13th International Conference on Science, Technology and Management
Mahratta Chamber of Commerce, Industries and Agriculture, Pune (India) (ICSTM-18)
01st-02nd December 2018 www.conferenceworld.in ISBN:978-93-87793-58-3

AN APPROACH OF INFORMATION RETRIEVAL FROM CLOUD USING SECURE DATA MINING PROCESS

Aditya Sharma[1], Dr. Amit Sharma[2],

[1]Mtech Scholar Computer Science & Engineering Department

Vedant College of Engineering and Technology Bundi, Rajasthan, India

[2] Professor, Computer Science & Engineering Department,

Vedant College of Engineering and Technology, Bundi, Rajasthan, India

ABSTRACT-

Cloud computing is an advanced term alludes to a model for rising processing, where it is conceivable to utilize machines in substantial server farms for conveying administrations in an adaptable way, so partnerships has moved toward becoming in requirement for vast scale modest figuring. As of late, a few governments have started to use distributed computing models, applications and stages for addressing the requirements of their constituents and conveying administrations. Security possesses the main rank of hindrances that confront distributed computing for administrative offices and organizations. Additionally, Cloud Computing is one of the promising innovation in which established researchers has as of late experienced. Distributed computing is identified with other research regions, for example, disseminated and network figuring, Service-Oriented Architecture, and virtualization, as distributed computing acquired their confinements and headways. It is conceivable to misuse new open doors for security. This paper point is to talk about and examine how accomplish alleviation for distributed computing security hazards as a fundamental advance towards getting secure and safe condition for distributed computing. Cloud computing is a decent stage for research and application of information mining, for the reason that it gives incredible limits of capacity and figuring, great asset the executives based on virtualization and asset sharing model, and complete administration framework. Be that as it may, examination on information mining in distributed computing condition is still in its early stages

Keywords : Cloud computing, Cloud security issue, Data mining, Classification, Naive Bayes; multilayer perception; Support vector machine; decision tree (C4.5); and Partial Tree (PART)

I. INTRODUCTION

At the point when cloud applications are mainstream and enormous information has collected, information mining is an imperative issue for cloud administrations, for example, Salesforce.com and YouTube. With the end goal to enhance nature of these cloud administrations, information mining on enormous information is urgent, since customaries administrations depend on important data obtained by information mining. YouTube conducts suggestion by breaking down verifiable information and mining clients' interests. For Salesforce.com, information mining is an essential strategy to give CRM benefit. Distributed computing has turned into a most well known research popular expression. Driving IT partnerships, for example, Google, Amazon and IBM has proposed some distributed computing models. Some exploration foundations have likewise created distributed computing stages. For example, in the Science Clouds Project started by University of Chicago and University of Florida, Nimbus Cloud and Florida Cloud were investigated to give rent assets to academic network [1]. The characteristic highlights of cloud can be finished up as: enormous putting away and registering limits, adaptable and adaptable assets and structure, and on-request benefit by means of virtualization and asset pool. These attributes make it conceivable to actualize information mining as a business application, and make information mining in distributed computing an examination zone significant in principle and practice. In spite of the fact that cloud is brilliant away and calculation, it is additionally fundamental to have devices and conditions that help investigation and revelation over these information.

Distributed computing is a decent stage for information mining. It can counterbalance imperfections of past strategies in breaking down system information. In cloud, assets for capacity and registering were dispersed. In this way, information mining in cloud is led fundamentally not quite the same as the conventional mining worked on neighborhood PCs, and meets the necessities of information mining in Internet. The framework of distributed computing is built of significant server groups, which supply the cloud with ground-breaking limits of registering, putting away, information examining and information the executives. These limits give basic preconditions to monstrous information mining on the Internet. What's more, IT assets and applications are given as open offices in cloud. As the manner in which you utilize water, power and gas, you can utilize assets and application in cloud without thinking about where they originate from and how to create them. This is an administration arranged IT application show, which can be more versatile to prerequisites of information mining improvement and application. Moreover, as indicated by the SaaS (Software as a Service) plan of action of distributed computing, information mining projects, programming or stages

are bundled as an administration and sold to clients and designers. Ventures can enhance the versatility of their administrations and manage barges in asset requests by utilizing cloud administrations [2]. This will encourage little and medium-sized endeavors decrease the expense of programming improvement when actualize information mining, and spread business utilization of information mining therefore.

In the previous couple of decades, parallel, conveyed and matrix methods were connected to information mining. For parallel and disseminated ideal models, database was partitioned into a few portions, which were dispersed to various figuring hubs for information mining. By such a methodology, the worldwide computational exertion is shared. Furthermore, the registering proficiency increments in light of the fact that the subtasks work on circulated information locales simultaneously [3]. Information matrix offers apparatuses and methods for disseminated mining and extraction of learning from information vaults accessible on the network [4].

Since information mining assignments turn out to be progressively perplexing as information amassing, investigate in the previous couple of decades was centered around parallel and dispersed mining procedures. In the majority of the exploration, database was separated into a few sections, which were appropriated to various figuring hubs for information mining. By such a system, the worldwide computational exertion is shared. Furthermore, the processing productivity increments in light of the fact that the subtasks work on circulated information destinations simultaneously [3]. In any case, the figuring hubs will trade exchange data among one another amid the mining procedure. The high effectiveness will be undermined by continuous and enormous information trading. In the mean time, data handling in system requires ongoing correspondence. Yet, parallel and dispersed information mining don't ensure astounding system of data sharing and collaboration to satisfy such a critical prerequisite. Furthermore, the information protection and security is likewise a noteworthy worry, since information might be unlawfully assaulted when the parallel and appropriated calculations copies the database to each hub [5]. With the end goal to defeat these issues, analysts have propelled examination on information the board and information investigation in distributed computing condition. Sakr et al. gives an exhaustive overview of various methodologies and components of sending information escalated applications in the cloud, and talks about some open issues and future difficulties relating to adaptability, consistency, conservative handling of extensive scale information on the cloud [6]. Distributed computing has opened up the test for structuring information the executives frameworks that give consistency ensures at a bigger granularity. Along

these lines, Agrawal et al. feature some structure standards for frameworks giving adaptable and predictable information the executives as an administration in the cloud [7]. With the end goal to help the ECG information examination, Pandey et al. structure an autonomic cloud condition that gathers wellbeing information and scatters them to a cloud-based data vault and encourages information investigation utilizing programming administrations in the cloud [8]. In spite of the fact that information the board and investigation in cloud have been investigated inside and out, examine concentrated on information mining in cloud isn't sufficient. Issues, for example, calculation and framework engineering of information mining in cloud, require further examination.

In any case, information mining in distributed computing condition is definitely not a novel field. It tends to be actualized in cloud as indicated by some customary systems. A few downsides of information mining can be undermined when abused in cloud. The test here is the manner by which to adjust existing information mining models and methods into the cloud. Thus, in this paper, we abuse the distributed computing condition naming Cloud based Genetic Classification Rules Mining Model (CGCRMM) to address grouping rules mining issue. The system of grouping is orchestrated considering the circulated and parallel cloud condition. What's more, the adjusted hereditary calculation, which makes great utilization of the registering intensity of distributed computing, is intended to illuminate this model. For preparing and testing the proposed model, we utilize information gathered from UCI dataset to lead an illustrative precedent. Rest of the paper is organized as pursues: Section 2 is a short survey of the writing pertinent to information mining in distributed computing condition; Section 3 depicts the essential strategies and point by point development of the CGCRMM demonstrate; Experiment in area 4 assess the legitimacy and execution of the proposed model; Section 5 finishes up the entire research.

II. RELATED WORKS

Distinguishing with the traditional mining paradigms, data mining in cloud is a novel area filled with valuable issues worthy of investigation. Basing on an intensive review on the relevant literature, most of the researchers concentrate on the following problems.

2.1. Data Mining Algorithm

Cloud computing, with its promise of virtually infinite computing and storage resources, is suitable to solve resource greedy computing problems. One problem of data mining in the cloud has been investigated from the data mining algorithm perspective. Wang et al. [9] utilized the powerful and huge capacity of cloud computing into data mining and machine learning. In their experiments, three

algorithms, i.e., global effect (GE), K-nearest neighbor (KNN) and restricted boltzmann machine (RBM) were performed in cloud computing platforms, which use the S3 and EC2 of Amazon Web Services. And they built two predictors based on KNN model and RBM model respectively with the order to testify their performance based on cloud computing platforms.

The MapReduce programming model was designed for processing massive data sets in a parallel network. Based on this programming model, Wang et al. [10] adapted the SPRINT algorithm which is ideal tool for data classification. SPRINT has been designed to be easily parallelized. Due to the parallelism, the original SPRINT was modified to be implemented in Hadoop architecture. The algorithm divided datasets in vertical direction and horizontal direction respectively, in accordance with the "Map" step in MapReduce. The vertical partition separated datasets by attribute, while horizontal partition produced many item sets. They applied the revised SPRINT algorithm to classify customer groups with different credit grades.

III METHODOLOGY

Classification is an important mission in data mining, and probably has become the most studied data mining task. In this task, the goal is to predict the value of a specified goal attribute (called the class attribute) based on the values of other attributes (called the predicting attributes).

Generally speaking: With pairing based cryptography each attribute is represented as a group element. By virtue of the bilinearity property it allows for two independent sets of operations to be performed upon a set of group elements representing each Pi 2 P. These operations hide the secret exponent among the group elements such that when the result of these operations are combined if the conditions are right the secret exponent to be recovered. These conditions are dictated by the LSSS.

The precise use of LSSSs to hide and recover the secret exponent is dependent not only upon the placement of the predicates within the PBE scheme but also upon the exact predicate used. For the remainder of this section, a general overview of how LSSSs are used as part of both CP and KP schemes. Section 8.5 provides a concrete example of how one can use LSSS precisely as

3.1 Cipher text-Policy

with CP schemes a message will be encrypted under an access policy A and can be decrypted from a key that is derived from a set of attributes S. The use of LSSS within CP schemes can be seen as

LSSS in its standard form. Encrypt The LSSS is used to generate piece vectors, from the secret exponent, for each group element that represents an attribute Pi 2 P as de

Recall from Section 7.3.2 that with CP schemes a message will be encrypted under an access policy A and can be decrypted from a key that is derived from a set of attributes S. The use of LSSS within CP schemes can be seen as LSSS in its standard form.

Encrypt The LSSS is used to generate piece vectors, from the secret exponent, for each group element that represents an attribute Pi 2 P as defined in A. These vectors along with a description of the LSSS are stored alongside the encrypted message.

Key Generation With the generation of decryption keys each user is assigned a set of attributes represented as a set of group elements modified by some random secret value.

Decrypt During decryption the LSSS, described in the cipher-text, will only reconstruct the secret exponent if an authorized set of attributes can be found within the elements of the decryption key.

Key-Policy

With Key-Policy schemes, a message will be encrypted under a set of attributes S and can be decrypted using a key that is derived from an access policy A. The use of LSSS deifiers from its standard use.

Encrypt After the message has been encrypted the secret exponent is hidden among a set of group elements that have been derived from each Si 2 S, the encryption key. These elements along with a description of S are stored along side the encrypted message.

Key Generation When generating the decryption key, the LSSS is used to distribute a secondary secret value among each attribute Pi 2 P from A. The resulting piece vectors and description of A are returned as the decryption key.

Decrypt With decryption the LSSS, taken from the decryption key, will only reconstruct the secret exponent if an authorized set of elements exists within the elements stored alongside the cipher-text.

IV. CLOUD ENVIRONMENT LAYERS

Cloud computing attracts many managers and organizations. There are many similar terminologies that are usually utilized for describing cloud computing, these terms such as: distributed, grid, cluster, virtualization, on-demand, utility, and software-as-a-service. In other words, cloud computing refers to end -users connecting with applications running on sets of shared servers, often hosted and virtualized, instead of a traditional dedicated server.

A. Deployment Models The deployment models can be categorized into four categories namely Public cloud, Private cloud, Hybrid cloud and Community cloud [6]. These categories will be described in details as follows:-

• **Public Cloud,** this model is owned by an organization for selling the cloud services and the design of infrastructure is made in order to be available for industries, organizations and businesses.

• **Private Cloud**, this model is managed by the organization itself or by a third party. Private cloud may be either off or on premises. The major characteristic of this model is that the infrastructure of the cloud is private, in addition to its availability to a single organization.

• **Hybrid Cloud, this model** is similar to the private cloud as it is managed by third party or by organization itself and may exist off or on premises. But the cloud infrastructure may combine two or more clouds (public, private or community).

• **Community Cloud, this model is similar to the previously mentioned private and hybrid cloud** as the organization or third party are allowed to manage it and also exists off or on premises. But in community cloud, multiple organizations with common interests, requirements, or considerations share the infrastructure.

The security of the cloud needs testing, it is important for organizations that want to ensure the optimal product before distributing it. The results are used in finding out security weakness points and to patch them before the occurrence of penetration. However organizations' lack of time and resources, computer related crime is usually on the rise. Consequently penetration investigators (testers) have to reduce the amount of resources. This motivates testers to widely adopt automatic tools, as it is demonstrated by the continuous release of platforms finalized to automate this process, discovering gaps in compliance, verifying secure Configurations, finding holes now before somebody else does, Report problems to management and testing new technology.

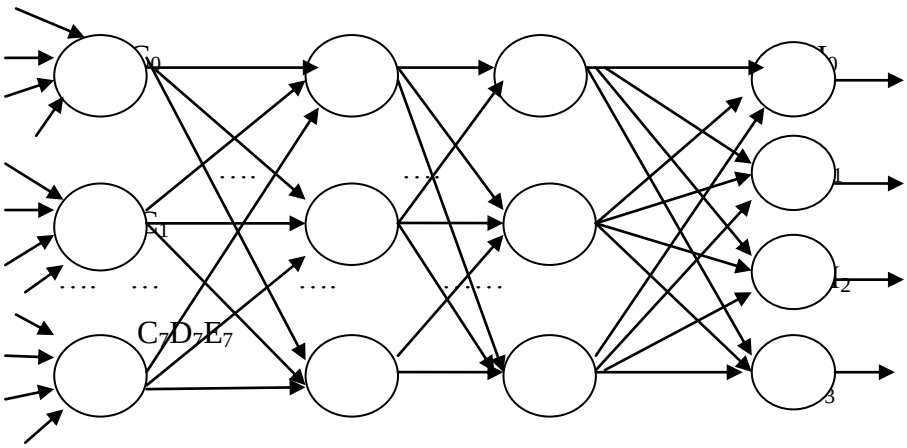
V Proposed Algorithms

Proposed Hash function based on key generated

Message is given as input to the four layers feed forward neural network. In the proposed hash function the network is used is composed of three layers: input layer, hidden layer1,hidden layer2, and output layer. Four neuronal layer are used to realize data confusion, diffusionand compression. Let the layers inputs and output be $M=[M_1M_2\ldots M_{31}]$,$C=[C_0C_1\ldots..C_7]$,$D=[D_0D_1\ldots D_7]$,$E=[E_0E_1\ldots E_7]$and $H=[H_0H_1H_2H_3]$ so functioning of neural network is defined as-

$H=$

$f_3(W_3E+B_3)=f_3(W_3f_2(W_2D+B_2)+B_3)=f_3(W_3f_2(W_2f_1(W_1C+B_1)+B_2)+B_3)=f_3(W_3f_2(W_2f_1(W_1f0(W_0M+B_0)+B_1)+B_2)+B_3)$



Input Layer | Hidden Layer | Hidden Layer | Output Layer

Figure 1: Four layer feed forward network

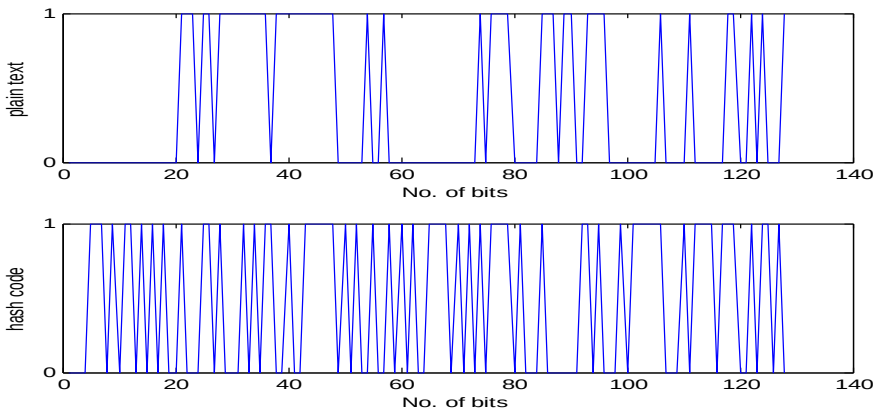
Where f_i, W_i and $B_i(i=0,1,2,3)$ are the piecewise linear chaotic map as transfer function, weight and bias of the ith neuron respectively.

And f is defined as: Z(k+1)=f(Z(k),Q) =

$$(Z(k),Q) = \begin{cases} \frac{Z(k)}{Q}, 0 \le Z(k) < 0.5 \\ \frac{Z(k)-Q}{0.5-Q}, Q \le Z(k) < 0.5 \\ \frac{1-Q-Z(k)}{0.5-Q}, 0.5 \le Z(k) < 1-Q \\ \frac{1-Z(k)}{Q}, 1-Q \le Z(k) \le 1 \\ Z(k)-1, \quad Z(k) \ge 1 \end{cases}$$

Where Q is control parameter and its value in [0, 0.5].Here map is piecewise linear and it is in chaotic state when 0<Q<0.5.This chaotic map has parameter sensitivity means a small change in initial values Z(k),Q causes drastic change in iterated values Z(k+T) which is suitable for constructing a cipher. Generally chaotic map is iterated for (T >=50) to keep the randomness of output for each layer. Chaotic map f() is applied on each layer and used as an activation function. Input and output layers are iterated T times in order to improve the randomness of output of hidden layers and D,E and H.

0EB548D1593F5254F55E881A2FC5EE5A



52FAA79169E235AA31B7A73FDF49CBD4

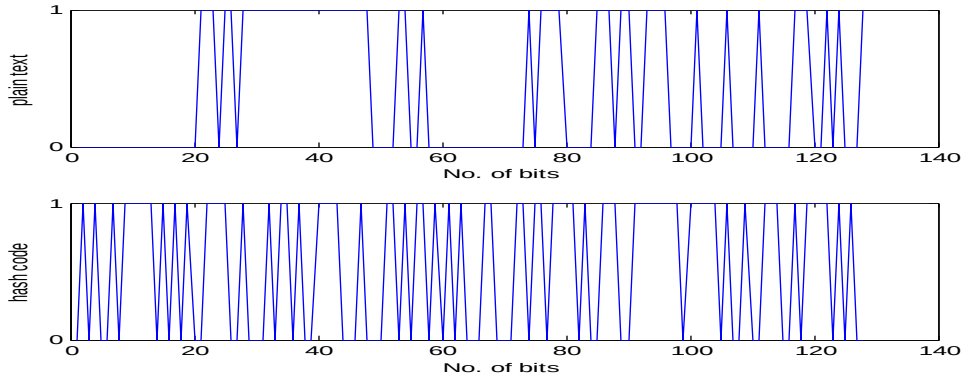


Figure 2: Key generation and Testing

VI Conclusion

Cloud Computing embodies the as-a-Service paradigm and allows for services to be provided en masse to consumers. The problems associated with the use of cloud based services can be summarized by the unknown risk profile and unknown expectation of privacy sees Section. When service users push data to the cloud they need to rely upon Cloud Service Providers (CSPs) adhering to their remit, and doing so dutifully. However, when looking to build solutions to protect data in the cloud it is important to remember that for the service user the CSP can be trusted, albeit at arms length see Section. The threat models presented in illustrate that threats to data occur both in the domain of the service user and the domain of the CSP. Traditional privacy models are too user-centric and CSP-fearing when trying to address the problem of protecting. A privacy model centered around Kafka's The Trial helps to address this problem, this privacy model indicates that when protecting one's data one should also have control over its use rather than solely preventing its collection: CSPs and service users need to work together.

Reference

[1] Abdullah, A. M., & Aziz, R. H. H. (2016, June). New Approaches to Encrypt and Decrypt Data in Image using Cryptography and Steganography Algorithm., International Journal of Computer Applications, Vol. 143, No.4 (pp. 11-17).

[2] Singh, G. (2013). A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. International Journal of Computer Applications, 67(19).

[3] Gaj, K., & Chodowiec, P. (2001, April). Fast implementation and fair comparison of the final candidates for Advanced Encryption Standard using Field Programmable Gate Arrays. In Cryptographers' Track at the RSA Conference (pp. 84-99). Springer Berlin Heidelberg.

[4] Stallings, W. (2006). Cryptography and network security: principles and practices. Pearson Education India.

[5] Yenuguvanilanka, J., & Elkeelany, O. (2008, April). Performance evaluation of hardware models of Advanced Encryption Standard (AES) algorithm. In Southeastcon, 2008. IEEE (pp. 222-225).

[6] Lu, C. C., & Tseng, S. Y. (2002). Integrated design of AES (Advanced Encryption Standard) encrypter and decrypter. In Application-Specific Systems, Architectures and Processors, 2002. Proceedings. The IEEE International Conference on (pp. 277-285).

[7] Mohamed, A. A., & Madian, A. H. (2010, December). A Modified Rijndael Algorithm and it's Implementation using FPGA. In Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on (pp. 335-338).

[8] Pramstaller, N., Gurkaynak, F. K., Haene, S., Kaeslin, H., Felber, N., & Fichtner, W. (2004, September). Towards an AES crypto-chip resistant to differential power analysis. In Solid-State Circuits Conference, 2004. ESSCIRC 2004. Proceeding of the 30th European IEEE (pp. 307-310).

[9] Deshpande, H. S., Karande, K. J., & Mulani, A. O. (2014, April). Efficient implementation of AES algorithm on FPGA. In Communications and Signal Processing (ICCSP), 2014 IEEE International Conference on (pp. 1895-1899).

[10] Nadeem, H (2006). A performance comparison of data encryption algorithms," IEEE Information and Communication Technologies, (pp. 84-89).

[11] Diaa, S., E, Hatem M. A. K., & Mohiy M. H. (2010, May) Evaluating the Performance of Symmetric Encryption Algorithms. International Journal of Network Security, Vol.10, No.3, (pp.213-219).

[12] Jain, R., Jejurkar, R., Chopade, S., Vaidya, S., & Sanap, M. (2014). AES Algorithm Using 512 Bit Key Implementation for Secure Communication. International journal of innovative Research in Computer and Communication Engineering, 2(3).

[13] Selmane, N., Guilley, S., & Danger, J. L. (2008, May). Practical setup time violation attacks on AES. In Dependable Computing Conference, 2008. EDCC 2008. Seventh European (pp. 91-96). IEEE.

[14] Berent, A. (2013). Advanced Encryption Standard by Example. Document available at URL http://www. networkdls. com/Articles/AESbyExample. pdf (April 1 2007) Accessed: June.

[15] Benvenuto, C. J. (2012). Galois field in cryptography. University of Washington.

[16] Lee, H., Lee, K., & Shin, Y. (2009). Aes implementation and performance evaluation on 8-bit microcontrollers. arXiv preprint arXiv:0911.0482.

[17] Padate, R., & Patel, A. (2014). Encryption and decryption of text using AES algorithm. International Journal of Emerging Technology and Advanced Engineering, 4(5), 54-9.

[18] Reddy, M. S., & Babu, Y. A. (2013). Evaluation of Microblaze and Implementation of AES Algorithm using Spartan-3E. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(7), 3341-3347.