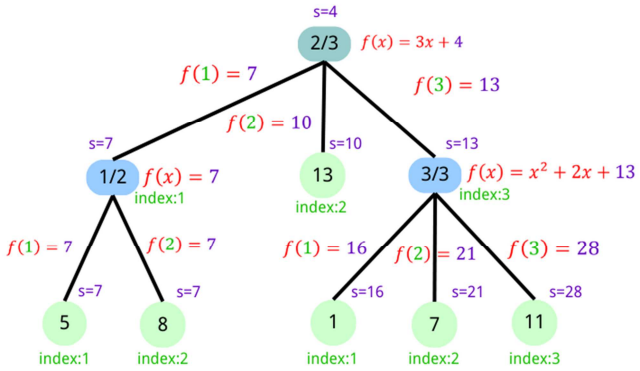

Figure 5. *Example of Access Tree.*

Table 1. *The Comparison Table of Attribute and Code.*

User Attribute	Attribute Code
Tsinghua University	1
Nankai University	2
Beijing University of Posts and Telecommunications	3
University of Chinese Academy of Sciences	4
School of Information Management	5
School of Marxism	6
School of Automation	7
School of Computing	8
School of Economics and Management	9
Professor	10
Graduate Student	11
Undergraduate	12
Dean	13
Monitor	14
Counselor	15

During decryption, if the user attribute conforms to the attribute stored in the leaf node, the corresponding secret value can be obtained. If the threshold value requirements of the parent node can be met, the secret value of the parent node can be calculated by using the obtained secret values through Lagrange interpolation, otherwise the decryption fails.

3. Experimental Results and Analysis

3.1. Machine Learning Model for ABE

First, a data set for user login is generated randomly, and the campus network is used as a simulation scenario. The administrator determines whether the login is abnormal manually according to the user's operation. The judgment methods for abnormal login include entering the password incorrectly for several times or logging in at unreasonable time intervals. In order to generate more reasonable data, the normal login period of the campus network is set from 8:00 to 18:00 in the program. The probability of each user logging in at this period is 80%, and the probability of logging in at other periods is 20%. The user enters the password for more than 8 times will be judged as abnormal login. In consideration of the administrator's operation error or special circumstances, the probability is set that the abnormal situation does not conform to the actual situation as 5% in the program. The data set generated according to the above conditions has a certain fault tolerance, which is more practical.

400 user login details are generated randomly. Next, python is used to implement three machine learning algorithms. First, the data set is divided into training set and test set. The training set is used to build the model, and the test set is used to verify the accuracy of the model. The indicators for partitioning the data set are the login time and the number of failed logins. Then, several functions in the scikit-learn package are used to implement standardization, generate predictors, calculate accuracy and f1-score. After standardization (the decision tree model does not need to be standardized.) and predictor generation, the login exceptions of the test set can be predicted. Then calculating the accuracy and f1-score. Finally, by comparing these two indicators, the advantages and disadvantages of the model can be judged. The following three tables show the test results of the three machine learning models:

Table 2. *The Test Result of Logical Regression Model.*

	Precision	Recall	F1-score
Abnormal Login	0.70	0.76	0.73
Normal Login	0.81	0.76	0.79
Accuracy		0.76	

Table 3. *The Test Result of KNN Model.*

	Precision	Recall	F1-score
Abnormal Login	0.95	0.83	0.89
Normal Login	0.89	0.97	0.93
Accuracy		0.91	

Table 4. *The Test Result of Decision Tree Model.*

	Precision	Recall	F1-score
Abnormal Login	0.95	0.88	0.91
Normal Login	0.92	0.97	0.94
Accuracy		0.93	

In the experimental results, the precision, recall, f1-score value of each category and the accuracy of the current model prediction results are mainly recorded. Next, let's analyze the experimental results of the following three machine learning models.

Because the accuracy rate and recall rate are a pair of contradictory variables, generally speaking, when precision is

high, recall is often low. When precision is low, recall is often high. In order to consider these two indicators comprehensively, f1-score is defined as the harmonic average of precision and recall, and its value range is from 0 to 1.

As the classification result only include "abnormal login" and "normal login", which is typical problems of binary classification, so the value of the accuracy and f1-score are main analysis objects.

From the test results of the three models, the accuracy is all more than 75%. It can be preliminarily determined that the characteristics of the three models are good. For the accuracy rate and recall rate of each category, it is easy to see that the gap between the KNN model is large, and the gap in the other two models is kept within 7%. As a whole, the precision and recall of the decision tree model are higher than those of the other two models, and it is difficult to compare the advantages and disadvantages of the other two models. However, by comparing f1-score indicators, it can be clearly seen that the comparison results of "abnormal login" and "normal login" of the three models are 0.91>0.89>0.73 and 0.94>0.93>0.79. To sum up, the comparison results of the model are that the decision tree model is better than the KNN model, and the KNN model is better than the logical regression model.

3.2. The Implementation for CP-ABE

According to the experiment of the machine learning model above, the decision tree model is selected to predict the user's login. If the predicted result is inconsistent with the real result, the data will be removed. If the predicted result is the same as the real result, only the data with the result of "normal login" will be retained. For the user who is judged as "normal login" by the administrator and machine learning model, it required to perform identity authentication next.

Java language is used to implement the CP-ABE algorithm. In the experiment, firstly, an access tree is built as shown in Figure 6 (The number code of the root node in Figure 6 represents the user attributes. For details, please refer to Table 1.), and then designed four functions: setup, key generation, encryption and decryption. The decryption part is used to simulate the user's operation of viewing plaintext. In order to conform to the actual situation, the secret value assigned in the preparatory stage needs to be cleared before the decryption operation, so that users can use their own attributes to decrypt.

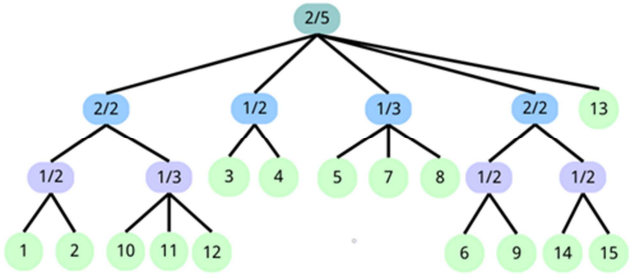

Figure 6. *The access tree of user attributes.*

Confronted with a series of user information that is not easy to operate in the program, the method of converting user information into number codes is adopted. The user's information is processed as the input of the program, and the decryption results of users are recorded. The experimental results are shown in the following table.

Table 5. *The Decryption of Users.*

User Name	Code of User Attributes	Results
Xiao Hong	4, 6, 14	success
Xiao Li	3, 5, 11	success
Xiao Wang	1, 8, 11	success
Zhang San	4, 6, 15	success
Wang Wu	4, 5, 15	success
Xiao Chen	2, 6, 13	failure
Li Zhang	4, 5, 14	success
Li Si	2, 9, 13	failure
Zhao Liu	3, 5, 15	success
Xiao Ming	3, 9, 12	failure

During the experiment, it can be seen whether the attributes of current user meet the threshold value of a node. However, because the information is lengthy, the experimental results are directly displayed here and the details are discarded.

By comparing the user attribute code with the access tree in Figure 6, it can be found if the user attribute finally reaches the threshold value of the root node, the decryption succeeds, otherwise the decryption fails.

4. Conclusion

This paper first introduces the research status of machine learning and attribute based encryption, then proposes the idea of combining machine learning with attribute encryption. Then three machine learning models based on the background of logging in to the campus network are set up. By comparing f1-score, it is concluded that the decision tree model has the best performance. On this basis, the CP-ABE algorithm is also implemented, which make use of user attributes as the key to obtain plaintext, so as to achieve the effect of authentication. After the experiment, it can be found that when the user attributes meet the access tree structure and reach the threshold value of the root node, the user can decrypt successfully, otherwise the decryption fails.

The solution proposed in this paper strengthens the security of network information from behavior detection and identity authentication, prevents and controls the abnormal behavior and access rights of users effectively.

The three machine learning models used in the experimental scheme are all suitable for solving the classification problem. The model with the best performance is selected, which prevents the problem that the single model leads to large experimental error.

The experimental scheme applies the machine learning algorithm to the encryption algorithm successfully and takes the output of the machine learning model as the input of the CP-ABE algorithm effectively. It achieves the connection of the interfaces between algorithms.

In the experiment, the data set is only processed by the machine learning model with the best performance, and then deleted the data whose prediction results are inconsistent with

the real results. In order to improve the security, the experiment also consider the intersection of the data with the correct prediction results of the three machine models. Since the processing results of the data set depend on the model with the worst performance (The f1-score of the logical regression model in this experiment does not exceed 80%), how to comprehensively improve the accuracy of the machine learning model is a key issue for us to study in the future.

Acknowledgements

This work was supported by the R&D Program of Beijing Municipal Education Commission (No. KM202311232014) and the Higher Education Research Project of Beijing Information Science and Technology University (No. 2021GJYB30).

References

[1] Wang S J. (2022). Application of User Identity Authentication Technology in Network Information Security. Information and Computer (Theoretical Edition) (08), 221-223.

[2] Liu Y, Guo S & Yang X Y. (2022). Threshold identity authentication scheme based on biometrics. Application Research of Computers (04), 1224-1227. doi: 10.19734/j.issn.1001-3695.2021.08.0388.

[3] Wang S S, Ma Z F, Liu J W & Luo S S. (2022). Research and Implementation of Cross-Chain Security Access and Identity Authentication Scheme of Blockchain. Information Network Security (06), 61-72.

[4] Sahai A, Waters B. Fuzzy Identity-Based Encryption [J]. International Conference on Theory & Applications of Cryptographic Techniques, 2005.

[5] Goyal V, Pandey O, Sahai A, et al. Attribute-Based Encryption for Fine-Grained Access Control of Encrypted Data [J]. ACM, 2006.

[6] Bethencourt J, Sahai A, Waters B. Ciphertext-Policy Attribute-Based Encryption [C] // IEEE Symposium on Security & Privacy. IEEE, 2007.

[7] Huang G S. (2022). Application of Machine Learning Algorithm in Artificial Intelligence. Integrated Circuits Design and Application (09), 192-193. doi: 10.19339/j.issn.1674-2583.2022.09.085.

[8] Guo G, Yang C, Chen M Z & Ma J F. App traffic identification under ShadowSocksR proxy with machine learning. Journal of Xidian University.

[9] Chen Y J, Gao H R & Ding Z J. (2023). Credit Evaluation Model Based on Dynamic Machine Learning. Computer Science (01), 59-68.

[10] Qin B D, Yu P H & Zheng D. (2022). Decision Tree Classification Model Based on Double Trapdoor Homomorphic Encryption. Information Network Security (07), 9-17.

[11] Yin F F. (2021). Research on Revocable and Searchable Attribute-Based Encryption Algorithms (Master Dissertation, Xidian University). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1022013684.nh

[12] Kurniawan A, Kyas M. Securing Machine Learning Engines in IoT Applications with Attribute-Based Encryption [C] // 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). 0.

[13] Zhou X, Xu K, Wang N, et al. A Secure and Privacy-Preserving Machine Learning Model Sharing Scheme for Edge-Enabled IoT [J]. IEEE Access, 2021, 9: 17256-17265.

[14] Wu X Y, Wang S H, Zhang Y D. Survey on theory and application of k-Nearest-Neighbors algorithm [J]. Computer Engineering and Applications, 2017.

[15] Dong X. Research on logistic regression and its parallel implementation on GPU [D]. Harbin Institute of Technology, 2016.

[16] Liu Y. Mathematical model of multiple linear regression [J]. Journal of Shenyang Institute of Engineering, 2005.

[17] Zhang X, Huang X, Zhong W H, et al. Implementation of Sigmoid Function and Its Derivative on FPGA [J]. Journal of Fujian Normal University (Natural Science Edition), 2011.

[18] Ruder S. An overview of gradient descent optimization algorithms [J]. 2016.

[19] Chong L U, Hui X U, Yang Y C. The research and application of classification algorithm based on decision tree [J]. Electronic Design Engineering, 2016.

[20] Chen L. (2013). Improvement and application of decision tree with Covariance information entropy information entropy (Master Dissertation, Yunnan University). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201401&filename=1013306255.nh