IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-15, May 2019 E-ISSN: 2347-2693

Advanced Encryption Standard Strategy for Big Data in Cloud

Praveen G[1*], Pratheek R[2], Rahul Mogar Y [3], Patil G B[4], Prasanna G[5]

[1,2,3,4,5]Department of Computer Science, East West Institute of Technology, Bengaluru, India

DOI: https://doi.org/10.26438/ijcse/v7si15.99104 | Available online at: www.ijcseonline.org

***Abstract*—**In the era of information age, due to different electronic, information & communication technology devices and process like sensors, cloud, individual archives, social networks, internet activities and enterprise data are growing exponentially .The most challenging issues are how to effectively manage these large and different type of data .Big data is one of the term named for this large and different type of data .Due to its extraordinary scale, privacy and security is one of the critical challenge of big data.Many current applications abandon data encryptions in order to reach an adoptive performance level companioning with privacy concerns. In this paper, we concentrate on privacy and propose a novel data encryption approach, which is called Dynamic Data Encryption Strategy (D2ES). Our proposed approach aims to selectively encrypt data and use privacy classification methods under timing constraints. This approach is designed to maximize the privacy protection scope by using a selective encryption strategy within the required execution time requirements.

Keywords-Privacy-preserving,data-encryption-strategy,BigData,mobile-cloudcomputing.

I. INTRODUCTION

Big Data is term for any collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges for the big data include capture, storage, search, sharing, transfer and security analysis and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data. Introduction to mobile cloud computing techniques has empowered numerous applications in people's life in recent years..Moreover, as an emerging technology, cloud computing has spread into countless fields so that many new service deployments are introduced to the public such as mobile parallel computing and distributed scalable data storage. Penetrations of big data techniques have further enriched the channels of gaining information from the large volume of mobile apps' data across various platforms, domains, and systems. Being one of technical mainstreams has enabled big data to be widely applied in multiple industrial domains as well as explored in recent researches.

Despite many benefits of using mobile cloud computing, there are great concerns in protecting data owners' privacy during the communications on social networks or mobile apps.One of the privacy concerns is caused by unencrypted data transmissions due to the large volume of dataavailable.

Storing and processing big volumes of data requires scalability, fault tolerance and availability [9]. The traditional infrastructure of storing and managing data is now proving to be slower and not easy to manage. Cloud computing delivers all the essential requirements for storing big data through hardware virtualization. Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant.

However, in the high speed connectivity era, moving large datasets on cloud and providing the details needed to access it, is a current issue. Because, these large sets of data often carry sensitive information like credit or debit card numbers, addresses, medical records and other details, raising data privacy concerns. In order to ensure data privacy, the data can be encrypted at the client side before outsourcing the data to the cloud server.

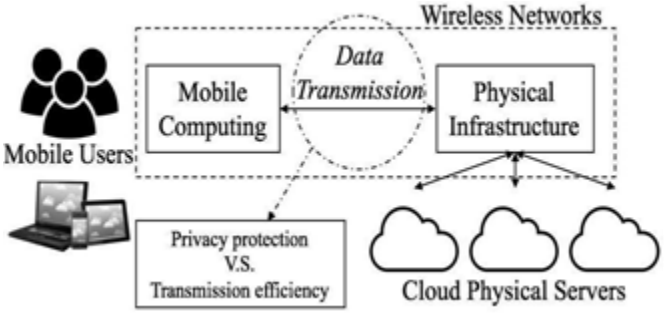

Fig. 1.1: High level architecture of mobile cloud computing illus-trating the balance between privacy protection and transmission efficiency.

The crucial issue is that most contemporary wireless transmissions carry plain-texts due to the workload volume and real-time service concerns. The implementation of big data further stops transmission from carrying cipher-texts. The target protection location is represented by the broken-line box in the figure, which depicts that the data transmissions between physical infrastructure and mobile computing in mobile cloud need to be protected.

In this paper we propose a novel approach that selectively encrypts data in order to maximize the volume of encrypted data under the required timing constraints. The proposed model is called Dynamic Data Encryption Strategy (D2ES) model, which is designed to protect data owners' privacy at the highest level.The basis for this method is a series of patents filed in [4] [5] and [6].The rest of this paper is organized as follows. Section II discusses the existing technologies. Section III presents the proposed methodology Section IV provides the system architecture for proposed method .Lastly, in Section V, we conclude this study along with all the references.

II. RELATED WORK

The privacy preserving of Big data in cloud can be carried out in a number of ways. Over the years, multiple variants of techniques have been identified and targeted with strategic solutions. Some of the methods are discussed below.

k-anonymity

The k-anonymity model was developed to achieve effective data privacy preservation.In k-anonymity techniques such as generalization and suppression are used to reduce the granularity of data representation. Any given k record maps onto at least k-other records in the data to reduce this granularity sufficiently.The term k-anonymity impliesthat with respect to a set of quasi identifiers attribute each record within anonymitized table must be indistinguishable with at least (k-l) other record within dataset i.e. the identifiers of each sanitized dataset is same as those of at least (k+l).Quasi identifier is defined as a set of attribute that can be used toidentify an information provider with a significant probability of accuracy. Quasi identifiers cannot be distinguished from others if each dataset linked to at least k-information providers.In order to reduce granularity generalization used togeneralize attribute value to range e.g. the data set of birthcan be generalized to range such as year of birth so risk ofidentifiers is reduced. In the suppression method, the valuesof the attribute is removed completely by replacing thosevalues by default one such that rare attributevalues merge into same group assigned the default value.

Perturbation Approach

Perturbation approaches protects privacy of data by distorting information of original dataset.Different data perturbation techniques are available for modifying dataset such that they are different from original.Data derived from perturbed dataset can be use to perform data. mining as perturbed datasets still retain features of the originals. There are two common approaches-for perturbing data that is n01se adding and random substitution.

Noise Addition

Noise Addition uses a random number or noise which is use for adding into the numerical attribute forcreating perturbed dataset: Random number ls generally drawn from. a small deViation and normal distribution With Q-mean. Thls strategy usually used for numeric values hence has only a little privacy.

Random Substitution

Random substitution replaces sample by randomly replacing values of attributes. This technique uses invertible matrix M of size n*n called the perturbation matrix where n is the umber of possible values of an attribute that is being perturbated.Perturbation based approaches do not make strict tradeoffs between preservation of data sample utility and privacy. This mechanism preserves privacy of each sample independently rather than depending on attribute values of every sample. The result illustrated by random substitution in terms of both privacy and data utility preservation can be equally random .

Data Set Complementation Approach

Dataset Complementation approach was designed for discrete value classification so continuous values are replaced with ranged values. The entire original dataset is replaced by unreal dataset for preserving the privacy via dataset complementation. This approach can be applied at any time during the data collection process so that privacy protection can be in effect even while samples are still being collected The original accuracy of training dataset is preserved Without linking the perturbed dataset to the information provider i.e accurate data mining result yield while preserving privacy of individual's records by dataset complementation approach.

III. METHODOLOGY

Our proposed system demonstrates how a dataset can be encypted for privacy preservation The proposed system creates a platform hosted on a cloud service.User registration need to be done and the user can login based upon the id an password credentials..The user is greeted with home page and are provided with options to view profile,for training data and to upload file and to download the file. The user

first selects for training dataset for sensitive and non-sensitivity.Then the user selects the Upload File option to upload a file which is in text format. In case the file is already uploaded then a prompt "file already exists" will appear..

The whole scenario of the system is showning Figure 3.1.The admin selects the training data after which the privacy preserving data encryption strategies are done.The classification and clustering process is carried on after which the user by registration and login can download the file.Every file thta is uploaded and encrypted can be downloaded by the user.

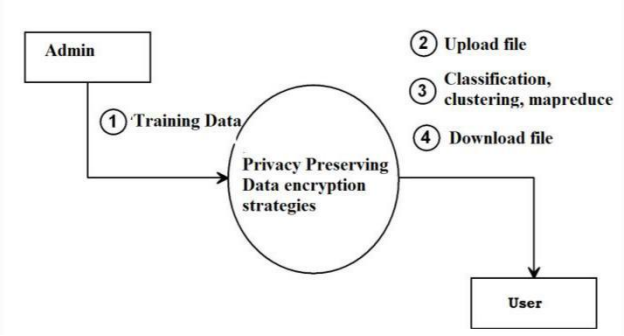

Figure 3.1: Admin actions

IV. SYSTEM ARCHITECTURE

Userfirst login to the network cloud by registering with fresh registration or login credentials. The user then can upload datasets by clicking Upload File domain. In that he selects from the category of Sensitive Data and Non-Sensitive Data.The selection of sensitive and non-sensitive data depends on the user,If the user wants to upload a sensitive data then sensitive text is selected from the category or else non-sensitive.The sensitive and non-sensitive dataset selection is based on keywords.

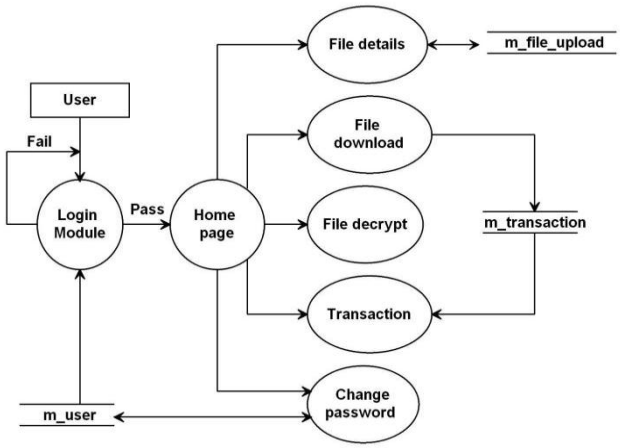

Figure 4.1: An Overview Architecture for user actions

The system takes the uploaded dataset from the admin and extracts the keywords from the data package.Afterthe keyword extraction these keywords are classified into two clusters. The two clusters of dataset will then get into the blocking process. After the blocking process the Map reduce process takes place.The datasets are then encrypted and will be uploaded into the cloud. System architecture of this is as shown in Figure 4.2. The admin can login by giving specific credentials**.**The admin can upload to training datasets by sensitive and non-sensitive based upon the keywords.The system architecture for the user module is shown in Figure 4.1

Figure 4.2: An Overview Architecture for encryption process

V. ALGORITHM

The Dynamic Data Encryption Strategy (D2ES) method is a combination of different encryption algorithms.These algorithms are used to check number of repeatationof words,for clustering,to remove sensitive words,to calculate sensitive weight and to generate hash block.

i. LBA(Linear Base Algorithm)
This Algorithm is used to collect the number of bases or total repeated words.

ii. K-means Clustering Algorithm Used for Cluster formation.

iii. Stop Word Removal Algorithm
Used to check whether the word is sensitive or not(ex. the is on etc).

iv. Weight Frequency Algorithm
Calculates the sensitive weight of the dataset.

v. MD5 Algorithm
Generates hash for the block.

Algorithm: K-means Clustering

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centers.

1: Randomlyselect 'c' cluster centers.
2: Calculate the distance between each data point and cluster centers.
3: Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4: Recalculate the new cluster center using $v_i = (1/c_i) \sum$
5: Recalculate the distance between each data point and new obtained cluster centers.
6: If no data point was reassigned then stop, otherwise repeat from step 3.

Algoirthm: Stop word removal

1: The target document text is tokenized and individual words are stored in array.
2: A single stop word is read from stopword list.
3:The stop word is compared to target text in form of array using sequential search technique.
4: If it matches , the word in array is removed , and the comparison is continued till length of array.
5: After removal of stopword completely, another stopword is read from stopword list and again algorithm follows step2.

6: Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

VI. Result and Disscusion

The D2ES model is way more effective in privacy preserving of a dataset in cloud because of the combination of algorithms used in the model.The five different algorithms used has differn mechanism to provide the privacy for a dataset.The main advantage in this model is that it categorises the dataset as sensitive and non-sensitive before applying the privacy preserving strategies.As the dataset is categorised at the beginning itself the the dataset which doesnot require privacy is discarded and so the time comsumption is avoided.

VII. Conclusion and Future Works

This paper focused on the privacy issues of big data and considered the practical implementations in cloud computing. The proposed approach, D2ES, was designed to maximize the efficiency of privacy protections.The Main support to the system is the D2ES model which is the combonation of different privacy preserving algorithms that was developed to dynamically alternative data packages for encryptions under different timing constraints.

D2ES maximizes the efficiency of privacy protections in cloud computing environment and provides user safe and protected services. Higher efficiency in encryptions for dynamic data packages under different timing constraints i.e Encryption of dynamic data packages are done irrespective of time period.Proposed approach has an adaptive and superior performance than the existing system.Taking into consideration ofthe privacy issues of big data and considered the practical implementations in cloud computing. The proposed approach, D2ES, was designed to maximize the efficiency of privacy protections

Acknowledgment

We would like to thank our principal Dr. K Chennakeshavalu and head of the department Dr. Arun Biradar, East West Institute of Technology , Computer Science and Engineering for supporting us to carry out our project with clear guidelines from Assistant Professor Mr Prasanna G.

References

[1] S. Yu, W. Zhou, S. Guo, and M. Guo. A feasible IP traceback framework through dynamic deterministic packet marking. IEEE Transactions on Computers, 2018.
[2] S. Yu, G. Gu, A. Barnawi, S. Guo, and I. Stojmenovic.Malware propagation in large-scale networks. IEEE Transactions on Knowl-edge and Data Engineering, 2018.

[3] S. Liu, Q. Qu, L. Chen, and L. Ni. SMC: A practical schema for privacy-preserved data sharing over distributed data streams. IEEE Transactions on Big Data, 2017.
[4] S. Rho, A. Vasilakos, and W. Chen. Cyber physical systems technolo-gies and applications. Future Generation Computer Systems, 2017.
[5] L.Wu, K. Wu, A. Sim, M. Churchill, J. Choi, A.Stathopoulos, C. Chang, and S. Klasky. Towards real-time detection and tracking of spatio-temporal features: Blob-filaments in fusion plasma. IEEE Transactions on Big Data,2017.
[6]S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Basar. Dependable demand response management in the smart grid: A stackelberg game approach. IEEE Transactions on Smart Grid, 2017.
[7] M. Qiu, M. Zhong, J. Li, K. Gai, and Z. Zong. Phase- change memory optimization for green cloud with genetic algorithm. IEEE Transactions on Computers, 2017.
[8] H. Liu, H. Ning, Y. Zhang, Q. Xiong, and L. Yang. Role- dependent privacy preservation for secure V2G networks in the smart grid. IEEE Transactions on Information Forensics and Security, 2016.
[9] F. Tao, Y. Cheng, D. Xu, L. Zhang, and B. Li. CCIoT- CMfg: cloud computing and internet of things-based cloud manufacturing service system. IEEE Transactions on Industrial Informatics, 2016.
[10]G. Wu, H. Zhang, M. Qiu, Z. Ming, J. Li, and X. Qin. A decentralized approach for mining event correlations in distributed system monitor-ing. Journal of parallel and Distributed Computing, 2016.
[11]S. Yu, W. Zhou, R. Doss, and W. Jia. Traceback of DDoS attacks using entropy variations. IEEE Transactions on Parallel and Distributed Systems, 2016..
[12]Y. Li, W. Dai, Z. Ming, and M. Qiu. Privacy protection for preventing data over-collection in smart city. IEEE Transactions on Computers, 2016.
[13]Yu, M. Au, G. Ateniese, X. Huang, W. Susilo, Y. Dai, and Min. Identity-based remote data integrity checking with perfect data privacy preserving for cloud storage. IEEE Transactions on Information Forensics and Security,2015.
[14]L. Weng, L. Amsaleg, A. Morton, and S. Marchand-Maillet. A privacy-preserving framework for large-scale content-based infor-mation retrieval. IEEE Transactions on Information Forensics and Security,2015.
[15]K. Gai, M. Qiu, H. Zhao, and J. Xiong. Privacy-aware adaptive data encryption strategy of big data in cloud computing. In The 2nd IEEE International Conference of Scalable and Smart Cloud, 2014.
[16]Y. Zhang, C. Xu, S. Yu, H. Li, and X. Zhang. SCLPV: Se-cure certificateless public verification for cloud-based cyber-physical-social systems against malicious auditors.IEEE Transactions on Computational Social Systems,2014.
[17] C. Wang, S. Chow, Q. Wang, K. Ren, and W. Lou.Privacy-preserving public auditing for secure cloud storage.IEEE Transactions on Computers,2013.
[18] K. Gai, L. Qiu, M. Chen, H. Zhao, and M. Qiu. SA-EAST: security-aware efficient data transmission for ITS in mobile heterogeneous cloud computing. ACM Transactions on Embedded Computing System, 2012.
[19] C. Lai, M. Chen, M. Qiu, A. Vasilakos, and J. Park. A RF4CE-based remote controller with interactive graphical user interface applied to home automation system. ACM Transactions on Embedded Computing Systems,2012.
[20] S. Backhaus, R. Bent, J. Bono, R. Lee, B. Tracey, D. Wolpert, D. Xie, and Y. Yildiz. Cyber-physical security: A game theory model of humans interacting over control systems. IEEE Transactions on Smart Grid, 2011.
[21]K. Gai, M. Qiu, H. Zhao, and W. Dai. Privacy- preserving adaptive multi-channel communications timing constraints. In The IEEE International Conference on Smart Cloud 2016, page 1, New York, USA, 2010.
[22]L. Tang, X. Yu, Q. Gu, J. Han, G. Jiang, A. Leung, and T. Porta. A framework of mining trajectories untrustworthy data in cyber-physical system. ACM Transactions on Knowledge Discovery from Data,2010.
[23]F. Schuster, M. Costa, C. Fournet, C. Gkantsidis, M.Peinado,Mainar-Ruiz,andM.Russinovich.VC3: Trustworthy data ana-lytics in the cloud using SGX. In IEEE Symposium on Security and Privacy, 2010.

Authors Profile

Mr. Pratheek R is pursuing his 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. His area of interest includes Artificial intellegence and Big Data.

Mr. Patil G B is pursuing his 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. His area of interest includes Big Data and cloud computing.

Mr. Praveen G is pursuing his 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. His area of interest includes Big Data and Machine Learning.

Mr. Rahul Mogar Y is pursuing his 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. His area of interest includes Machine Learning and Image Processing.

Mr. Prasanna G received the B.E degree in Computer Science and Engineering from JVIT, Bengaluru, VTU and received M.Tech degree in Computer Science from EWIT, Bengaluru, India. He is currently working as Assistant Professor in East West Institute Of Technology.

