AN EFFICIENT SECURITY MODEL FOR PRIVACY ENCHANCED DISTRIBUTED DATA MINING USING CLOUD FEDARATION APPROACH

Sukumar[1], Ajitha[2] , Akkshaya[3] , Janupriya[4]
Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, India.

ABSTRACT

Cloud computing is an alternative to conventional IT Outsourcing. As a result, Cloud computing migration between organizations is rapidly growing. The existing security model provides security to sensitive data through third party and that has levels of security and it work for various kinds of dataset. The main drawback is the security of information may be compromised from third party. The proposed system proceeded using RASP model along with classification, clustering, decision tree and feature selection for processes. With the help of cloud federation approach, it stores the information safely without third party access. Then the user gets the information from the cloud using KNN query which serves as a path between sender and the receiver. The dataset transformations are effective by using linear algorithm Receiver RASP model.

KEYWORDS:RASP, KNN, cloud federation, linear.

l.INTRODUCTION

Computing is the on-demand availability of computer system resources, especially data storage and computing power, withoutdirect active management by the user. The term is generally used to describe data centres available to many users over the Internet. Large clouds, predominant today, often have functions distributed over multiple locations from central servers. If the connection to the user is relatively close, it may be designated an edge server. Clouds may be limited to a single organization (enterprise cloud), or be available to multiple organizations. Cloud computing relies on sharing of resources to achieve coherence and economies of scale. Advocates of public and hybrid clouds note that cloud computing allows companies to avoid or minimize up-front IT infrastructure costs. Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and that it enables IT teams to more

rapidly adjust resources to meet fluctuating and unpredictable demand, providing the burst computing capability, high computing power at certain periods of peak demand. Cloud providers typically use a "pay-as you-go" model, which can lead to unexpected operating expenses if administrators are not familiarized with cloud-pricing models. The availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture and autonomic and utility computing has led to growth in cloud computing. By 2019, Linux was the most widely used operating system, including in Microsoft's offerings and is thus described as dominant.

Cloud federation is the practice of interconnecting the cloud computing environments of two or more service providers for the purpose of load balancing traffic and accommodating spikes in demand. Cloud federation requires one provider to wholesale or rent computing resources to another cloud provider. Those resources become a temporary or permanent 2 extension of the buyer's cloud computing environment, depending on the specific federation agreement between providers. Cloud federation offers two substantial benefits to cloud providers. First, it allows providers to earn revenue from computing resources that would otherwise be idle or underutilized. Second, cloud federation enables cloud providers to expand their geographic footprints and accommodate sudden spikes in demand without having to build new points-ofpresence. Service providers strive to make all aspects of cloud federation—from cloud provisioning to billing support systems (BSS) and customer support— transparent to customers. When federating cloud services with a partner, cloud providers will also establish extensions of their customer-facing service-level agreements (SLAs) into their partner provider's data centres.

II.LITERATURE SURVEY:

Many applications that require the ability to manipulate the sequences of data and motivate the importance of sequence query processing, and present a framework for the optimization of sequence queries based on several novel techniques. These include query transformations, optimizations that utilize meta–data, and caching of intermediate results.Many real life applications manipulate data that is inherently sequential. Such data is logically viewed and queried in terms of a sequence abstraction and is often physically stored as a sequence. Databases should allow sequences to be queried in a declarative manner, utilizing the ordered semantics of the data, and take advantage of the opportunities available for query optimization. Consequently, expressing sequence queries is tedious, and evaluating them is

inefficient. Sequence databases therefore require techniques that are distinct from established relational database techniques.

The implementation of SEQ, a databases system to support or sequence data. SEQmodels have n ordered collection of records, and supports declaratives sequence query language. SEQ has been built as a component of the PREDATOR database system that provides support for relational and other kinds of complex data as well.SEQ is a component of the PREDATOR multi-threaded, client-server database system which supports sequences, as well as relations and other kinds of complex data. The system uses the SHORE storage manager library for low-level database functionality like buffer management, concurrency control and recovery. A novel design paradigm provides query processing support for multiple data types, including both sequences and relations Much real-life information contains logical ordering relationships between data items.Traditional relational databases provide no abstraction of ordering in the data model, and do not support queries based on logical sequentially in the data

The need for and research issues arising from a new model of data processing. In this model, data does not take the form of persistent relations, but rather arrives in multiple, continuous, rapid, time varying data streams. In addition to reviewing past work relevant to data stream systems and current projects in the area, the paper explores topics in stream query languages, new requirements and challenges in query processing, and algorithmic issues. Recently a new class of data-intensive applications has become widely recognized applications in which the data is modeled best not as persistent relations but rather as transient data streams. Examples of such applications include financial applications, network monitoring, security, telecommunications data management, web applications, manufacturing, sensor networks. In the data stream model, individual data items may be relational tuples, e.g., network measurements, call records, web page visits, sensor readings.

III.RESEARCH METHOLOGY:

PROPOSED METHODOLOGY

There are n-parties holding their secret data, and each party wants to enhance its data mining results by using the data of all the parties.

The data are safe in privacy distributed environment under Random Space Perturbation. In RASP the use of matrix multiplication does not protect the dimensional values so no need to go through from the distribution-based attack. RASP does not preserve the distances between

records, so it prevents the data that are perturbed from distance-based attacks. And also, it won't protect more difficult structures it may be a matrix and other components. The range queries can be sent to the RASP perturbed data and this range query describe open bounds in the multidimensional space. In Random space perturbation, the perturbation is used to do collapsing this process will take place according to the key value that is specified by the owner. In this module the data owner has to register like owner and have to provide owner name as well as key value.

RASP has some important features such as,used for safe handling for privacy distributed data. The calculation of energy consumed will calculated and displays. Safe upload and retrieval of the information done by RASP. Speed up range query processing. The output which sends to the receiver will calculate accurately. Using cloud federation approach, we provide security without third party access. Cloud federation is the practise ofinterconnecting the cloud computing environments of two or more service providers for the purpose of load balancing traffic and accommodating spikes in demand. Using cloud federation approach, we provide security without third party access and secures the data confidentiality. Also, Increased security and control.

Reduction in IT costs to support growing federation infrastructure. Eliminates federated application deployments KNN- query denotes k-Nearest Neighbour query. This query is usually used to retrieve the nearest neighbour values of k. Here k is used to denote positive integer value. KNN algorithm is mainly used for classification and regression. use of kNN algorithm is to process the range query to kNN query. This algorithm consists of two methods. That is used to make interaction between the client and the server. The client will send the query to the server with initial upper bound and lower bound.

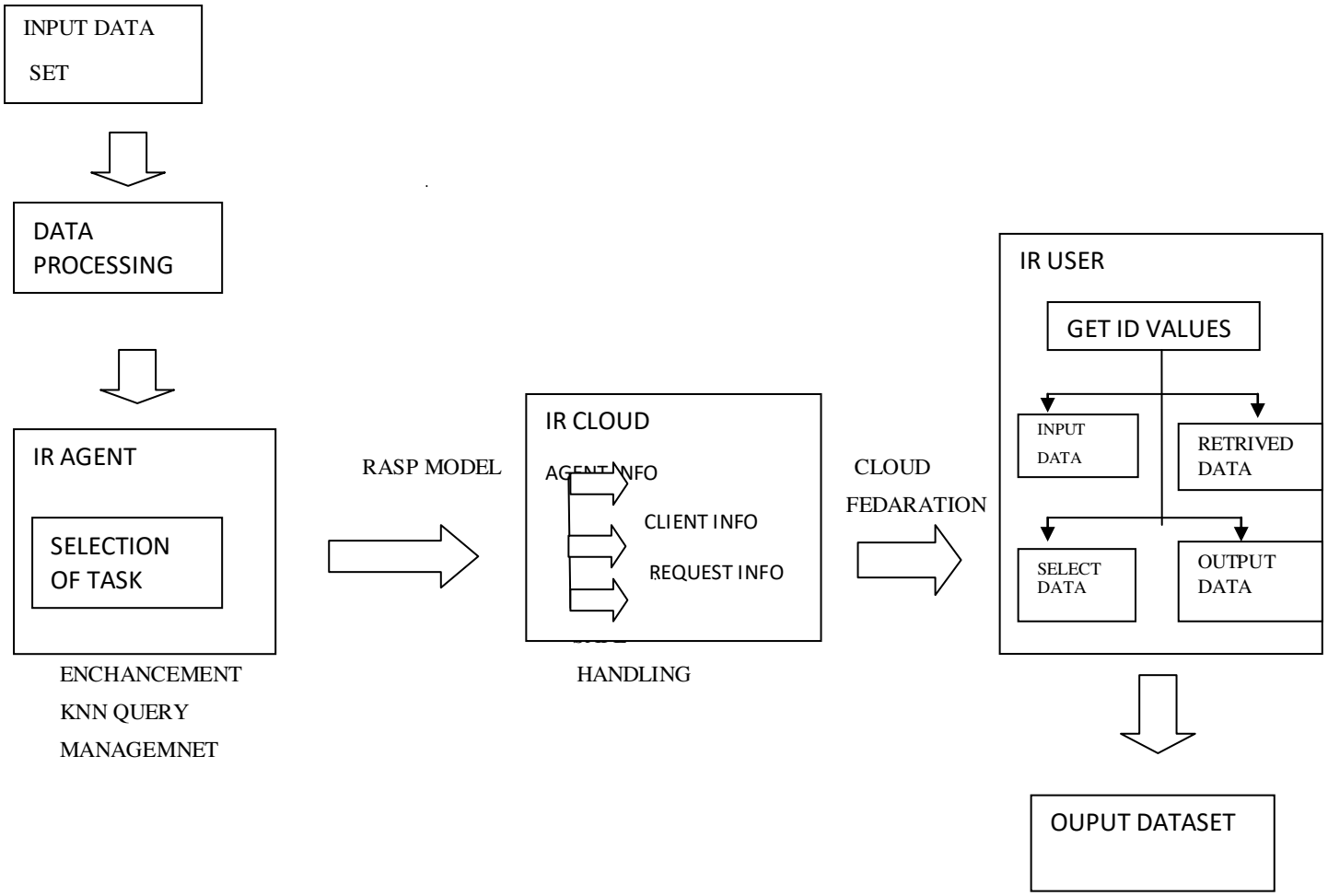
IV.THEORY AND CALCULATIONS:

Based on the RASP perturbation method, the services for two types of queries: range query and k-Mean query. This section will dedicate to range query processing. The range query in the original space can be transformed to a polyhedron query in the perturbed space, and then to develop a secure way to do the query transformation. Then develop a two-stage query processing strategy for efficient range query processing to maximize their throughput. The maximum user keyword search results, not shown here, followed a similar pattern.

Three datasets are used in experiments. (1) A synthetic dataset that draws samples from uniform distribution in the range [0, 1]. (2) The Adult dataset from UCI machine learning

dataset. then assign a numeric value to the categorical values using a simple one-to-one mapping scheme. (3) The 2-dimensional North East location data from rtreeportal.org. Many secure approaches cannot use indices for query processing, which results in poor performance. For example, the secure dot-product approach encodes the points with random projections and recovers dot-products in query processing for distance comparison. The way of encoding data disallows the index-based query processing. Without the aid of indices, processing a k-Means query will have to scan the entire database, leaving many optimizations impossible to implement.

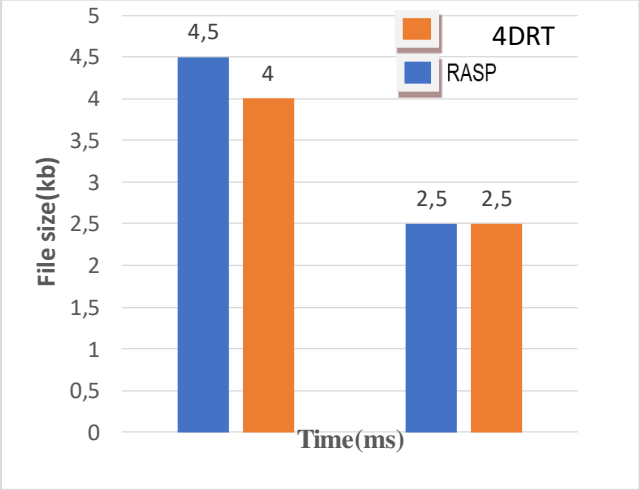
V.FLOW DIAGRAM:



VI.COMPARISION CHART:

comparison between existing and proposed algorithm

Algorithm	Over all Data Handling Accuracy in percentage	Data integrity per block(for 100 percentage)
Existing	78	93
Proposed	83	98



ADVANTAGE OF PROPOSED SYSTEM:

- A secure multiparty computation (SMC) scheme will have multiple parties in a distributed environment.
- The cloud federation approach secures the data confidentiality
- By using four dimensional rotation transformation algorithm enhances security and robustness of data
- The data transform are effective by using linear algorithm Receiver RASP model

VII.RESULT AND DISCUSSION:

Cloud-federation algorithms perform in terms of different settings that may also involve the tradeoff between model quality and client-side costs. Specifically, it will show the scalability of our approach with client-side computation and communication costs on real datasets. Experiments to understand which of the four private learning methods is the best in terms of costs and model quality and evaluate model confidentiality with the proposed method.

Dataset	Records	Dimensions	Link
German Credit	1000	20	https://goo.gl/lVy34O
Ozone Days	2536	73	https://goo.gl/Si6aDh
Spambase	4601	57	https://goo.gl/WPyXTi
Bank Marketing	45211	17	https://goo.gl/vvgj3M
Twitter Buzz	140000	77	https://goo.gl/Yfy80u

TABLE 1

Datasets for experiments.

VIII.CONCLUSION:

Cloud mining should address four aspects: data confidentiality, model confidentiality, model quality, and low client-side costs. The RASP perturbation to guarantee the data confidentiality. The RASP-B framework aims to provide practical confidential classifier learning with the cloud service provider (cloud federation). To develop the boosting-based RASP-Boost framework to obtain high-quality classifiers with these non-optimal linear classifiers. Client provides a set of encoded base classifiers, and the cloud computes a boosting model from the set. This pattern does not require the client to stay online during boosting iterations, which is convenient for the client. The confidentiality of data, query, learning process and models is formally analysed, and we show the confidentiality of data and query is satisfactorily guaranteed. In have conducted an extensive evaluation of the proposed algorithms and studied the effect of the major factors in the RASP-Boost framework. Derived DS can generate high-quality models with accuracy very close to the optimal boosting models. To evaluate the concept of model confidentiality for real data and

models and show that the model confidentiality is well preserved under the security assumption.

REFERENCE:

1. P. Seshadri, M. Livny, and R. Ramakrishnan, "Succession question processing,"in SIGMOD, 1994.
2. P. Seshadri, M. Livny, and R. Ramakrishnan, "The plan and execution of an arrangement information base framework," in VLDB, 1996.
3. B. Babcock, et al., "Models and issues in information stream frameworks," in PODS, 2002.
4. J. Chen, et al., "NiagaraCQ: a versatile nonstop question framework for web information bases," SIGMOD Rec., 2000.
5. F. Wang, et al., "Fleeting administration of RFID information," in VLDB, 2005.
6. H. Gonzalez, J. Han, X. Li, and D. Klabjan, "Warehousing and Analyzing Massive RFID Data Sets," in ICDE, 2006.
7. H. Gonzalez, J. Han, and X. Li, "FlowCube: Constructing RFID FlowCubes for Multi-Dimensional Analysis of Commodity Flows," in VLDB, 2006.
8. E. Lo, B. Kao, W.- S. Ho, C.- K. Chui, and D. Cheung, "OLAP on grouping information," in SIGMOD, 2008, pp. 649–660.
9. Z. He, P. Wong, B. Kao, E. Lo, and R. Cheng, "Quick assessment of chunk of ice design based total inquiries," in CIKM, 2013, pp. 2219–2224.
10. K. Beyer, P. J. Haas, B. Reinwald, Y. Sismanis, and R. Gemulla, "On outlines for particular worth assessment under multiset tasks," in SIGMOD, 2007, pp. 199–210.