CUSTOMIZING SECURITY OF CLOUD DATA

Mohd Anwar Ali, ABBA CHETHANA, J RANJITH
4.KALYAN KUMAR, 5.TAMAKANTH ROOHI
Assistant Professor, , Ellenki college of Engineering and Technonlogy, patelguda (vi), near BHEL ameenpur (m), Sangareddy Dist. Telangana 502319

.

ABSTRACT

In the current age of smart devices and smart phones, any image taken using these devices are immediately auto uploaded to the cloud (Google Photos, iCloud, etc.) or internet (Social media sites like Facebook, Twitter, etc.).Unfortunately, people knowingly or unknowingly upload images containing sensitive data. All this sensitive information falls under three categories: 1) Personal and Private Information 2) Confidential Business Information 3) Classified Information. Loss, misuse, modification, or unauthorized access to sensitive information can adversely affect the privacy or welfare of an individual, trade secrets of a business or even the security and international relations of a nation depending on the level of sensitivity and nature of the information. So, we have designed a machine learning system that classifies the data in the images as sensitive or non-sensitive. Modern machine learning and pattern matching techniques help identify sensitive files such as consent forms, financial statements. Group files into categories such as sales, finance, academia.

Keywords: Machine learning , Confidentiality, data classification, privacy preserving

I. INTRODUCTION

Cloud Computing is an internet founded allotted digital atmosphere. As cloud computing helps companies to sharpen their development and performance. Besides this, it also hosts many users to furnish entry to shared resources with less effort. But security issues or threats are nonetheless a stumbling block in the success route of cloud computing. Numbers of factors are the subject. First intent is that users and many businesses store their knowledge on cloud storage, so the most important focus is the info ought to be comfy, and the info are not being lost and tampered even as traveling from one situation to yet another over the network. So it is main that confidentiality, availability and integrity of data will have to be ensured. Secondly, unauthorized access where an attacker tries to be the impersonator of the legal person. In the current age of smart devices and smart phones, any image taken using these devices are immediately auto uploaded to the cloud (Google Photos, iCloud, etc.) or internet (Social media sites like Facebook, Twitter, etc.).Unfortunately, people knowingly or unknowingly upload images containing sensitive data. All this sensitive information falls under three categories: 1) Personal and Private Information 2) Confidential Business Information 3) Classified Information. Loss, misuse, modification, or unauthorized access to sensitive information can adversely affect the privacy or welfare of an individual, trade secrets of a business or even the security and international relations of a nation depending on the level of sensitivity and nature of the information. So, we have designed a machine learning system that classifies the data in the images as sensitive or non-sensitive. Modern machine learning and pattern matching techniques help identify sensitive files such as consent forms, financial statements. Group files into categories such as sales, finance, academia. As shown in Fig 1 how the dataset is been taken from the various systems like tourism, Government and Business organization. These information is been processed various data classification algorithms and the with the that dataset the system is been trained to know

which data is less confidential , average confidential and highly confidential.
Sample Dataset:

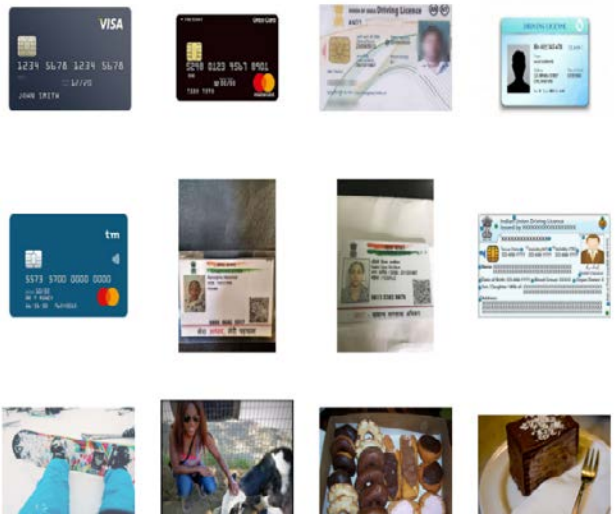

Fig 1:Sample dataset containing sensitive and non-sensitive images

data, now we have proposed an information classification mannequin to classify the data according to its sensitivity stage and then encrypting the one knowledge which is required to comfy using an encryption procedure in cloud atmosphere.Data classification is a laptop studying process used to foretell the category of the unclassified understanding. Knowledge mining makes use of specified instruments to grasp the unknown, respectable patterns and relationships within the dataset. These tools are numerical calculations, factual models and prediction and evaluation of the info. Hence, data mining contains management, collection, prediction and analysis of the data. ML algorithms are described in to 2 classes: supervised and unsupervised. In supervised studying, courses are already outlined. For supervised studying, first, a test dataset is defined which belongs to individual courses. These lessons are appropriately labeled with a specific identify. Lots of the data mining algorithms are supervised finding out with a designated intention variable. In unsupervised learning classes aren't without difficulty characterized but as a substitute arrangement of the know-how is performed automatically. The unsupervised algorithm looks for similarity between two gadgets in order to find whether they are able to be characterized as forming a cluster. In simple words, in unsupervised learning, “no goal variable is identified”. The classification of know-how within the context of confidentiality is the classification of expertise headquartered on its sensitivity level and they have an effect on to the organization that capabilities be disclosed handiest licensed users. The info classification helps investigate what baseline security standards/controls are correct for safeguarding that knowledge. The knowledge is labeled into two classes, personal and non-exclusive (non-distinct) understanding. The classification of the information relies on the attributes of the knowledge. The values of the sensitive attributes are labeled as “confidential “and “highly confidential” and values of the non- touchy attributes are categorized as “basic”.

The remainder of this paper is organized as follows: In section 2, related work is mentioned. In section 3, proposed work is presented. In section 4, results and discussions are discussed. The document has been concluded in section 5 with future research directions.

II. RELATED WORK

Sinha N et.al [8] This paper provides the brief history of cloud computing with its benefits architecture implementation and all issues in cloud computing. It provides the basic idea of all the different kind of issues related to security, data and performance in cloud. Diwan V et.al [9] This paper different cryptographic algorithm are been compared which are been taken into consideration to provide the confidentiality of the data. In this different cryptographic algorithms are being compared by considering different parameters like block size, key length type and features. This paper had provided the idea of different cryptographic algorithm which can be used to ensure the security of data in cloud. Zardari MA et.al[10] In this paper they had used the K-NN approach for in order to do classification of data in order to provide the confidentiality of data. The main aim to classify data is to provide security. In this approach they classify the data into labels sensitive and non-sensitive data using K-NN algorithm. On the sensitive data the encryption is done inorder to provide the security. Classification is mainly performed because it become easy to select an appropriate security for data according to need of data. So this way

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-3, ISSUE-10, 2016
DOI: DOI:10.21276/ijcesr.2016.3.10.7

it will enhance the security. Shaikh, Rizwana et.al [11] This paper had contributed another important technique inorder to enhance the security of data in cloud by considering the classification technique. In this different parameters are been taken into consideration to provide the classification of data and then on the classified data the encryption is performed. Different classification properties considered are access control, content and storage. On these properties the data classification is done and then theencryption is performed to enhance more security and better efficiency. Tawalbeh L et al. [12] In this paper they had contributed the secure cloud computing model based on the classification which basically minimizes the overhead and processing time needed to secure the data through using different security mechanism with variable keysizes to provide the confidentiality level required for data. Classification is done by user manually and the encryption is been performed in three different level with different cryptographic algorithm. The levels are based on the sensitivity of data which includes Basic, Confidential and highly confidential level.The different cryptographic algorithm are been used at different levels to provide the security of data upto a great extent.

III. PROPOSED WORK

The research involves exploring various data classification algorithms.

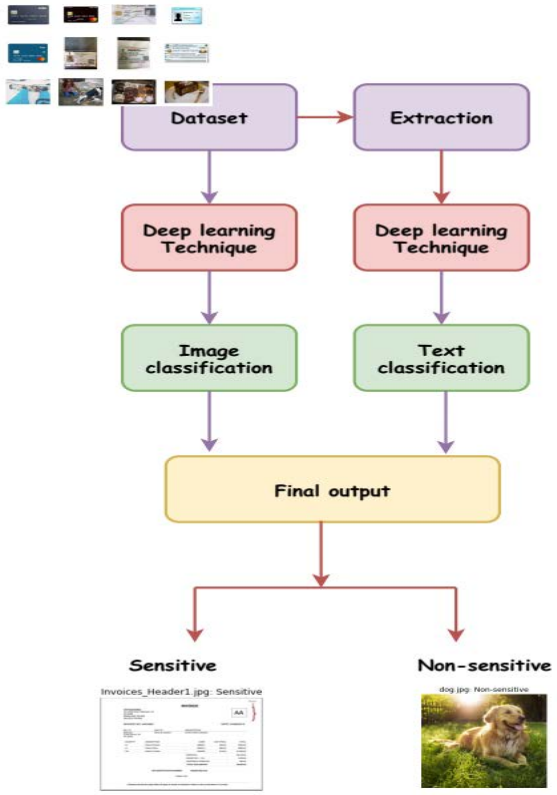

Fig 2: Architecture overview

Data classification for Cloud:

Successful data classification in an organization requires broad awareness of your organization's needs and a thorough understanding of where your data assets reside.Data exists in one of three basic states:

At rest

In process

In transit

All three states require unique technical solutions for data classification, but the applied principles of data classification should be the same for each. Data that is classified as confidential needs to stay confidential when at rest, in process, and in transit.

Data can also be either structured or unstructured. Typical classification processes for the structured data found in databases and spreadsheets are less complex and time-consuming to manage than those for unstructured data such as documents, source code, and email. Classification process:

This work introduces two generalized terminology models that are based on well-used and industry-respected models. These terminology models, both of which provide

ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-3, ISSUE-10, 2016
DOI: DOI:10.21276/ijcesr.2016.3.10.7

three levels of classification sensitivity, are shown in the following Table.1.

Sensitivity	Terminology model 1	Terminology model 2
High	Confidential	Restricted
Medium	For internal use only	Sensitive
Low	Public	Unrestricted

Protecting confidential data:

After data is classified, finding and implementing ways to protect confidential data becomes an integral part of any data protection deployment strategy. Protecting confidential data requires additional attention to how data is stored and transmitted in conventional architectures as well as in the cloud.As the following figure shows, these technologies can be deployed as on-premises or cloud-based solutions—or in a hybrid fashion, with some of them deployed on-premises and some in the cloud. (Some technologies, such as encryption and rights management, also extend to user devices.)

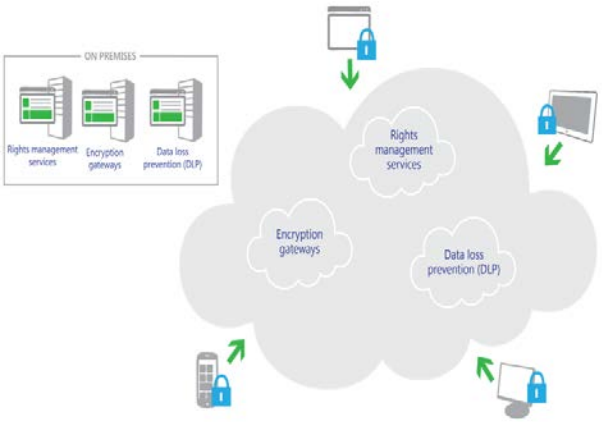

Fig 3: Overview of Protecting confidential data

Encryption gateways operate in their own layers to provide encryption services by rerouting all access to cloud-based data. This approach should not be confused with that of a virtual private network (VPN). Encryption gateways are designed to provide a transparent layer to cloud-based solutions.Encryption gateways can provide a means to manage and secure data that has been classified as confidential by encrypting the data in transit as well as data at rest. Encryption gateways are placed into the data flow between user devices and application data centers to provide encryption/decryption services. These solutions, like VPNs, are predominantly on-premises solutions. They are designed to provide a third party with control over encryption keys, which helps reduce the risk of placing both the data and key management with one provider. Such solutions are designed, much like encryption, to work seamlessly and transparently between users and the service.

Performance evaluation:
We evaluate the detection accuracy in simple and complexleaking scenarios. First, we test the detection rate and false positive rate in three simple experiments where the sensitive datais leaked in its original form or not leaked. Then we presentaccuracy evaluation on more complex leaking experiments toreproduce various real-world leaking detection scenarios. Therefore, in order to reduce the encryption time on cloud data is classified according to its security needs using machine learning algorithms.

IV. CONCLUSION

In this research, a technique for data confidentiality in cloud environment using data classification is proposed. The basic contribution of this security model is data confidentiality and classification of data using machine learning classification approach. The classified confidential data into three classes named as basic,confidential and highly confidential on the basis of the sensitivity of data is then encrypted using different cryptographic techniques and is stored in the cloud server .

ACKNOWLEDGMENT

I would like to place on record my deep sense of gratitude to Assistant Professor Mr. M Prashanth for his valuable suggestions in my research work. I would like to thank all the people whose encouragement and support has made the fulfilment of this work conceivable

REFERENCES

[1] Munwar ali zardari, Low Tang Jung, Nordin Zakaria," K-NN Classifier for Data

Confidentiality in Cloud Computing", IEEE, pp.1-6, 2014.

[2] Almorsy, M., Grundy, J., & Ibrahim, A. S., "Collaboration- Based Cloud Computing Security Management Framework" IEEE conference of cloud computing, Washington (DC), pp. 364-371,2011.

[3] Song, D., E. Shi, I. Fischer and U. Shankar, "Cloud data protection for the masses", IEEE Computer. Soc., Vol. 45, Issue 1, pp.39-45, 2012

[4] 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), 2010 1st International Conference on 2010 Oct 28 (pp. 211-216). IEEE.

[5] Dubey AK, Dubey AK, Namdev M, Shrivastava SS. Cloud-user security based on RSA and MD5 algorithm for resource attestation and sharing in java environment. InSoftware Engineering (CONSEG), 2012 CSI Sixth International Conference on 2012 Sep 5 (pp. 1-8). IEEE.

[6] Yellamma P, Narasimham C, Sreenivas V. Data security in cloud using RSA. InComputing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on 2013 Jul 4 (pp. 1-6). IEEE.

[7] Rewagad P, Pawar Y. Use of digital signature with Diffie Hellman key exchange and AES encryption algorithm to enhance data security in cloud computing. InCommunication Systems and Network Technologies (CSNT), 2013 International Conference on 2013 Apr 6 (pp. 437-439). IEEE.

[8] Sinha N, Khreisat L. Cloud computing security, data, and performance issues. In2014 23rd Wireless and Optical Communication Conference (WOCC) 2014 May 9 (pp. 1-6). IEEE.

[9] Diwan V, Malhotra S, Jain R. Cloud security solutions: Comparison among various cryptographic algorithms. IJARCSSE, April. 2014 Apr.

[10] Zardari MA, Jung LT, Zakaria N. K-NN classifier for data confidentiality in cloud computing. InComputer and Information Sciences (ICCOINS),

[11] Shaikh, Rizwana, and M. Sasikumar. "Data Classification for achieving Security in cloud computing." Procedia Computer Lo'aiTawalbeh NS, Raad S. Al-Qassas and Fahd AlDosari,"A Secure Cloud Computing Model based on Data Classification". InFirst International Workshop On Mobile Cloud.