A novel Security Aware Sensitive Encrypted Storage approach to improve the encryption of big data

Gitanjali Gupta (gitagupta32@gmail.com)
Punjab College of Technical Education https://orcid.org/0000-0002-2551-4660
Kamlesh Lakhwani
Lovely Professional University

Research

Keywords: Intelligent cryptography, Cybersecurity, Encryption, Cloud computing, Big data

Posted Date: October 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-80029/v1

Abstract

The data security and privacy have become a critical issue that restricts many cloud applications. One of the major concerns about security and privacy is the fact that cloud operators have the opportunity to access sensitive data. This concern dramatically increases user anxieties and reduces the acceptability of cloud computing in many areas, such as the financial industry and government agencies. This paper focuses on this issue and proposes an intelligent approach to cryptography, which would make it impossible for cloud service operators to reach sensitive data directly. The suggested method divides the file with precision using an intelligent classification technique. An alternative approach is designed to determine whether data packets need splitting to shorten operating time and reduce storage space. Our experimental assessments of both safety and efficiency performance and experimental results show that our approach can effectively address major cloud hazards and that it requires an acceptable computing time using an intelligent machine learning classification technique. We have proposed a novel approach entitled as a model for Security Aware Sensitive Encrypted Storage (SA-SES). In this model, we used our proposed algorithms, including Convolution Neural Network with Logistic Regression (CNN-LR), Elliptic-curve Diffie–Hellman-Shifted Adaption Homomorphism Encryption (ECDH-SAHE) and Elliptic-curve Diffie–Hellman-Shifted Adaption Homomorphism Decryption (ECDH-SAHD) .

1. Introduction

Cloud computing is a very popular and successful technology in this period. Earlier, applications were built on the local server, but if the local network was blocked, the whole system and the application got failed immediately. Cloud Computing came into use to solve this problem and to store data digitally. So many famous companies like Google, Microsoft, Amazon, Facebook, and others have their clouds[7]. Because of minimal investment, low costs and so many different access services, many companies are shifting into the cloud. Cloud Computing offers services such as application services (e.g., SaaS), operator platform (e.g., PaaS), and operator-service network (e.g., IaaS).

Most companies that provide cloud services like Amazon(AWS), Dropbox, Google Drive, and One Drive for Microsoft offer different storage services packages and adjustable cloud storage spaces for customers. However, the security problem that cloud operations create remains a problem for the use of cloud services. Many cloud users care about their sensitive data accessed by cloud operators. Security risks create a lot of problems in cloud computing's direction of progress.

In cloud storage, users do not know the physical location of the data because they stored it on unknown servers where there is always a chance of user's private data getting leaked. This research shows a security architecture for cloud security. This system helps build a partnership between cloud service providers and consumers to effectively manage security. Big data is safe in cloud computing using the Smart Cryptography approach. In this method, the file is divided into pieces or packets and these packets are is distributed to the cloud servers and stored there. Another method is often used in this work, which is

to find out the data packet required to break to get less operating time. This method provides good security services with an effective time of computation[9].

1.2 Types of classification algorithms

Classification is a Supervised method used for learning, which is used in machine learning and statistics. Classification is done using the principle of learning based on the data input provided to it. It classifies the data based on bi-class and multiclass such as male or female gender classification, classifies the emails in the spam or not spam. In Machine Learning, there are various types of classification algorithms:

Naive Bayes Classifier (Generative Learning Model):

The classifier Naïve Bayes is based on the Bayesian probability theorem. It is a Supervised Learning Algorithm used for classification purposes. It solves the problem in attributes of both continuous and categorical nature.

- It is used mainly in word detection and spam filtering.
- This classifier was also used in recommendation-based systems.

Logistic Regression (Predictive Learning Model):

The logistic regression technique is used to assess the data set outcomes in which one or more independent variables are present. The outputs were only calculated for two possible outcomes.

Decision Trees:

In this approach, data is divided into sub-nodes into a tree-like structure and allows the model for regression and classification. The tree nodes are linked to each one and give the decision in the tree form. Tree subnodes are called leaf nodes.

Random Forest:

The logistic regression technique is used to assess the data set outcomes in which one or more independent variables are present. The outputs were only calculated for two possible outcomes.

Neural Network:

The neural network is based on the neuronal biological method. This network consists of the basic unit neurons, which are arranged in a layer and modify the input according to the decided amount and submit the output. This is used to find out the classification and patterns. Neural Network is capable of getting the relevant data from the complex data. It is very difficult for humans and also computer techniques to get information from complex data. Solving that problem is a solution.

Nearest Neighbor:

This algorithm is used based on similarity to store the current cases and to use those cases for potential classification. It is mostly used in statistical estimation and pattern recognition. It classifies the data by the nearest class of neighbors.

1.3 Encryption Techniques

Encryption is a technique that is used to safeguard the data. In encryption, meaningful data are turned into meaningless data that the normal person cannot understand. It is often found in military and other data centers where classified data is stored for data protection purposes. Below are the different algorithms that can be used in the data encryption process.

Triple DES

The more up-to-date, the improved version of DES is Triple DES or 3DES as it is written now and again, and its name suggests what it does. In three stages it runs DES three times on the information: scramble, unscramble, and then encode again. It doesn't give the efficiency of the cipher a triple increase.

RSA

RSA is an asymmetric cryptographic algorithm used to encrypt the message, without separately exchanging the private key. Uses the principle of large numbers factoring.

Blowfish

Blowfish has a 332–448 bits variable-length key and is a 64-bit square number. The two techniques consisting of the Blowfish algorithm are the introduction of the key and the scrambling stage of the details. In the first stage, a client variable key is consumed to 4168/8336-byte sub-key varieties, which is given a 4-byte component cluster size.

Two fish

This is a method of symmetric encryption in which two blowfish algorithms are combined for effective security. This algorithm matches the length of the keys up to 256 bits but only one key is used in the encryption process.

AES

Advanced Encryption Standard is an encryption algorithm that is based on an asymmetric key, and it works effectively on hardware and software. It supports 128bit, 192bit, and 256bit block capacity. It obeys the substitution-permutation network theory.

The rest of this paper follows the structure described below. Recent work related to this work is discussed and summarized in Sect. 2. We also provide a motivating illustration to illustrate the method of execution in Sect. 3. Also, the proposed model and the main principles used in the model are set out in Sect. 4. Section 5 then interprets the main algorithms with pseudo-codes and algorithm explanations. Besides, we

are testing our proposed model through expert demonstrations in Sect. 6. Finally, our conclusions are set out in Sect. 7.

2. Related Work

This section summarizes recent research achievements in the field of big data classification and cloud security issues that help our research history and theoretical foundation representation.

2.1 Review of Classification Techniques

Zardari et al. [1] *discussed that the data classification method depends on the confidentiality of the data. The KNN algorithm is used to identify the data as per the safety needs. It classifies the data into the sensitive and non-sensitive type that presents the data's need for protection. The RSA algorithm is used for encryption, and the CloudSim Simulator simulates this.*

Moghaddam et al. [2] *discussed that a variable data classification index must be used to ensure cloud data security and privacy. The index value is determined through the use of various parameters and the key parameters are confidentiality, honesty, and availability.*

Zardari et al. [3] *discussed the cloud computing issue of data protection and data recovery. Such problems are addressed by using the data and cloud model classification. The challenge is to solve it with data classification using the hybrid multi-cloud model. This model is worked on multiple clouds, grouping, and various cluster numbers.*

Tawalbeh L et al. [4] *presented a classification-based model that provides safe cloud computing. This model reduces the overhead and processing time included in the safety mechanism. For variable key sizes, it determines the protection at a different level. The proposed model is evaluated with different safety measures and produces positive results with high efficiency in the proposed work.*

Shaikh et al. [5] *proposed an approach of classification based on different parameters. The different dimensions are defined by those parameters. The security of the data can be given according to the level of protection needed. The proposed method solves the issue of privacy security and data leakage.*

Zardari et al. [6] *proposed that the K-nearest neighbor classifier suggested the confidentiality of the data in cloud-based data. The method is extended to the virtual cloud, and the data are categorized according to its security needs. KNN classifier classifies the data into two sensitive and non-sensitive classes. The data classification discusses which data would need to be more secure.*

Balachandran et al [7] *Discussed that the most challenging task in today's scenario is to choose the right Institute. Any student browsing through the social network sites for the reviews, ratings about the specific institution to get the approximate information about the particular institute. But the statistical dimension from the feedback is hard to examine. In this Aspect based Sentiment Analysis is applied directly to the comments that offer us negative and positive feedback of the institution in question. The different techniques are used to classify aspects such as NLP-based methodology, Machine Learning based (ML),*

unsupervised approach, Dictionary-based, Corpus-based approach. The NLP and the ML classifier give the best possible results analytics to classify each factor into their respective category.

Hagge Marvin et al. [8] *described that the micro-blogging service identified consumer sentiment i.e. This is Facebook. Twitter is a social media on which every person expresses their opinions. Consumer preferences are the basis for evaluating the view of customers of the individual product. In this user view, aspect-based sentiment analysis is done by part-of-speech tagging, and, for exchange for its excerpts, the optimistic, neutral, negative aspects of the tweets are parsed directly from natural language processing. The software toolkit in the proposed approach is designed to extract the tweets first, then filter to evaluate the polarity of the feelings and then show the result. In this people over the web platform will rent out their homes to each other. The aspects are Airbnb, place, time, house, day, people, night, view, apartment, space, for study. The results are shown by the table. The future will operate on Airbnb's website feedback.*

Pannala, NipunaUpeka, et al. [9] *Existing opinion mining work defined shall be performed at the word level, not at the sentence level. He includes the views articulated explicitly. The paper proposed is based on the qualified data set that analyzes and offers positive, favorable, and negative feedback for different products. The Aspect-based sentiment analysis (ABSA) operates on the various aspects of the object and reveals the polarity in returns. Techniques are used for applying ABSA machine learning (ML), and Natural language (NL). The dataset used in this proposed paper has annotations of the 1654 aspect category in the training dataset and annotations of the 845 aspect category in the test data set.*

KeumheeKang et al. [10] *Proposed a novel way to identify stressed mood users by monitoring their frequent tweets for a long time. They manipulate all forms of tweets on the internet, i.e. photos, emoticons, and texts. To assess the validity of the proposed method, two types of experiments were performed: 1) the proposed multimodal approach has been validated with several tweets and its output has been compared with SentiStrength; 2) it has been used to identify 45 mental states of users as depressive and non-depressive. The experimental results indicated that the proposed method of multimodal analysis has higher precision than existing methods, and it can more accurately predict the moods of individuals.*

Rongrong et al. [11] *Proposed approach to the approaches to visual sensation research. This is presented with a survey that describes the different techniques used for the study of visual sentiment. In this kind of research, photographs are used to assess the person's feelings. The survey concentrated largely on cutting-edge approaches that are used in the field of image analysis. This survey explains the current researcher framework since research is done mainly on the text, but the ontology of visual sentiment is a new concept for doing something else. The principle of deep learning is useful in the firm's successful visual sentiment analysis.*

P. D. Turney et al [12] *proposed a supervised learning algorithm that classifies the analysis as thumbs up and thumbs down. The mean semantic orientation is used to determine a review's classification. The positive and negative interaction with the review is indicative of the review's orientation. The semantic orientation is determined using PMI-IR which is this research's core step. The proposed algorithm provides*

different accuracy on different tweet forms including 74 percent on movies, 80 percent on banks and vehicles, and 84 percent on traveler reviews.

R. Socher, et al. [13] *worked on the prediction of label distribution through the sentence level using a new approach based on the estimation of the sentence level in the recursive autoencoder. The suggested work is done on the criteria for improving feeling and lexica. The dataset used in this study is personal user stories that have been annotated with several labels and aggregated from multinomial distribution capturing the emotional responses.*

A.-M. Popescu et al. [14] *proposed an unmonitored method of extraction of information used to derive the opinion from the comments. This work is done in the section below. Firstly, the product's characteristics are identified secondly, the product-related opinion is established, and thirdly, the opinion polarity. The final step of the proposed methods will be to rate opinions based on their power. Use the relaxation marking method specifies the semantic orientation. The tests of the suggested approach's accuracy and recall indicate the success in recognizing sentiments.*

M. Abdul-Mageed et al. [15] *worked on standard Arabic data for a study of sentiments. In this collection of work, data is collected, and then the automatic classification phase in which tokenization is performed on the data is started. The method of classifying the two stages is carried out on the data set. The outcomes of the proposed solution are evidence of the method's effectiveness.*

Donglin et al. [16] *worked on approaches to visual sense analysis. This is presented with a survey that describes the different techniques used for the study of visual sentiment. In this kind of research, photos are used to assess the person's feelings. The survey focused essentially on cutting-edge methods that are used in the process of image analysis.*

Gitanjali et al.[17] *discussed text classification is a basic approach to text mining and the processing of natural languages. In the previous usage, classifiers use human interface features such as frequency base and n-gram features that cannot find non-linearity in features and increase variance in features that directly impact classifier performance. The convolution-based approach refines the traditional features in the layered approach by an activation function. This method improves the effective learning pattern that is learned by logistic regression and is optimized through the boosting approach. The results showed that the proposed CNN-Logistic regression method significantly improves the accuracy due to improving feature pattern.*

2.2 Review of encryption techniques

Sehra et al. [18] *discussed that the role-based approach of access control is an efficient way of managing information access and reducing ambiguity in large network applications. It also helps to lower safety costs in large applications. In this RBAC work policy, on the cloud as migration policy is considered, which allows the user to migrate the database schema with effective security. Restriction policy helps limit the number of cloud-based transactions. The new backup and restore policy is being introduced to provide*

the data lost and restore policy helps to recover the data even if the local system crashes and the migration policy helps to transfer data from one cloud to another using XML.

Almorsy et al. [19] *proposed a cloud security management system based on the FISMA standard that enables security certification for customers and cloud providers. To control the protection it improves collaboration or cloud providers and service users. Using that method is applied. NET platform, and SaaS network security management.*

Yibin et al. [20] *presented a smart cryptographic approach that allows the cloud provider not to access partial data. This method separates the file into subfiles and stores certain files on cloud servers distributed. Another strategy is also proposed for determining when to split the data packets to reduce running time.*

Diwan et al. [21] *proposed various cryptographic algorithms that were compared and taken into account to ensure the confidentiality of the data. In these various cryptographic algorithms, different parameters such as block size, key length type, and characteristics are compared. He provided the idea of a different cryptographic algorithm that can be used to ensure data security in the cloud.*

Sood et al. [22] *A hybrid solution providing data security in cloud computing was suggested. In this job, various techniques are combined to provide successful protection from the sender to the ends of the receiver. Data security is given to the user based on confidentiality, honesty, and availability of the information. The safe socket layer provides data protection using the encryption method, and integrity is provided by Media Access Control. Using the login Id and password method to all users will enhance the protection.*

Sengupta et al. [23] *discussed a Cloud computing protection framework using cryptography. For this work, the cryptography is performed using the form of hybrid Ceaser cipher encryption. This offers security at the client, server, and network location for the cloud. That method provides hackers with effective security.*

Somani et al. [24] *proposed an RSA algorithm used to ensure confidentiality as part of protection while using Digital Marks to improve security by verifying it through Digital Signatures. The solution used five-stage carryout encryption. The key is generated in an initial step. In the second step, advanced labeling is carried out, and encryption and decoding in stage 3 and stage 4*

Rewagad et al. [25] *discussed the specification for maintaining the confidentiality of information placed in the cloud by manipulating the use of the computerized mark and Diffie Hellman for key exchange with the Advanced Encryption Standard encryption algorithm. Regardless of whether the transmission key is hacked, Diffie Hellman's key trade office makes it useless because the traveling key is of no use without the private key of the customer, which is only issued to the true blue customer. This proposed design of a three-way instrument makes it extreme for programmers to breach the security system, ensuring information is put in the cloud in this way.*

Prabhakar et al. [26] *proposed an information encryption protocol with the AES algorithm in mind. In cloud conditions, the AES approach covers the knowledge for the entire life cycle from start to finish. This encryption process uses an AES-256 encryption algorithm and a Secure Socket Layer to ensure information records in the cloud interchange. This method prevents data from being targeted by force and provides efficient protection for data in the cloud. It's not relying much on data protection and data effectiveness. The proposed approach ensures that knowledge is finished in all stages and is separated into two stages. In the first stage, information encryption is finished by AES – 256 encryption. In the second stage the client should be verified, the client sends the username and secret word to the cloud. At the point when the cloud gets the demand from the client at that point confirms the client's subtle elements, if the client is substantial at that point begin the procedure of information recovery.*

3. Motivational Example

An example of motivation in this section shows the important part of the suggested model, which is to secure data packets with sensitive information. The method consists of broken data packet and data packet retrievals. This situation is taking place in the financial sector, where sensitive information on cloud users' needs to be strongly secured. The data volumes have exploded and a huge amount of data has been generated in the last two years than in the human race's entire history. In big data, the major challenge is resources, because static and non-adjustable resources cannot support big data. Therefore, developing a suitable classification method is the key requirement.

In this research work, a unique classification strategy will be proposed to resolve the different problems with current techniques in use. The program suggested will use innovative machine learning methods and cryptography to handle the big data. Firstly, data will be categorized in sensitive and non-sensitive and classified without calculating the data non-linearity and dynamic information. The duration of encryption and the storage will be difficult because, after encryption, capacity still increases. So, the size of the data for encryption should be improved. The performance of classification will be increased with the use of the approximating function. This scheme will be capable to protect user data, as the main value is generated at random and no content information is contained in any split data. Attackers(hackers) are unable to get sensitive information even if they focus on details.

From an industry point of view, this scheme will be a very good method as the data protection system will be helping to categorize data to protect critical, sensitive, and classified information. If sensitive data is not managed properly, organizations will have to pay fine for breaking laws and regulations and may face financial loss or damage to reputation. From a society's point of view, the scheme of classifying big data will be cost-effective, as it will decrease the computation cost and will provide a more useful and effective method to information technology security.

4. Concepts And The Proposed Model

4.1 Graphical representation of the proposed methodology

4.2 Phases of the proposed methodology

The proposed methodology involves exploring various phases:

Phase 1

This phase of the proposed work will be focused on the model of secure data classification, which will further be based on the level of sensitivity of the data and graded according to this point. Precision, recall, and accuracy are evaluated in this process.

Phase 2

This process will only encrypt and save sensitive data in the cloud, and use the same server to store the non-sensitive data for efficient data use. In this step, the encryption Technique, calculation-time, and storage space will be analyzed.

Phase 3

This proposed research process would aim to achieve better results than current algorithms by using accuracy, time, and parameters as well as improving cloud data security and integrity as well as enhancing total execution time and reducing overall storage space.

5. Methods

In this section, we'll present descriptions of our proposed algorithms. Our proposed model is supported by three main algorithms, including the Convolution Neural Network with Logistic Regression (CNN-LR), Diffie–Hellman-Shifted Adaptation Homomorphism Encryption (ECDH-SAHE), and elliptic curve Diffie–Hellman-Shifted Adaptation Homomorphism Decryption (ECDH-SAHD). The sections below explain the detailed structure of the algorithms respectively.

- **Convolution Neural Network with Logistic Regression (CNN-LR)** *algorithm*

CNN-LR algorithm uses Convolution Neural Network with Logistic Regression to present the pseudo-code for the proposed method. Algorithm input is tweet text, first, it is pre-processed and then the function is extracted and the first step is to reduce non-linearity through the mechanism of convolution, pooling, and activation after the part of learning begins. Then use logistic regression. Then measure loss and accuracy in different numbers of EPOCH which improves the accuracy and reduces the loss iteratively.

In the following statement, we define the principal steps of Algorithm 1:

Algorithm 5.1

Convolution Neural Network with Logistic Regression (CNN-LR)

Input: feature vector with a class label

Output: Learning model for text classification

While (Number of Rows (i) > 0)

Start

While (Number of column (j) > 0)

Start

Perform Convolution X_i

$X_i = -y\text{-}a^{(nl)} .f' (z^{(n)})$.....(5)

X_i = Convolution of i Layer

y = features

$a^{(nl)}$ = n text features

$f' (z^{(n)})$ = transpose of features

Perform Polling and Sigmoid mapping

$X_i^{(l)} = ((W^{(l)})^T \delta^{(l+1)}). f' (z^{(l)})$.....(6)

$X_i^{(l)}$ = Sigmoid mapping of i layer l instances

$W^{(l)}$ = Weight of l instances

$\delta^{(l+1)}$ = Partial differentiation

$z^{(l)}$ = Bias Value

Compute features

$X_i . X_i^{(l)} = \delta^{(l+1)} . (a^{(l)})$.....(7)

Learn logistic refression

Learn $X_i . X_i^{(l)}$by loss function

$f_{LR}^{(W)} = \log (1 + e)$ ……..(8)

$f_{LR}^{(W)}$ = logistic function of w features

y_i= i^{th} Instances

X_i= X_i Text

w = weight of layer

Stop

In Algorithm 5.1 the pseudo-code of feature extraction uses both machine learning approaches and deep learning approaches. Two sections are used in the extraction of the function one is the frequency base feature and this method is called TF-IDF as the application termbase features. In another section, the features of n-grams are combined and the feature vector is obtained which is used in machine learning and deep learning for learning text.

- **Elliptic-curve Diffie–Hellman-Shifted Adaption Homomorphism Encryption (ECDH-SAHE)**

ECDH-SAHE Algorithm is designed to perform data processing before it is forwarded to the cloud side

Pseudo codes of the ECDH-SAHE algorithm is given in Algorithm 5.2.

• Algorithm 5.2

• Elliptic-curve Diffie–Hellman-Shifted Adaption Homomorphism Encryption (ECDH-SAHE)

Algorithm (Encryption)

Input: Text

Output: Encrypted Text and Key

1. 1. Start
2. 2. Elliptic curve [EC] (Text)
3. 3. Key⇓EC (Text)
4. 4. Client⇓DH(Key)
5. 5. Str⇓String (Text)
6. 6. Gen_bin⇓binary(str)
7. 7. s = 0, i = length (Gen_bin)
8. 8. while (i is not zero) do
9. 9. x = i mod 10
10. 10. y = x mod 10 and i/10
11. 11. s = x + y
12. 12. end while
13. 13. T_{text}= left_shift (str, s)
14. 14. Stop

- **Elliptic-curve Diffie–Hellman-Shifted Adaption Homomorphism Decryption (ECDH-SAHD)**

ECDH-SAHD Algorithm is designed to perform data processing before it is forwarded to the cloud side. This algorithm aligns with the Until Sent to Cloud Process process. Algorithm 3 Diffie – Hellman-Shifted Adaptation Homomomorphism Decryption (ECDH-SAHD) elliptic curve.

• Algorithm 5.3

• Elliptic-curve Diffie–Hellman-Shifted Adaption Homomorphism Decryption (ECDH-SAHD)

DECRYPTION ALGORITHM

Input: Key and Encrypted Text

Output: Original Text

1. 1. Start
2. 2. L = 32
3. 3. i=$|key/L|$
4. 4. s = 0
5. 5. while (i is not zero) do
6. 6. x = i mod 10
7. 7. y = x mod 10 and i/10
8. 8. s = x + y
9. 9. End while
10. 10. z = right_shift (T_{Text}, Sum)
11. 11. O_{Text}=$(z)^{c}$
12. 12. O_{Text}= decimal (O_{Text})
13. 13. Stop

The next section describes our tests and the outcomes of our studies.

6. Results

In this section, we described our experimental setup and partial experimental findings. The experimental design focused on the adoption of the proposed model in terms of the time of execution and storage space.

This segment shows a few experimental results from our success evaluations. First of all, there was a comparison of the accuracy of different machine learning techniques. We used input data of the same size and we analyzed the accuracy rate of different classification techniques. Figure 5 demonstrates a comparison of different encryption methods with execution time and storage space. We used the same size input data and analyzed different encryption time execution and storage space. Figs. Figs. Our

proposed scheme was 4 and 5 more effective than other techniques and Fig. 7 shows that ECDH-SAHE takes less time than other encryption techniques.

6.1 Reason for Selecting CNN-LG:

• As we go to KNN to Neural Network, in experimental analysis. Table 1 shows that the neural network enhances outcomes that enable the selection of the layered network of convolution.

• Features rely on linear structures in machine learning approaches and there is no nonlinearity.

• The classification by activation function is optimized in machine learning.

Table 1
comparison of different classifier accuracy

Algorithm	Accuracy
KNN	77.28
SVM	78.53
Hybrid(SVM-KNN)	68.65
neural network	77.51
CNN-logistic regression	80.66

6.2 Reason for Improving Performance of Proposed approach

It improves feature extraction behavior by layered convolution-based approach with the convoluted feature and maps it in an abstract way that reduces nonlinearity. Non-linearity comes through the CNN approach with the function Sigmoid and TANH which effectively increases the learning.

7. Discussions

The output of the suggested and current solution is shown in Fig. 3 and Fig. 4, based on different dataset sizes. The range of the experiment goes from 50 MB-512 MB. The resulted graphs show that the suggested technique highly improves encryption, decryption time, and safety. When we compare the performance of a small dataset and large dataset, we may find that increase in size only contains such overhead. Therefore, when the data size increases, storage and time don't increase as much. This experiment offers great benefits for a large dataset and the output of this solution is completely checked. The change in this solution is seen from the following reasons:

• This removes overhead time using binary stream rather than unary text stream.

• Improves storage by moving to the left during encryption and to the right during decryption.

• Shifting analysis generally depends on binary streaming and it depends on security-based improvement indirectly.

8. Conclusions

This paper is based on the topic of cloud data management and found a solution that doesn't allow cloud users to access private data from the customers. To meet this goal, we introduced a unique solution named Security Conscious Sensitive Encrypted Storage (SA-SES) model. In this model, we used our proposed algorithms, including Convolution Neural Network with Logistic Regression (CNN-LR), Elliptic-curve Diffie-Hellman-Shifted Adaption Homomorphism Encryption (ECDH-SAHE), and Elliptic-curve Diffie-Hellman-Shifted Adaption Homomorphism Decryption (ECDH-SAHD). Our tests have shown that our suggested scheme is capable to protect the main cloud-side problems. The paper is further focused on the topic of encryption and offered a summary of the turnaround time for data recovery, although the range of the data was limited. The time used for decryption was close to data encryption. Our suggested method provided a shorter turnaround time than the other strategies that are already in use. Future research must deal with the problem of data replication to improve the amount of data access, as data recovery will be failed due to any upgrade in the data center.

Declarations

- Availability of data and material: Dataset have been taken from twitter.com. Twitter API has been used to obtain the twitter dataset.
- Competing interests: The authors declare that they have no conflict of interest.
- Funding:- No Funding for this editorial support was taken by anyone except your journal and they have said 20% discount will be given to me.
- Authors' contributions: 1 lakh tweets have been taken and novel technique has been proposed that will firstly classify data into sensitive and non sensitive part with more accuracy and then sensitive data will be encrypted with less storage space and time. So this model could be used by industry or society also.
- Acknowledgements: It is a pleasure for me to thank all those who have helped me to accomplish this Ph.D. thesis. Firstly, I wish to express my deepest gratitude to Dr. Kamlesh Lakhwani for guiding me throughout this research work. My supervisors have been a continuous source of knowledge, inspiration, motivation and encouragement during the entire course of this research work.

References

1. Munwar ali zardari. Jung LT. Nordin Zakaria," K-NN Classifier for Data Confidentiality in Cloud Computing", IEEE, pp. 1–6, 2014.
2. Moghaddam FF, Yezdanpanah M, Khodadadi T, VDCI: Variable Data Classification Index to Ensure Data Protection in Cloud Computing Environments, IEEE Conference, Process and Control (ICSPC), pp. 53–57, 2014.
3. Jung MAliZLowT, Zakaria N. Hybrid Multi-cloud Data Security (HMCDS) Model and Data Classification IEEE Advanced Computer Science Applications and Technologies (ACSAT), pp. 166–171, 2013.
4. Lo'aiTawalbeh NS, Raad S. Al-Qassas, AlDosari F, "A Secure Cloud Computing Model based on Data Classification". In First International Workshop On Mobile Cloud Computing Systems, Management and Security (MCSMS-2015) Vol. 52, pp. 1153–1158, 2015.
5. Shaikh R, Sasikumar M. "Data Classification for achieving Security in cloud computing.". Procedia Computer Science. 2015;45(Elsevier):493–8.
6. Jung MAliZLowT, Zakaria N. K-NN Classifier for Data Confidentiality in Cloud Computing, IEEE Computer and Information Sciences (ICCOINS), pp. 1–6, 2014.
7. Balachandran L, Kirupananda A. "Online reviews evaluation system for higher education institution: An aspect based sentiment analysis tool," 2017 11th Int. Conf. Software, Knowledge, Inf. Manag. Appl., pp. 1–7, 2017.
8. Hagge M, Von Hoffen M, Betzing JH, Becker J, "Design and implementation of a toolkit for the aspect-based sentiment analysis of tweets," Proc. – 2017 IEEE 19th Conf. Bus. Informatics, CBI 2017, vol. 1, pp. 379–387, 2017.
9. Pannala NU. "Supervised Learning Based Approach to Aspect Based Sentiment Analysis," 2016.
10. Kang K, Yoon C, Kim EY, "Identifying depressive users in Twitter using multimodal analysis," 2016 Int. Conf. Big Data Smart Comput. BigComp 2016, pp. 231–238, 2016.
11. Ji R, Cao D, Zhou Y, Chen F. Survey of visual sentiment prediction for social media analysis. Front Comput Sci. 2016;10(4):602–11.
12. Turney PD, "Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews," Proc. 40th Annu. Meet. Assoc. Comput. Linguist., no. July, pp. 417–424, 2002.
13. Socher R, Pennington J, Huang EH, Ng AY, Manning CD, "Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions," EMNLP 2011 - Conf. Empir. Methods Nat. Lang. Process. Proc. Conf., no. ii, pp. 151–161, 2011.
14. Popescu AM, Etzioni O. "Extracting product features and opinions from reviews," Nat. Lang. Process. Text Min., no. October, pp. 9–28, 2007.
15. Abdul-Mageed M, Diab MT, Korayem M, "Subjectivity and Sentiment Analysis of Modern Standard Arabic," Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., vol. 27, no. 1, pp. 587–591, 2011.

16. Ji SL, Cao D, Ji R, Lin D. “Visual sentiment topic model based microblog image sentiment analysis,” Springer, vol. 75, no. 15, pp. 8955–8968.
17. Gitanjali, Lakhwani K. “A novel approach of sensitive data classification using convolution neural network and logistic regression,” Int J Innov Technol Explor Eng, 8, 8, 2883–6, 2019.
18. Gitanjali SS, Sehra. Jaiteg Singh. Article: Policy Specification in Role-based Access Control on Clouds. International Journal of Computer Applications 75(1):39–43, 2013.
19. Almorsy M, Grundy J, Ibrahim AS, “Collaboration- Based Cloud Computing Security Management Framework” IEEE conference of cloud computing, Washington (DC), pp. 364–371, 2011.
20. Li Y, Gai K, Qiu L, Qiu. Meikang & Zhao, Hui, Intelligent Cryptography Approach for Secure Distributed Big Data Storage in Cloud Computing. Information Sciences. 387. 10.1016/j.ins.2016.09.005, 2016.
21. Diwan S, Malhotra V, Jain S. R. Cloud security solutions: Comparison among various cryptographic algorithms. IJARCSSE, 2014.
22. Sood SK. A combined approaches to ensure data security in cloud computing, ACM. Journal of Network Computer Applications. 2012;35(6):1831–8.
23. Nandita Sengupta J, Holmes. Designing of Cryptography Based Security System for Cloud Computing, IEEE International Conference on Cloud & Ubiquitous Computing & Emerging Technologies pp. 52–57, 2013.
24. Somani U, Lakhani K, Mundra M. Implementing digital signature with RSA encryption algorithm to enhance the Data Security of cloud in Cloud Computing. InParallel Distributed and Grid Computing (PDGC), 1st International Conference. IEEE, 2010.
25. Rewagad P, Pawar Y. Use of digital signature with Diffie Hellman key exchange and AES encryption algorithm to enhance data security in cloud computing. communication Systems and Network Technologies (CSNT), International Conference(pp. 437–439). IEEE, 2013.
26. Prabhakar DM, Joseph KS, A new approach for providing data security and secure data transfer in cloud computing, International Journal of Computer Trends and Technology (IJCTT) pp 1202 – 120, 2013.
27. framework for secure. cloud computing environments." In *Cloud security: Concepts, methodologies, tools, and applications*, pp. 249–263. IGI Global, 2019.
28. Amoore L. "Cloud geographies: Computing, data, sovereignty. Prog Hum Geogr. 2018;42(1):4–24.
29. Mishra S, Kumar. BibhudattaSahoo, and PritiParamitaParida. "Load balancing in cloud computing: A big picture." *Journal of King Saud University-Computer and Information Sciences* (2018).
30. Stergiou C, Psannis KE, Kim B-G, Gupta B. Secure integration of IoT and cloud computing. Future Generation Computer Systems. 2018;78:964–75.
31. Mishra P, Pilli ES. Vijay Varadharajan, and UdayaTupakula. "Intrusion detection techniques in cloud environment: A survey. Journal of Network Computer Applications. 2017;77:18–47.
32. Mahboob T, Zahid M, Ahmad G. "Adopting information security techniques for cloud computing—a survey." In 2016 *1st International Conference on Information Technology, Information Systems and*

Electrical Engineering (ICITISEE), pp. 7–11. IEEE, 2016.

33. Meharia P. and Dharma Prakash Agarwal. "Securing the Human Cloud: Applying Biometrics to Wearable Technology." In *Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security*, pp. 303–316. IGI Global, 2016.
34. Sharma S, Gupta G, and P. R. Laxmi. "A survey on cloud security issues and techniques." *arXiv: preprintarXiv:1403.5627* (2014).
35. Grover J, Sharma M. "Cloud computing and its security issues—A review." In *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–5. IEEE, 2014.
36. Hashizume K, Rosado DG. Eduardo Fernández-Medina, and Eduardo B. Fernandez. "An analysis of security issues for cloud computing. Journal of internet services applications. 2013;4(1):5.
37. Behl A, and KanikaBehl. "An analysis of cloud computing security issues." In *2012 world congress on information and communication technologies*, pp. 109–114. IEEE, 2012.
38. Finance/legal Industry Slowest to Adopt Cloud. [Online]. Available at: https://www.unitrends.com/blog/finance-legal-industry-slowest-to-adopt-cloud [Accessed on: 23-08-2019].
39. Balasubramanian R, Aramuthan DrM. "Security problems and possible security approach in cloud computing.". Int J Sci Eng Res. 2012;3(6):1–4.
40. Multitenancy and physical security. [Online]. Available at:https://www.owasp.org/index.php/Cloud10_Multi_Tenancy_and_Physical_Security [Accessed on: 23-08-2019].
41. Various types of network attacks. [Online]. Available at:https://www.symantec.com/connect/articles/security-11-part-3-various-types-network-attacks [Accessed on: 24-08-2019].
42. Almulla S, Abdulrahman, Chan, YeobYeun. "Cloud computing security management." In *2010 Second International Conference on Engineering System Management and Applications*, pp. 1–7. IEEE, 2010.

Figures

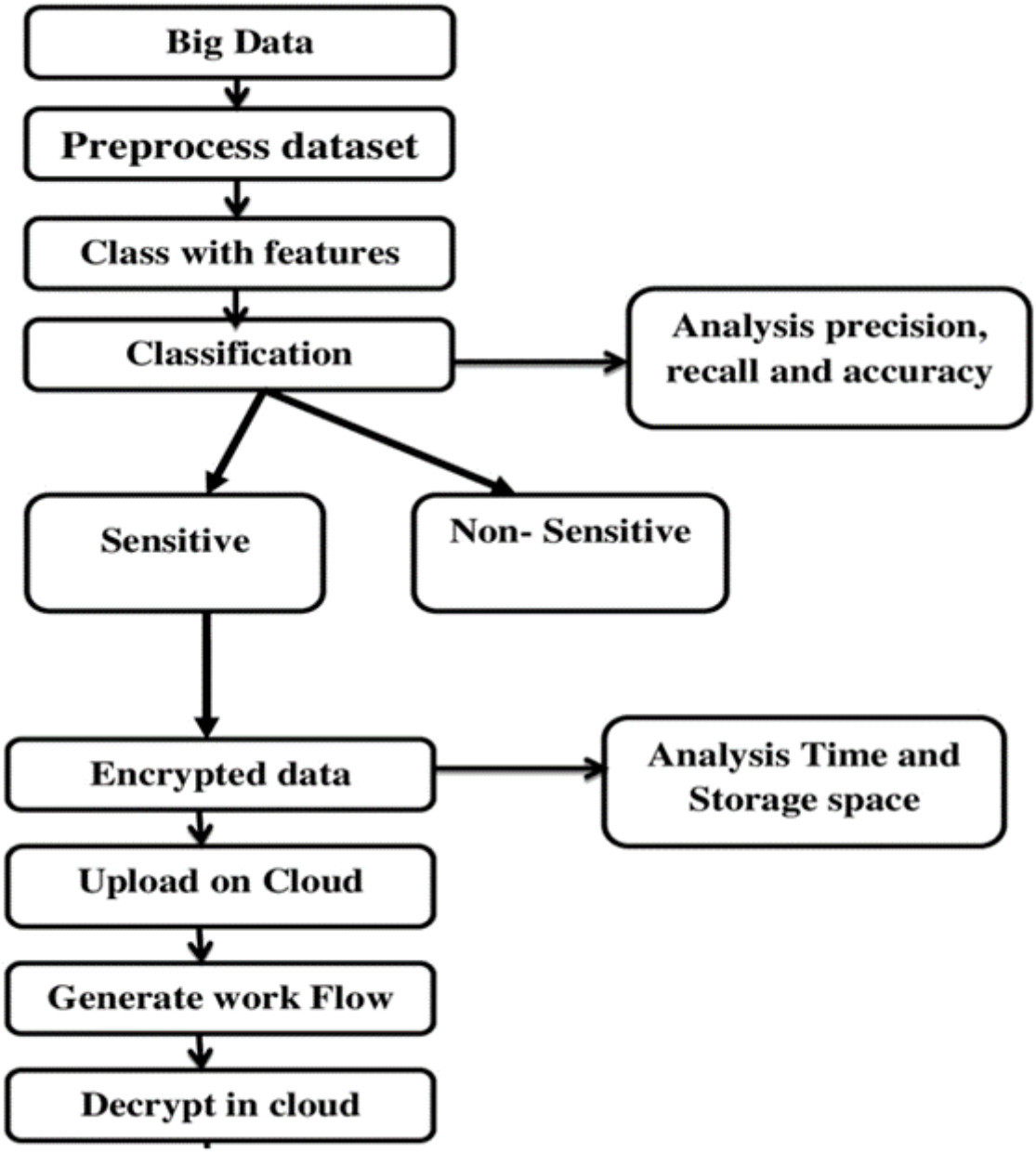

Figure 1

Proposed Methodology

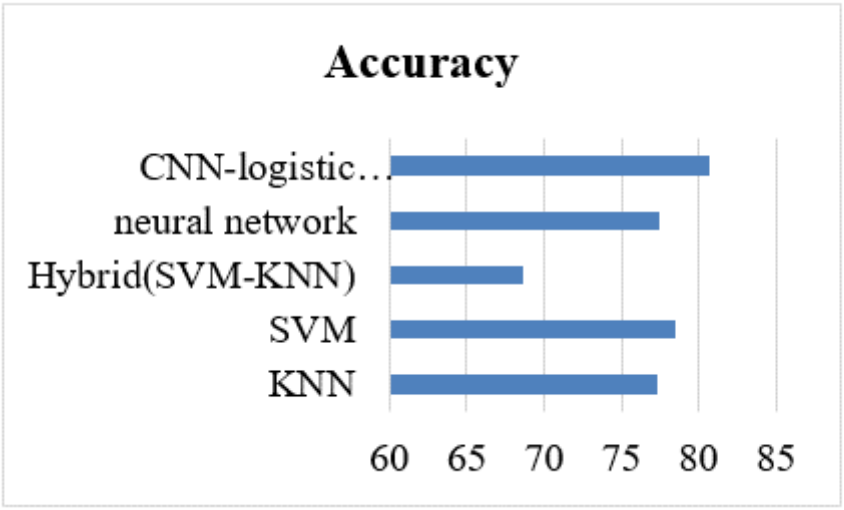

Figure 2

Comparison of different classifier accuracy

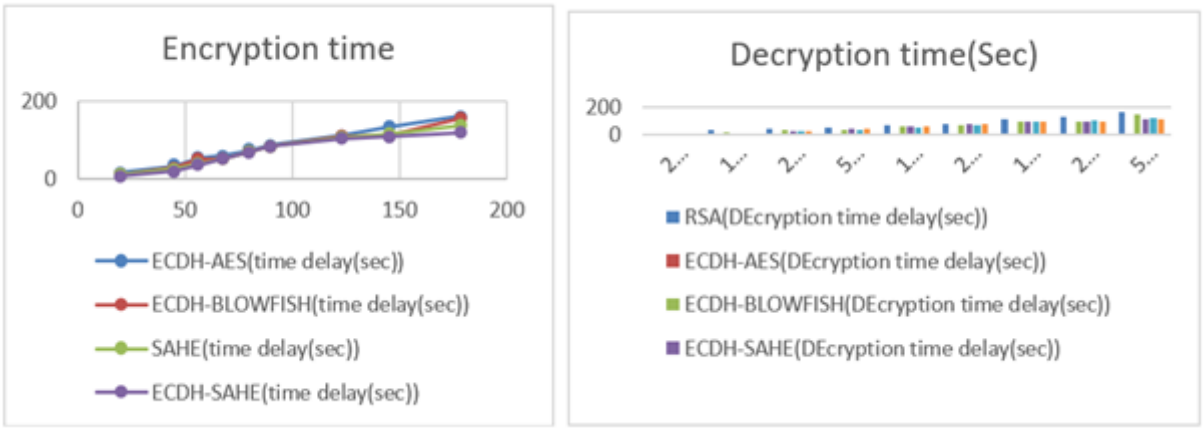

Figure 3

Comparison of proposed and existing approach encryption & decryption time based on different size dataset

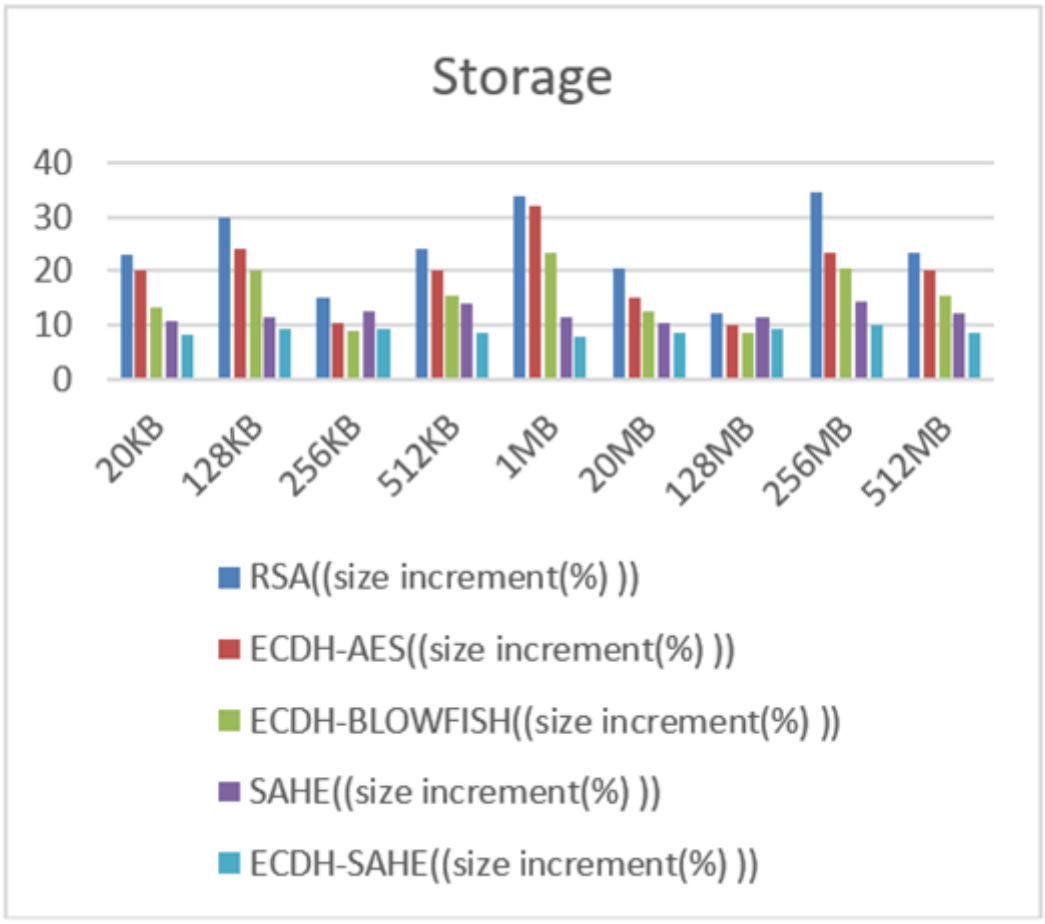

Figure 4

Comparison of proposed and existing approach storage on different size dataset