TECHNIQUES FOR DEALING WITH CONCEPT DRIFT IN PROCESS MINING

Avinash[1] , N. Sukhendar Reddy[2], Maraty Meena[3]
[1]*M.Tech student, CSE Dept, Vishnu sree Institute Of Science and Technology,Telangana, India.*
[2]*Asst.Professor, CSE Dept Vishnu sree Institute Of Science and Technology, Telangana, India .*
[3]*Assistant professor ©, CSE, University College of Technology, O.U., (Autonomous), HYD, Telangana.,India.*

Abstract- **The projects defines and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system wise privacy in the cloud computing paradigm. Data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. Among various multi-keyword semantics, choosing the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query.**

Index Terms- **- Cloud computing, searchable encryption, privacy preserving, keyword search, ranked search Anonymization.**

I. INTRODUCTION

Process Mining is a relatively novel discipline which has received a lot of attention in the last decade [1]. Although it shares many features with Data Mining, it has originated from different concerns and communities, has a set of distinctive techniques, and produces slightly different outcomes. Historically, process mining arises from the observation that many organizations record their activities into logs which describe, among others, the real ordering of activities of a given process, in a particular implementation. Software engineering techniques have mainly focused on the specification part of the processes within an information system. In reality, this may cause a big gap between a system specification's and the final implementation, hampering the use of the models that specify the main processes of an information system. As another example, designers of hardware or embedded, concurrent systems, need to compare behavior and specifications, typically they can passively or actively generate large amounts of logs from their target system and/or their prototypes, so a logical approach is to use these logs for the verification task.

Considering potentially huge number of on-demand data users and large amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability. Document ranking is provided for fast search, but the priorities of all the data documents is kept same so that the cloud service provider and third party remains unaware of the important documents, thus, maintaining privacy of data. Besides, to improve search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keyword search, as single keyword search often yields far too coarse results.

.

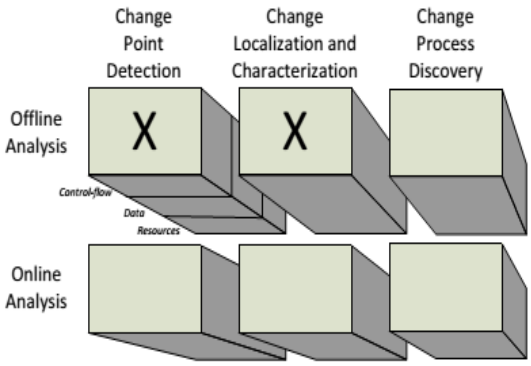


Fig. 1. Different dimensions of concept drift analysis in process mining.

As a common practice indicated by today's web search engines (ex. Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data. We can differentiate between two broad classes of dealing with concept drifts when analyzing event logs (Fig. 1).

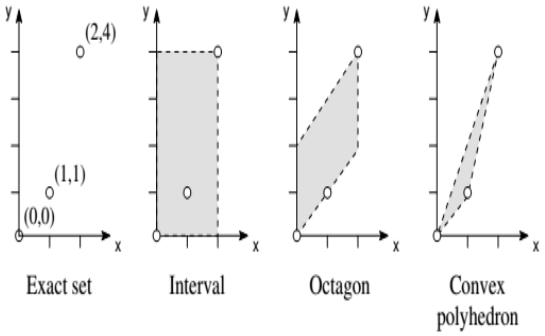
Offline analysis: This refers to the scenario where the presence of changes or the occurrence of drifts need not be uncovered in a real time. For example, offline concept drift analysis can be used to better deal with seasonal effects (hiring less staff in summer or skipping checks in the weeks before Christmas).

Online analysis: This refers to the scenario where changes need to be discovered in near real time. Such real-time triggers (alarms) will enable organizations to take quick remedial actions and avoid any repercussions.

II. BACKGROUND

Abstract Interpretation Intuitively, abstract interpretation defines a procedure to compute an upper approximation for a given behavior of a system that still suffices for reasoning about the behavior itself. An important decision is the choice of the kind of upper approximation to be used, which is called the abstract domain. For a given problem, there are typically several abstract domains available. Each abstract domain provides a different trade-off between precision (closeness to the exact result) and computational efficiency.

There are many problems where abstract interpretation can be applied, several of them oriented towards the compile-time detection of run-time errors in software. For example, some analysis based on abstract interpretation can discover numeric invariants among the variables of a program. Several abstract domains can be used to describe the invariants: intervals [2], octagons [3], and convex polyhedra [4], among others. These abstract domains provide different ways to approximate sets of values of numeric variables. For example, Figure 2 shows how these abstract domains can represent the set of values of a pair of variables x and y. For space reasons, we focus on the abstract domain of convex polyhedra. In the experiments, the domain of octagons is also used.



Interval	Octagon	Convex polyhedron
$0 \le x \le 2$ $0 \le y \le 4$	$0 \le x \le 2$ $0 \le y - x \le 2$	$y - x \ge 0$ $x - 2y \ge 0$ $3x - y \le 2$

Fig. 2. Approximating a set of values (left) with several abstract domains

In [5] and [6] collections of typical change patterns are described. In [7] and [8] extensive taxonomies of the various flexibility approaches and mechanisms are provided. Ploesser et al. [9] have classified business process changes into three broad categories: 1) sudden; 2) anticipatory; and 3) evolutionary. This classification is used in this paper, but now in the context of event logs.

A notable exception is the approach in [33]. This approach uses process mining to provide an aggregated overview of all changes that have happened so far. Recently, Carmona and Gavaldà [11] have proposed an online technique for detecting process changes. They first created an abstract representation of the process in the form of polyhedra

using the prefixes of some initial traces in the event log.

III. PROPOSED SYSTEM

In the proposed work, we will explore checking the integrity of the rank order in the search result assuming the cloud server is untrusted. To propose OTP (one Time Password) as our future work. This OTP used to see data in cloud and it can be used once only in a time, when you search a file and tend to see the file the OTP will send to email and you get the OTP and apply to see the file.

System Architecture

A. Data Flow Diagram

The framework identifies the following steps:

Feature extraction and selection: This step pertains in defining the characteristics of the traces in an event log. Depending on the focus of analysis, we may define additional features, e.g., if we are interested in analyzing changes in organizational/resource perspective, we may consider features derived from social networks as a means of characterizing the event log.

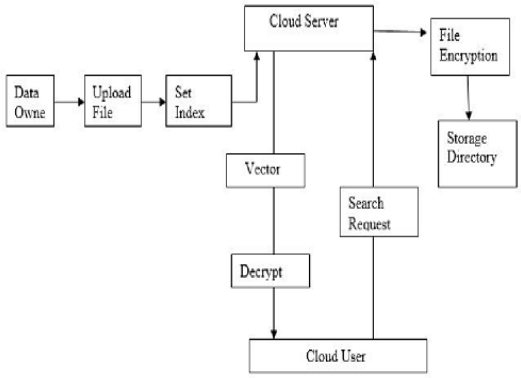


Fig.2 Architecture diagram of the MRSE Implementation.

Generate populations: An event log can be transformed into a data stream based on the features selected in the previous step. This step deals with defining the sample populations for studying the changes in the characteristics of traces.

Compare populations: Once the sample populations are generated, the next step is to analyze these populations for any change in characteristics.

The following modules are implemented in this technique

a. Cloud Setup

b. Cryptography cloud Storage

c. Vector Model

Cloud Setup

In this module we have setup data owner and cloud server. So the data owner is going push the data into the cloud sever. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud will be secured

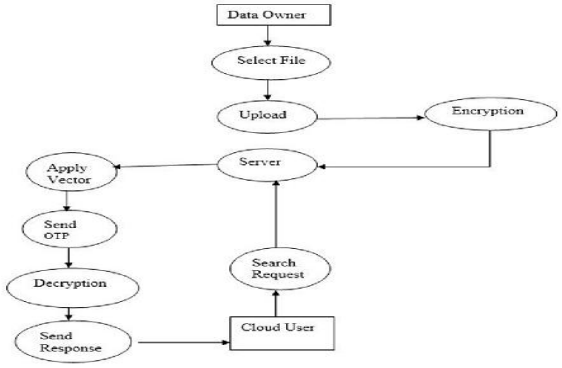


Fig.3 Data Flow Diagram of MRSE Implementation

Cryptography cloud Storage

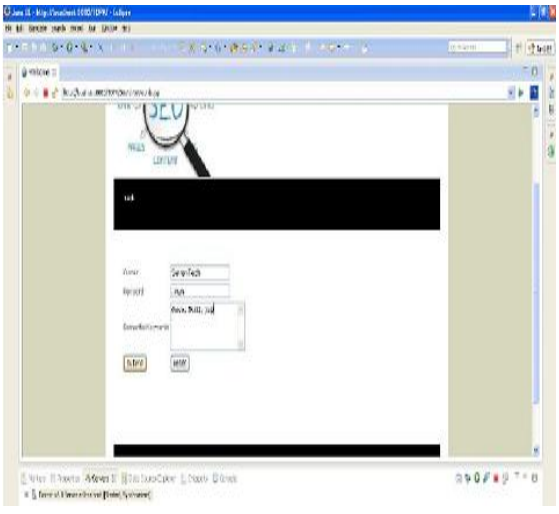
In this module while the data is uploaded into the Estorage and retrieve services. Since data may contain sensitive information, the cloud servers cannot be fully entrusted in protecting data. For this reason, outsourced files must be encrypted. Any kind of information leakage that would affect data privacy are regarded as unacceptable

Vector Model

In this model we used a series of searchable symmetric encryption schemes have been enable search on cipher text. In the former, files are ranked only by the number of retrieved keywords, which impairs search accuracy

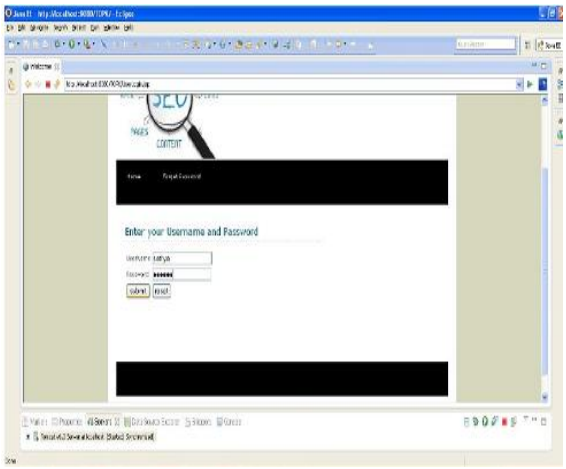
A. Home Page

B. Admin Login

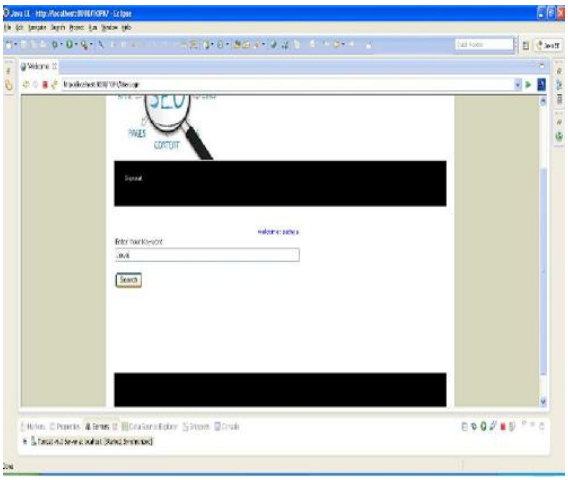


C.Upload the Multi keyword

D.File Upload



E. User login



F. Search Keyword

IV. CONCLUSION

Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we propose a basic

REFERENCES

[1]. P. Cousot and R. Cousot. Static determination of dynamic properties of programs. In 2nd Int. Symposium on Programming, pages 106–130. Paris, France, 1976.

[2]. P. Cousot and N. Halbwachs. Automatic discovery of linear restraints among variables of a program. In Proc. ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages, pages 84–97. ACM Press, New York, 1978.

[3]. A. Min´e. The octagon abstract domain. In Analysis, Slicing and Tranformation, IEEE, pages 310–319. IEEE CS Press, October 2001.

[4]. IEEE Task Force on Process Mining. Process mining manifesto. In Florian Daniel, Kamel Barkaoui, and Schahram Dustdar, editors, Business Process Management Workshops (1), volume 99 of Lecture Notes in Business Information Processing, pages 169–194. Springer, 2011.

[5]. N. Mulyar, "Patterns for process-aware information systems: An approach based on colored Petri nets," Ph.D. dissertation, Dept. Comput. Sci., Univ. Technol., Eindhoven, The Netherlands, 2009.

[6.] B. Weber, S. Rinderle, and M. Reichert, "Change patterns and change support features in process-aware information systems," in Proc. 19th Int., 2007, pp. 574–588.

[7]. G. Regev, P. Soffer, and R. Schmidt, "Taxonomy of flexibility in business processes," in Proc. 7th Workshop BPMDS, 2006, pp. 1–4.

[8]. H. Schonenberg, R. Mans, N. Russell, N. Mulyar, and W. M. P. van der Aalst, "Process flexibility: A survey of contemporary approaches," in Proc. Adv. Enterprise Eng. I, 2008, pp. 16–30.

[9]. K. Ploesser, J. C. Recker, and M. Rosemann, "Towards a classification and lifecycle of business process change," in Proc. BPMDS, vol. 8. 2008, pp. 1–9.

BIODATA

AUTHOR 1

Avinash currently pursing his M.Tech (Computer Science Engg) from VSIT, Bommalaramaram, Nalgonda,Telangana and India.

AUTHOR 2

Narra Sukhendar Reddy,has 6years experience in teaching and currently working as a Sr.Asst.Prof..in CSE Dept at VSIT,Nalgonda,Telangana,India.

AUTHOR 3

Maraty Meena received her B.Tech in ECE in Gokaraju Rangaraju institute of Technology and Sciences, in the year 2006 Miyapur, R.R.Dist, Telangana, and P.G. received in CSE IN JNTU-H, in the year 2011, Hyderabad, Telangana from .Currently working as a Assistant professor(c) in CSE Dept., U.C.T (A)–O.U Hyderabad Telangana.