Secure Data Mining in Cloud using Homomorphic Encryption for numeric and text data

Jaydeep Jayprakash Patil, Prof. S P Medhane
PG Student Department of Information Technology, Bharati Vidyapeeth (Deemed to be) University College of Engineering Pune, India
Prof. Department of Information Technology, Bharati Vidyapeeth (Deemed to be) University College of Engineering Pune, India

Abstract- The Purpose of Homomorphic encryption is to confirm privacy of knowledge in communication, storage or in use by processes with mechanisms almost like standard cryptography, however with further capabilities of computing over encrypted information, looking out Associate in Nursing encrypted information, etc. homomorphy may be a property by that a drag in one algebraical system may be reborn to a drag in another algebraically system, be solved and therefore the resolution later may be translated back effectively. Thus, homomorphy makes secure delegation of computation to a third party potential. several standard encoding schemes possess either increasing or additive homomorphic property and presently in use for various applications. Yet, a Homomorphic encoding theme with some Categorical Attributes (HECA) that may perform any capricious computation over encrypted information appeared in 2009 as Gentry's work. during this paper, we have a tendency to projected information partition exploitation HECA including some numeric as well as some categorical attributes using big data**.**

Keywords- Homomorphic Encryption, multi-cloud; privacy; fully homomorphic encryption; distributed System; confidentiality.

I. INTRODUCTION

Cryptosystems supply mechanisms to ensure data confidentiality and integrity. If the data is usually encrypted within the cloud, then management isn't lost, and therefore the issues area unit removed. once Associate in Nursing secret writing rule doesn't permit absolute computation over encrypted information, the encrypted information should be decrypted before the computation, and therefore the decrypted information is not any longer in check. Cloud computing is impossible for several business organizations if they have to transfer sensitive information from the cloud to a trustworthy pc so as to perform operations, so send the encrypted results backed to the cloud. Encrypted information has traditionally been not possible to control on while not 1st decrypting them. There area unit some secret writing algorithms that permit absolute computation on encrypted information. as an example, RSA could be a multiplicatively homomorphic secret writing rule wherever the coding of the merchandise of encrypted information are the merchandise of the 2 plain information. However, RSA doesn't permit addition operation nor the mix of multiplications and additions. Later, FHE has appeared [1] to perform unlimited chaining of algebraically operations within the cipertext, which suggests that Associate in Nursing absolute variety of additives and multiplications will be applied to encrypted operands. sadly, all implementations of FHE schemes showed that this technique is still much too slow for practical applications.

The aim of this Somewhat homomorphic Scheme (SHS) is to construct AN coding theme that's "almost" bootstrappable with relevancy a universal set of gates. the primary step is to style a SHE theme that may be a theme that supports some computations over encrypted knowledge. upper class then showed that if you'll manage to style a SHE theme that supports the analysis of its own cryptography algorithmic rule (and a bit more), then there's a general technique to rework the SHE theme into a HECA theme. A SHE which will measure its own cryptography algorithmic rule homomorphically is named bootstrappable and therefore the technique that transforms a bootstrappable SHE theme into a FHE scheme is called bootstrapping

II. LITERATURE SURVEY

According to Mr. V. Biksham and Dr. D. Vasumathi [1] proposed security and Privacy in Cloud Computing: they need declared that, starting with these attributes, they gift the relationships among them, the vulnerabilities which will be exploited by attackers, the threat models, similarly as existing defense methods during a cloud state of affairs. once outsourcing the info and business application to a 3rd party causes security and privacy problems to become a crucial concern. Throughout the study at hand, the authors get a standard goal to supply a comprehensive review of the prevailing security and privacy issue in cloud environments.

According to S. SelvaRatna and Dr. T. Karthikeyan [2] Survey security and Privacy in Cloud Computing: they need declared that, starting with these attributes, they gift the relationships among them, the vulnerabilities that will be

exploited by attackers, the threat models, similarly as existing defense ways during a cloud situation. once outsourcing the info and business application to a 3rd party causes security and privacy problems to become a crucial concern. Throughout the study at hand, the authors acquire a typical goal to supply a comprehensive review of the present algorithm is presented.

Shashank Bajpai and Padmija Shrivastava [3] Data Mining Approach in Security Information and Event Management: This paper provides an summary info} mining field& security information event management system. however numerous data processing techniques may be utilized in security info and event management system to reinforce the capabilities of the system . data processing is turning into more and more common in each the non-public and public sectors. the important drawback in today's enterprise security is quantity of logs generated by numerous systems. Organizations usually place an excessive amount of religion in their new shiny firewalls. The drawbacks of this technique use different data processing technique like classification, clump to reinforce the system capability. numerous techniques are introduced to reduce false positive alerts and reduce CPU loads on system.

Deepti Mittal et. Al. [4] Security threat issues and measures in cloud computing: they need expressed that this technique is especially focuses on security threats of cloud automatic data processing system additionally they mention some answer and countermeasure on this security downside, it highlighted of these issue of cloud computing. it's whole web based mostly technology wherever the resources and knowledge shared on a distributed network, thus it's vital for each supplier also as shoppers to produce the safety and trust to shared .the data for developing cloud computing application. as a result of organization ar currently moving quick towards the cloud thus there's an occasion of threats that will harm the data on the cloud.

Sunanda Ravindran and Parsi KalpanaImproving [5] Cloud Security Using knowledge Mining: they need explicit that, this method propose and economical distributed design to mitigate the risks. New attacks square measure being discovered each day and new step got to be develop to stay knowledge secure. Attackers and suppliers economical data processing technique to extract info regarding the user from the information hold on in cloud. With increase in sharing of knowledge over internet there's a rise in risk of knowledge being subjected to malicious attacks .Attackers /provider will extract sensitive info by analyzing the shopper knowledge over a protracted amount of your time. thence the privacy and security of user data is compromised.

Hemalatha and Dr. R. Manickachezian [6] Cloud Computing security- Trends and analysis Directions: they need explicit that, it take a holistic read of cloud computing security-spanning across doable issue and vulnerabilities connected with virtualization infrastructure software package platform; identity management and access control; information integrity ; confidentiality and privacy; physical and method security aspects; and legal compliance in cloud. Cloud computing as a platform for outsourcing and remote process of application and information is gaining fast momentum. Security concern; especially those around platform, data and access; can prove to be hurdles for adoption public and hybrid cloud.

System [7] provides secured mining of data in cloud exploitation "homomorphic encoding" encryption technique. Here planned approach perform k-means cluster algorithmic rule of an information set, that is divided horizontally exploitation "HDFS" and save 2 nodes. this technique 1st run regionally so implement "k-means" for combined knowledge on encrypted result to get complete(final) result.

System [8] propose security ensurement both in public and conjointly personal cloud. The projected system is employed to send knowledge to the cloud suppliers. so sanctionative of cloud computing bourgeois is finished to perform operations on the info as per user decision, like analyzing sales patterns, on the far side exposing the important knowledge. this could be achieved by cryptosystems supported "Homomorphic Encryption". 3 variety of homomorphic encryptions ar potential.1)partial HE, 2) somewhat HE, 3) absolutely HE. Authors advocate to boost HE algorithm's quality and latency to requests gets calculated according to public key length.

In paper [9] they are provide data security in cloud is terminated. Real service supplier Company, for instance Google, Yahoo, Microsoft have behind supplemental secret writing to end-to-end data facilitating and administration for shoppers. as an example, Google Cloud Storage currently consequently scrambles each single new information written on disk. As indicated by specialists, measures area unit essential for securing the event of data between the client organizations and therefore the suppliers of cloud services. Ordered archives from the United States intelligence agency demonstrate that they are attempting to debilitate encryption calculations in general utilized by people.

System [10] proposed different Homomorphic Encryption systems (Partially, Somewhat and Fully Homomorphic Encryption), the few challenges resulting of using this idea and new Client-Cloud supplier design, which will improve the performance of this system. Writers suggested to primarily target the analysis expand of the quality of existing HE algorithms by sanctioning cloud server for operational numerous operations by the purchasers

System [11] the study of information security mistreatment RSA rule. Cloud computing is that the idea enforced to unravel the Daily Computing issues, conduct of Hardware computer code and Resource availableness unhurried by laptop client. The distributed computing offers AN lenient and

non-incapable answer for cyclic Computing. the conventional drawback combined with Cloud Computing is that the Cloud security and also the correct Implementation of Cloud over the Network.

III. RESEARCH METHODOLOGY

The Proposed system improves security and data integrity within the cloud setting. It guess that the user's information isn't hold on in centralized location however is distributed location and performs a combined k-means agglomeration formula. The projected system may be a hybrid of antecedently studied systems wherever the protection are going to be provided exploitation primarily 2 cryptography techniques. Digital Signature is employed to supply believability of a message or method that may be a one sort of mathematical theme. If the given digital signature is correct then that may offer detail that the message is made by notable sender and also the information residing therein isn't altered. Kmeans agglomeration is employed to mine from given information facet in conjunction with beside at the side together with that it maintains the privacy of knowledge from each the side to forestall leak of intermediate result by associate wrongdoer. it'll offer input and final outputs to the host and user does not have to bother about intermediate results.

Homomorphic Encryption: Homomorphic encoding, it's the conversion of information into cipher text which will be analyzed and worked with as if it were still in its real type [4]. Homomorphic encryptions enable sophisticated mathematical operations to be performed on encrypted data without compromising the encryption.

IV. PROPOSED SCHEME

In mathematics, homomorphic define the transformation of 1 information set into another information set whereas conserving relationships between parts in each information sets. The term springs from the Greek words for "same structure." as a result of the information during a homomorphic cryptography theme retains the equivalent structure, identical mathematical operations whether or not they area unit performed on encrypted or decrypted information can yield equivalent results. Homomorphic cryptography is predicted to play a very important half in cloud computing, permitting corporations to store encrypted information during a public cloud and profit of the cloud provider's analytic services. Here may be a terribly plain example of however a homomorphic cryptography theme may add the cloud computing:

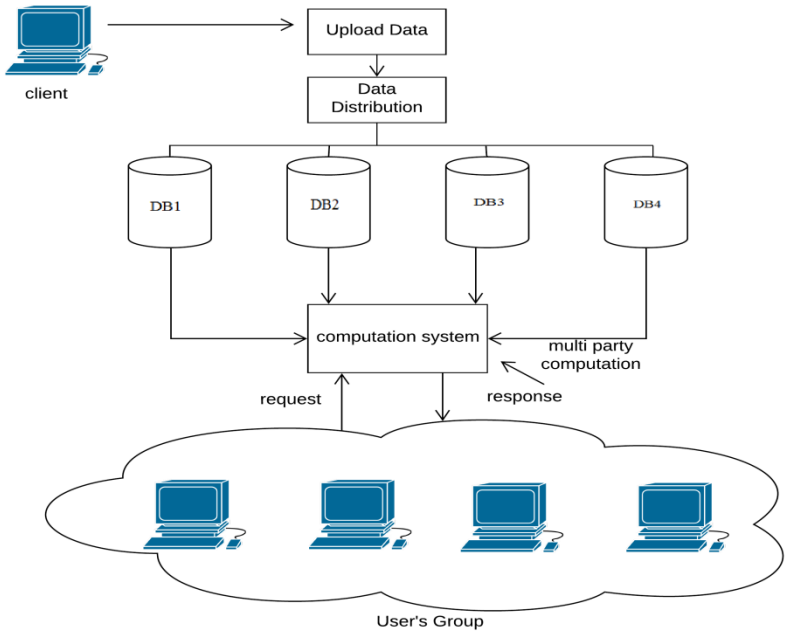

Fig.1: Proposed System Architecture

1) University basic principle encompasses a vital knowledge set which consists of the numbers ten and twenty. currently to encipher those knowledge set, university basic principle multiplies every component within the set through four. when with success finished this operation created new set whose members ar forty and eighty.

2) University basic principle sends the encrypted vital knowledge set to the cloud for secure knowledge storage. a

couple of months or years later, the govt. contacts university basic principle and requests the total of vital knowledge set parts.

3) University basic principle is simply too a lot of busy at that point, thus it asks the cloud supplier to perform the and operation. The cloud supplier, WHO solely has access to the encrypted knowledge set, realize the total of forty + eighty so recovery the solution a hundred and twenty.

4) University basic principle decrypted that cloud provider's reply (decrypt that result with divide by 4) and provides the govt. with the decrypted last final answer thirty.

There are sorts of homomorphic encryption: absolutely Homomorphic encoding (FHE) and Somewhat Homomorphic encoding (SHE) and partial HE. every kind differs within the variety of operations that may be performed on encrypted knowledge. FHE permits for a limiteless, discretional variety of computations (both addition and multiplication, usually minus not possible) to be performed on encrypted knowledge. SHE cryptosystems support a restricted variety of operations (i.e., any quantity of addition, however only 1 multiplication) and ar quicker and a lot of compact than FHE cryptosystems.

The system provides concurrent support for text as well as numeric dataset, To improve the efficiency of the generalization operation, propose a data structure, called Taxonomy Encoded Anonymity (TEA) index for QID = D1, . . . , Dm. TEA is a tree of m levels. The ith level represents the current value for Dj . Each root-to-leaf path represents a qid value in the current data table, with a (qid) stored at the leaf node. the TEA index links up the qids according to the generalizations that generalize them. When a generalization g is applied, the TEA index is updated by adjusting the qids linked to the generalization of g. The purpose of this index is to prune the number of candidate generalizations to no more than |QID| at each iteration, where |QID| is the number of attributes in QID. For a generalization g : child(v) → v, a segment of g is a maximal set of sibling nodes, {s1, . . . , st}, such that {s1, . . . , st} & child(v), where t is the size of the segment. All segments of g are linked up. A qid is generalized by a segment if the qid contains a value in the segment. A segment of g represents a set of sibling nodes in the TEA.

The below procedure we follow when deals with heterogeneous dataset.

Fingerprint Encoding:

Tardos Code Generation:

Fingerprint is created by using Tardos Code (fc). These bits are given as input to Fingerprint insertion process.

Data Partitioning:

The Database NDB is partitioned into m non-overlapping partitions by using secret key (Ks) concatenated with cryptographic secure hash function H().

Subset Selection:

In this process, few tuples are selected for fingerprinting to minimize the distortions.

Fingerprint Insertion:

Fingerprint bits are embedded in the selected tuple of each partition by using fingerprinting function.

Fingerprint Decoding:

It is the process of capturing the embedded fingerprint codeword from pirated copy (PDB) using secret key Ks, buyer ID and primary key. After extracting embedded fingerprint Traitor tracing algorithm is used to find out guilty user (Gu).

Data Partitioning:

Same data partitioning algorithm is used to partition the data as used in the fingerprint encoding phase. Marked Rows Identification: Fingerprinted rows are identified by the same procedure used while inserting fingerprint in encoding phase.

Fingerprint Detection:

As of mentioned decoding does not violate the requirement blind decoding. The decoding algorithm decodes the inserted fingerprint. Only the modifications are taken into consideration.

Traitor Tracing:

Captured fingerprints are the input to the traitor tracing process. Using captured fingerprints, one can detect guilty user by comparing captured fingerprint to each buyers fingerprint.

Mathematical Model

Here overall system set as s= {s1,s2,s3….sn} Here S1 provide Data provider module, S2 denote the load balancing and fingerprint insertion. S3 execute Data privacy mechanism. S4 access control and s5 denote the attack as well as analysis phase.

Now,

S1={D1,D2…..Dn} set of documents

Rec={AA1,AA2….AAn} select each attribute

Access={D1a, D2a….. Dna}

Ins={AA1(a*)…….Aan(a*)} // adding fingerprint to

Ext={AA1(e*)…. AAn(e*)}

P= {DBp1,DBp2…….DBpn}

// database i.e data provided by providers

// Apply F on it.

F = {Fingerprint adding(FA), slicing algorithm(SA), binary algorithm(BA), privacy verification algorithm(PA) }

T* = {Ruˆ DBpn}

// collaborative data according to user request and database which we have. F provides privacy and security to input data.

e = output in table format according to user authentication.

Success condition,

Ru[i] ≠ NULL, DBpn ≠ NULL

Failure condition,

Ru[i] = = NULL, DBpn = = NULL

V. PERFORMANCE EVALUATION

The For the system performance evaluation, calculate the matrices for accuracy. The system is executed on java 3-tier architecture framework with INTEL 2.8 GHz i3 processor and 4 GB RAM with public cloud Amazon EC2 consol. For the system evaluation we create 2 machines on physical environment with distributed environment. After implementing some part of system we got system performance on reasonable level. The below table 1 shows the proposed homorphic algorithm performance for user plain text conversion as well encryption decryption.

Table 1: System performance

Data Size in MB	Encryption time (Milliseconds)		Decryption time (Milliseconds)	
	Existing (AES)	Proposed	Existing (AES)	Proposed
5	595	515	724	612
10	1120	1026	1132	1033
15	1680	1547	1687	1556
20	2260	2064	2231	2033

In second experimentation system show the user verification time with different approaches. In current system we consider as four different authorities for runtime verification. The below Fig. 3 shows the performance measures using different parameters with some existing approaches.

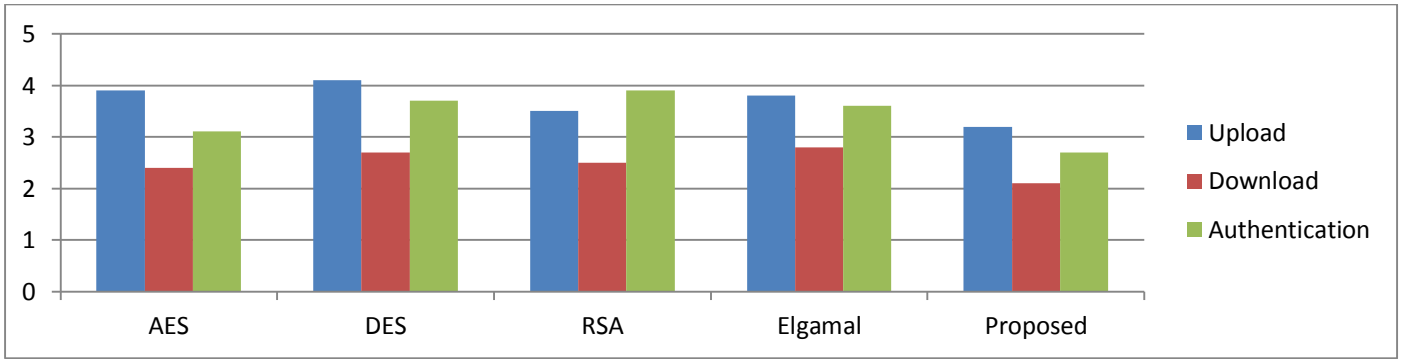

Fig.3: System Performance Measures proposed vs Existing approaches

VI. CONCLUSION

The proposed system describes a data partition using HECA including some numeric as well as some categorical attributes using big data**.** Security and privacy is the major issue concerning the clients as well as the providers of cloud services as a lot of confidential and sensitive data is stored in cloud which can provide valuable information to an attacker. This paper proposes a method to solve the privacy issues of the cloud. It assumes that the user data is distributed on two hosts and performs a combined k-means clustering using the Pallier Homomorphic encryption system for security purpose therefore on forestall Associate in Nursingy interpretation of intermediate results by an offender. The projected approach will any be extended by adding a digital signature or hashing technique to demonstrate the third party therefore on forestall Associate in Nursing someone from movement because the third party to host's. conjointly it will be generalized or extended to a lot of range of hosts if needed.

VII. REFERENCES

[1]. Mr. V. Biksham, Dr. D. Vasumathi, "Security and Privacy in Cloud Computing", International Journal ofEngineering Research and Applications (IJERA) ISSN: 2248-9622 NATIONAL CONFERENCE onDevelopments, Advances & Trends in Engineering Sciences

[2]. S. SelvaRatna , Dr. T. Karthikeyan, "Survey on recent algorithms for privacy preserving data mining",S.SelvaRathna et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol.6 (2) , 2015, 1835-1840.

[3]. ShashankBajpai, PadmijaShrivastava, "Data Mining Approach in Security Information and Event Management", International Journal of Information & Computation Technology, ISSN 0974-2239 Volume 4, Number 8.

[4]. Deepti Mittal, Damandeep Kaur, Ashish Aggarwal, "Security threat issues and countermeasures in cloudcomputing". IEEE Cloud Security.

[5]. Sunanda Ravindran, Parsi Kalpana, "Improving Cloud Security Using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4.
[6]. Hemalatha ,Dr. R. Manickachezian, "Cloud Computing security- Trends and research Directions", International Journal of Advanced Research in Computer and Communication Engineering ,Vol. 3, Issue 11,
[7]. Deepti Mittal, Damandeep Kaur, Ashish Aggarwal. " Secure Data Mining in Cloud using Homomorphic Encryption" 2014 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM).
[8]. Shashank Bajpai andPadmija Srivastava. "A Fully Homomorphic Encryption implementation on Cloud Computing" International Journal of Information & Computation Technology. ISSN 2014.
[9]. Aws Naser Jaber1, Mohamad Fadli Bin Zolkipli2. " A Study in Data Security in Cloud Computing" International Conference on Computer, Communication, and Control Technology 2014.
[10]. Khalid EL MAKKAOUI, Abdellah EZZATI. "Challenges of Using Homomorphic Encryption to Secure Cloud Computing" 2015 International Conference on Cloud Technologies and Applications (CloudTech). Uma Somani,Kanika Lakhani,Manish Mundra. " Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing" 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC – 2010s).
[11]. M. TEBAA and S. EL HAJII. "Secure Cloud Computing through Homomorphic Encryption, " International Journal of Advancements in Computing Technology (IJACT), Vol.5, No.16, pp. 29 –38.