International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Segregation of Sensitive Data in Cloud Storage

R. Sivagami, A. Nagarajan

Abstract: ***Big data is a huge collection of data, which are larger in size. It assembles many techniques and technologies to uncover the needed values from a larger data set. Big data needs a large server to store the data which is higher in cost and also there is a need for maintenance. Cloud server can be a key for this problem. It has the capability of large scale storage management. But it is a third party service, so the apprehension here is the data security. Data can be secured from the cloud server by strong encryption methodologies. All data doesn't need a high data security, so first we need to classify the data into sensitive and insensitive data. Sensitive data alone needs a proper attention over threats. This paper focuses on the identification of sensitive data within an acceptable computation time.***

Key Terms: ***Big data, Cloud server, Data security, Sensitive data, Third party services.***

I. INTRODUCTION

Cloud Computing is a new generation technology which enables its users to access various available resources which are shared remotely in cloud servers using internet based web services. Various familiar companies and organizations are stirring into cloud based servers due to the storage and access from anywhere with any device. Also the cloud server has the centralized data storage centers and it has the ability to store a huge amount of data. Sensitive data is an information or group of information's. All the web applications are requesting various data from its users to utilize their services [3]. The web user's data can be classified into sensitive data and general data. By comparing the both types the sensitive data should be handled safely than the other one.

The sensitive information's are should be protected against the hackers and unauthorized access. The present web servers and web applications are offering various security approaches to handle sensitive data as more secretly. The familiar securing approach which is followed in today's web applications is authorizing the access. The web mechanisms are implementing access roles based data accessing approaches [4]. But these approaches are not efficient and the data can be tracked from the backdoor such as database DBA roles. The web services are needed to protect the sensitive data from all type of threats and also for ethical needs, ensuring personal privacy, data regularity and illegal access.

The big data is the recent trend and more number of web services are using the big data which enables to store huge data, organize and analyze.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
Ms.R.Sivagami*, Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi, India
Dr.A.Nagarajan, Assistant Professor, Department of Computer Applications, Alagappa University, Karaikudi, India

To store the data efficiently the storage management approaches are implemented for every datasets. The storage management involves in classification of sensitive data and normal data, data access roles, format of data and so on. The big data services and approaches are connected to the cloud servers to handle huge data with faster performance [5]. The major problem in cloud based big data handling is detecting and securing the sensitive data. Various approaches are implemented to classify the normal data and sensitive data like clustering, classification techniques and data detection approaches [6].

This proposed scheme is aimed to detect sensitive data from a wide range of data in a less computational time than the existing approaches. After the successful detection of sensitive data, it will be secured with cryptographic techniques by using high secure algorithm and divide and conquer approaches. The portioned data is transferred to the data sets into different cloud servers to ensure the data privacy and reliability. This paper demonstrates the sensitive data detection with data labels and partitioning the sensitive data to encrypt.

II. RELATED WORK

1. Big Data Security Approach in Cloud [1] – The authors concluded that data security is an important aspect. Threats like data loss and data phishing are the main problem for storing the data and security of data. Many research works are developed to guard the data stored in cloud servers. The proposed approach of split process encryption is a noble method to secure data. The sensitive data is detected by various level filtering. The proposed approach uses identify based encryption which detects the sensitive data and encrypt. The encrypted data is split with SA-ED. This proposed approach is concluded with as it is securing the data from various attacks.
2. Intelligent Cryptography Approach [2] - The paper highlights the data storage problems in a cloud and also intends to propose a scheme to the cloud operators to handle sensitive data. The authors have put forward a new approach as SA-EDS (Security Aware Efficient Distributed Storage). AD2, SED2 and EDCON algorithms are proposed by the authors for this approach. The authors have concluded that the techniques will protect the threats from cloud and also the time consumed will be shorter than the existing approaches.

III. PROPOSED APPROACH

This proposed approach is mainly concentrates about isolation of sensitive data from a wide range of data. The approach is aimed for cloud servers and services which uses big data approaches. A cloud based application is designed with various input fields.

The online users input's are organized in temporary heap without transferring into the dataset. The data isolation is done in this level. The sensitive data detection is the essential step to secure from threats. Also the previous works prove that the data detection may consume huge time.

The proposed approach includes a scheme which can perform the sensitive data detection in a faster way called DLDetect. The DLDetect is expanded as *Data Label Detection.* The approach can detect the sensitive data by using the data labels / data heads. For example the input data under the data label PIN is assumed as sensitive data. The approach can work in any type of data. The approach detects concludes the data is sensitive or normal data by the data label which is either fixed or dynamic.

The approach of DLDetect is demonstrated with a sample cloud based web service. The web service receives various inputs from the online user by both determining fixed data labels and dynamic data labels. The proposed approach architecture is illustrated below:

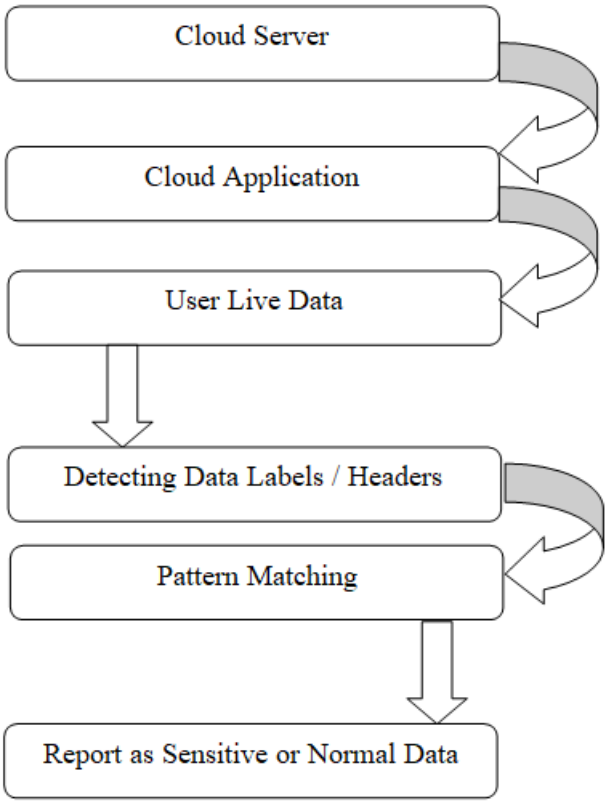

Fig 1: Proposed Architecture - DLDetect

The application receives the various inputs from online user. The data input can be asked by fixed data header labels and the runtime data labels given by the user. After the inputs are submitted by the user, the proposed approach performs a pattern matching process. The pattern matching is a separate unit and it checks the every data label for every user data. If the data label has the sensitive head then the data is treated as sensitive data. On the other hand, the dynamic data label which is given by the user is analyzed with existing patterns. The matching probabilities are counted at behind. The approach concludes the data is sensitive or not by based on the pattern matching probability results.The proposed DLDetect Approach algorithm is described below:

Step 1: Begin
Step 2: Input Live User Data
Step 3: Transfer the data to Temporary Dataset
Step 4: Analyze the data Header or Data Label
Step 5: Decide either Sensitive or Normal
Step 6: Create secondary Dataset and move if sensitive data
Step 7: Load Next data Header or Data Label
Step 8: Repeat Step 4 to Step 7 for every data headers.
Step 9: End

The Proposed DLDetect approach algorithms for dynamic data Labels is below:

Step 1: Begin
Step 2: Input Live Data Header and User Data
Step 3: Transfer the data to Temporary Dataset
Step 4: Partition the Data Header by words
Step 5: Match with Predefined Patterns
Step 6: Calculate Matching Probability
Step 7: Repeat Pattern matching to all the words in the data header.
Step 8: Verify the total probability and decide either sensitive or normal data
Step 9: End

The above two proposed algorithms can equip the server to detect the given data is sensitive or normal data. Both of the above algorithms are structured under the DLDetect Scheme.

IV. EXPERIMENTAL RESULTS

The Proposed DLDetect scheme is implemented with a sample dataset and cloud based web application. The DLDetect scheme can be enhanced into the cloud sever and big data environment. The performance result is illustrated below:

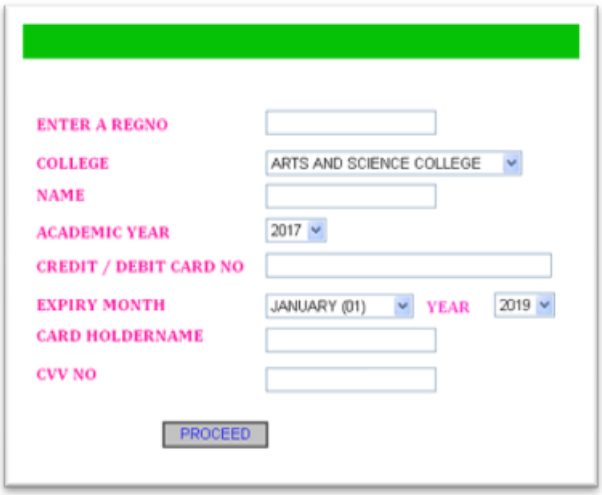

Fig 2: The User Form with Predefined Data Labels / Data Headers

Fig 3: The User Form with Sample Data Preview

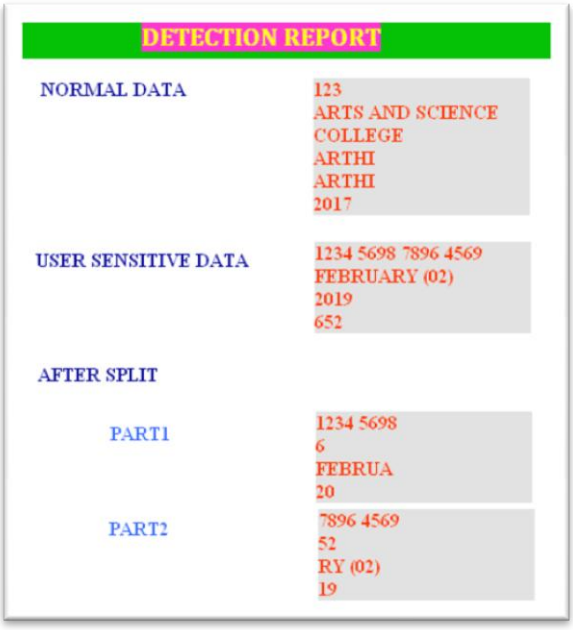

Fig 4: The DLDetect Scheme Detection Report

The performance analysis table is shown below which illustrates the time requirements by the proposed approach and existing approaches.

Fig 5: The Time Consumption Table between the Existing SA – EDS scheme and the proposed DLDetect Scheme.

Process – Time	SA – EDS (In Sec)	DLDetect (In Sec)
Data Read	2.45	1.56389
Data Classification	3.50	2.845
Sensitive Data Detection	4.789	2.489
Report Generation	3.48	1.25412

V. CONCLUSION

In this proposed approach a new scheme is introduced to detect sensitive data. The scheme is called as "*DLDetect*". The Data Label detect scheme is a new and easy way to implement in any web application. The DLDetect scheme is the concept of detecting sensitive and normal data based on its data headers or data labels. The scheme is proposed to handle any situations like fixed form headers and dynamic data headers. The dynamic data headers are handled with pattern matching techniques. The DLDetect scheme detects the user data in fast and accurate manner. The detected data can be applied with cryptographic techniques and partition techniques to secure from threats. This can be done later after the sensitive data detection. The performance analysis report illustrates the DLDetect scheme consumes less computational time than the existing approach.

ACKNOWLEDGMENT

This research work has been supported by RUSA Phase 2.0, Alagappa University, Karaikudi, Tamilnadu, India.

REFERENCES

1. "Big Data Security Approach in Cloud: Review", Kemal Anshar Elmizan, Antoni Wibowo, Finka Ria Damayanti, Yoel Frans Alfredo, Zidni Nurrobi Agam, Data & Knowledge Engineering, pp. 428, 2018.
2. Yibin Li, Keke Gai, Hui Zhao, Longfei Qiu, Meikang Qiu, "Intelligent cryptography approach for secure distributed big data storage in cloud computing" Information Sciences, pp. 103-115, 2016.
3. "Big data technologies and Management: What conceptual modeling can do," ,V. C. Storey and I.-Y. Song, Data & Knowledge Engineering, pp. 50-67, 2017.
4. "Big data processing in cloud computing environments", C. e. a. Ji, in 12th International Symposium on IEEE, 2012.
5. "The rise of big data on cloud computing: Review and open research issues," I. A. T. e. a. Hashem, Information Systems, vol. 47, pp. 98-115, 2015.
6. M. V. A. K. R. , Jyotsna S. Garg, "Reviewing Security Concerns in Cloud Environment," 2017.
7. "Cloud security issues and challenges: A survey," A. Singh, and K. Chatterjee, Journal of Network and Computer Applications, vol. 79, pp. 88- 115, 2017.
8. H.J. Watson, "Tutorial: Big Data Analytics: Concepts, Technologies, and Application" Communications of the Association for Information Systems: Vol. 34 , Article 65, 2014.

AUTHORS PROFILE

Ms.R.Sivagami, have completed B.Sc Computer Science in Madurai Kamaraj University, M.Sc Computer Science in Alagappa University and M.Phil in Bharathidasan University. Now Research scholar in Department of Computer Applications, Alagappa University, Karaikudi. Working as a Assistant Professor in Arts & Science College and have 14 years of teaching experience. Have attended three International conferences.

Dr.A.Nagarajan, Assistant Professor, Department of Computer Applications, Alagappa University, Karaikudi. He has 14 years of teaching experience. He has published 32 papers in International journals. He has attended more than 30 conferences.