International Journal of Engineering In Advanced Research Science and Technology ISSN: 2278-256

DEALING WITH CONCEPT DRIFTS IN PROCESS MINING USING SECURITY PRIMITIVES

Shaik. Rasheeda [1], ***Mr.K.Bhaskar*** [2]

[1]Department of Computer Science and Engineering
Gudlavalleru Enginnering College, Gudlavalleru
Email id: srasheeda9@gmail.com

[2]Assistant Professor
Department of Computer Science and Engineering
Gudlavalleru Enginnering College, Gudlavalleru
Email id: bhaskarjnvkp@gmail.com

Abstract

This paper is an attempt to enhance the existing Drift Detection with Change process discovery in complex Datasets. After detecting the change points and the regions of change, it is necessary to put them together in perspective. In addition, there are other applications such as deriving a configurable model for the process variants. A configurable process model describes a family of similar process models. The process variants discovered using concept drift can be merged to derive a configurable process model. A recent study revealed that different diversity levels in an ensemble of learning machines are required in order to maintain high generalization on both old and new concepts. Inspired by this study and based on a further study of diversity with different strategies to deal with drifts, we propose a new online ensemble learning approach called Diversity for Dealing with Drifts (DDD) and also explore how to make secure environments with less synchrony and show how it can be used to solve asynchronous Secure Multiparty Computation (SMC). Within the redesign we investigate the problem of solving consensus in a General omission failure model augmented with failure detectors.

Index Terms — Concept drift, flexibility, hypothesis tests, security, trust, process changes, process mining.

I. INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

While large-scale information technology has been evolving separate transaction and

analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- **Classes**: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

- **Clusters**: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

- **Associations**: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

- **Sequential patterns**: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

II. LITERATURE SURVEY

R. P. J. C. Bose [10]-"Handling concept drift in process mining", Operational processes need to change to adapt to changing circumstances, e.g., new legislation, extreme variations in supply and demand, seasonal effects, etc. While the topic of flexibility is well-researched in the BPM domain, contemporary process mining approaches assume the process to be in steady state. When discovering a process model from event logs, it is assumed that the process at the beginning of the recorded period is the same as the process at the end of the recorded period. Obviously, this is often not the case due to the phenomenon known as concept drift. While cases are being handled, the process itself may be changing. This paper presents an approach to analyze such second-order dynamics. The approach has been implemented in ProM and evaluated by analyzing an evolving process.

J. Carmona and R. Gavaldà [11] - "Online techniques for dealing with concept drift in process mining", Concept drift is an important concern for any data analysis scenario involving temporally ordered data. In the last decade Process mining arose as a discipline that uses the logs of information systems in order to mine, analyze and enhance the process dimension. There is very little work dealing with concept drift in process mining. In this paper we present the first online mechanism for detecting and managing concept drift, which is based on abstract interpretation and sequential sampling, together with recent learning techniques on data streams.

J. C. A. M. Buijs, B. F. van Dongen [12]- "Towards cross-organizational process mining in collections of process models and their executions",Variants of the same process may be encountered in different organizations, e.g., any municipality will have a process to handle building permits. New paradigms such as Software-as-a-Service (SaaS) and Cloud Computing stimulate organizations to share a BPM infrastructure. The shared infrastructure has to support many processes and their variants. Dealing with such large collections of similar process models for

multiple organizations is challenging. However, a shared BPM infrastructure also enables cross-organizational process mining. Since events are recorded in a unified way, it is possible to cross-correlate process models and the actual observed behavior in different organizations. This paper presents a novel approach to compare collections of process models and their events logs. The approach is used to compare processes in different Dutch municipalities.

J. J. C. L. Vogelaar, H. M. W. Verbeek[13]-"Comparing business processes to determine the feasibility of configurable models: A case study", Organizations are looking for ways to collaborate in the area of process management. Common practice until now is the (partial) standardization of processes. This has the main disadvantage that most organizations are forced to adapt their processes to adhere to the standard. In this paper we analyze and compare the actual processes of ten Dutch municipalities. Configurable process models provide a potential solution for the limitations of classical standardization processes as they contain all the behavior of individual models, while only needing one model. The question rises where the limits are though. It is obvious that one configurable model containing all models that exist is undesirable. But are company-wide configurable models feasible? And how about cross-organizational configurable models, should all partners be considered or just certain ones? In this paper we apply a similarity metric on individual models to determine means of answering questions in this area. This way we propose a new means of determining beforehand whether configurable models are feasible. Using the selected metric we can identify more desirable partners and processes before computing configurable process models.

H. Schoenberg, R. Mans, N. Russell, N. Mulyar[14]- "Process flexibility: A survey of contemporary approaches", Business processes provide a means of coordinating interactions between workers and organizations in a structured way. However the dynamic nature of the modern business environment means these processes are subject to a increasingly wide range of variations and must demonstrate flexible approaches to dealing with these variations if they are to remain viable. The challenge is to provide flexibility and offer process support at the same time.

Many approaches have been proposed in literature and some of these approaches have been implemented in flexible workflow management systems. However, a comprehensive overview of the various approaches has been missing. In this paper, we take a deeper look into the various ways in which flexibility can be achieved and we propose an extensive taxonomy of flexibility. This taxonomy is subsequently used to evaluate a selection of systems and to discuss how the various forms of flexibility fit together.

III EXISTING SYSTEM

- The process is stable and enough example traces have been recorded in the event log, it is possible to discover a high quality process model that can be used for performance analysis, compliance checking, and prediction.

- Unfortunately, most processes are not in steady-state. In today's dynamic marketplace, it is increasingly necessary for enterprises to streamline their processes so as to reduce costs and to improve performance.

DISADVANTAGES OF EXISTING SYSTEM:

- Characterization in an offline setting.
- Change point detection: To detect concept drift in processes, i.e., to detect that a process change has taken place.
- Change localization and characterization.
- Change process discovery: Having identified, localized, and characterized the changes, it is necessary to put all of these in perspective.

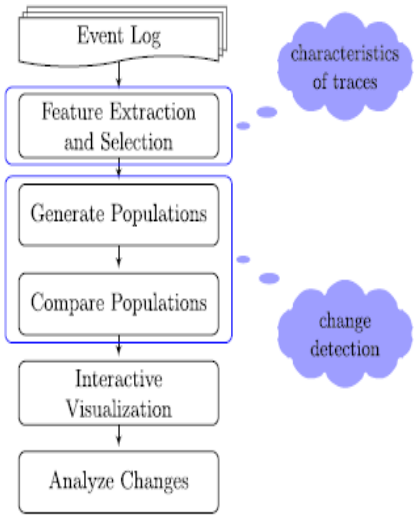

Fig 1: System Architecture

Algorithm:

Event logs are characterized by the relationships between activities. Dependencies between activities in an event log can be captured and expressed using the follows (or precedes) relationship, also referred to as causal footprints. For any pair of activities a, b ∈ *A*, and a trace **t** = **t**(1)**t**(2)**t**(3) . . . **t**(*n*) ∈ *A*+, we say b follows a if and only if for all $1 \leq i \leq n$ such that **t**(*i*) = a there exists a *j* such that $i < j \leq n$ and **t**(*j*) = b. In temporal logic notation: _(a ⇒ (♦b)). We say a precedes b if and only if for all $1 \leq j \leq n$ such that **t**(*j*) = b there exists an *i* such that $1 \leq i < j$ and **t**(*i*) = a,i.e., ¬a*W*b where *W* is the *weak until* in linear temporal logic notation.

The follows and precedes relationships can be lifted from traces to logs. If b follows a in all the traces in an event log, then we say that b *always follows* a. If b follows a only in some subset of the traces, then we say that b *sometimes follows* a. If b does not follow a in all traces,then we say that b *never follows* a. Consider an event log *L* = [acaebfh, ahijebd, aeghijk] containing three traces defined over *A* ={a, b, c, d, e, f, g, h, i, j, k}. The following relations hold in *L*: e always follows a, e never follows b, and b sometimes follows a.

IV PROPOSED SYSTEM:

- In this paper, we have introduced the topic of concept drift in process mining, i.e., analyzing process changes based on event logs.
- We proposed feature sets and techniques to effectively detect the changes in event logs and identify the regions of change in a process.

ADVANTAGES OF PROPOSED SYSTEM:

- Heterogeneity of cases arising because of process changes can be effectively dealt with by detecting concept drifts.

- Supporting or improving operational processes and to obtain an accurate insight on process executions at any instant of time.

Algorithm:

Bilinear Map/Pairing: Let G1, G2, and GT be cyclic groups of prime order p. Let g1 and g2 be generators of G1 and G2, respectively. A bilinear pairing is a map e : G1 × G2 →GT with the properties

1) Bilinear: ˆe(ua, vb) = ˆe(u, v)ab ∀ u ∈ G1, v ∈ G2,and a, b ∈ Zp

2) Non-Degenerate: ˆe(g1, g2) = 1

3) Computable: there exists an efficient algorithm for computing e

− H(·) is a map-to-point hash function : {0, 1}∗ →G1.

− EK is an encryption algorithm with strong diffusion property, e.g., AES.

Architecture:

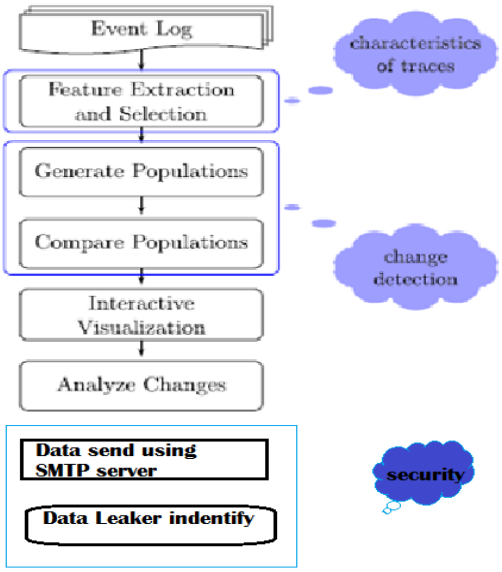

Fig 2: Proposed System Architecture

In the above architecture we are providing security and sending the data to the user by using SMTP server.After getting the data, user is successfully accessing the data from the SMTP server. If the proper data is not accessing by the user that user is called annonymous user the user will get the mail from the server.

V. Results – Sample screen shots:

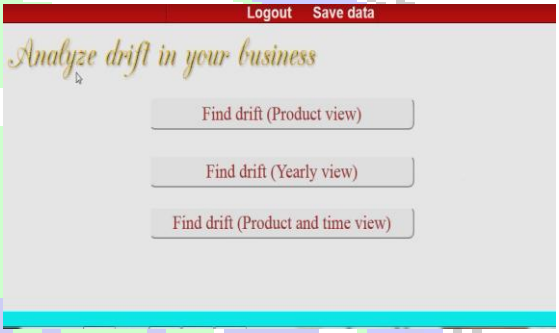

Fig 3: Different options of finding out drifts in offline mode.

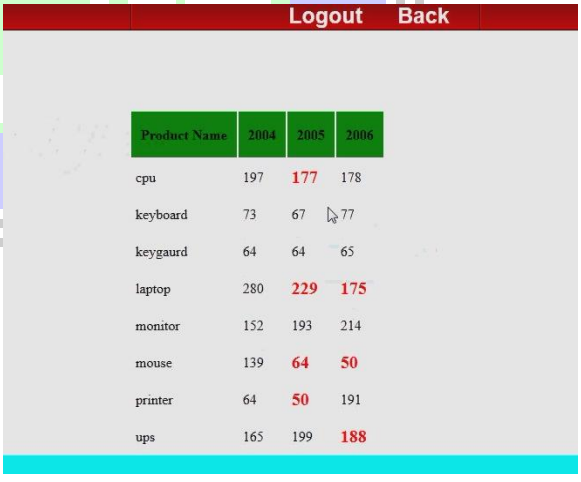

Fig 4: Analyzing of drifts in product and time view.

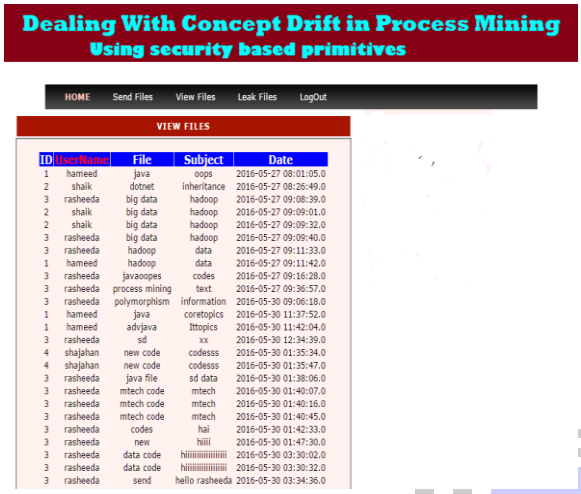

Fig 5: finding of change points of knowing who were viewed the files in online mode.

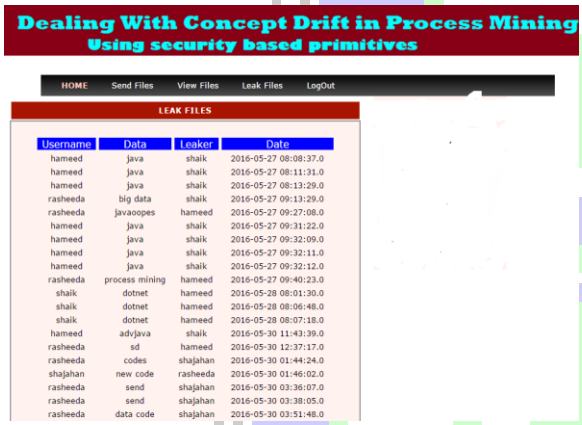

Fig 6: Finding of change points of knowing the changes in knowing who were leaked in online mode.

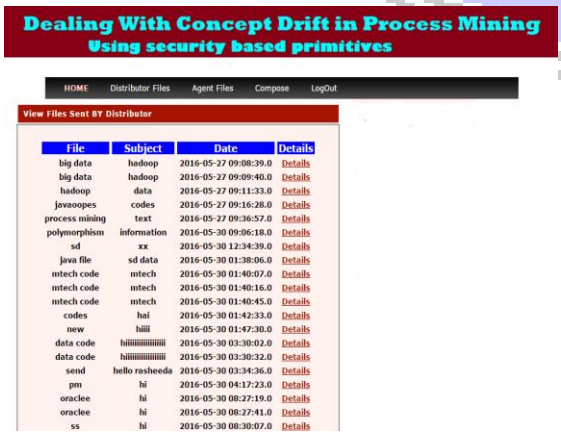

Fig 7: knowing the change points of view files sent by distributor of particular user.

VI. Conclusion

In this paper, we have introduced the topic of concept drift in process mining, i.e., analyzing process changes based on event logs. We proposed feature sets and techniques to effectively detect the changes in event logs and identify the regions of change in a process. Our initial results show that heterogeneity of cases arising because of process changes can be effectively dealt with by detecting concept drifts. Once change points are identified, the event log can be partitioned and analyzed. This is the first step in the direction of dealing with changes in any process monitoring and analysis efforts. We have considered changes only with respect to the control flow perspective manifested as sudden and gradual drifts. Therefore, our analysis should only be observed as the starting point for a new subfield in the process mining domain and there are lots of challenges that still need to be addressed. We transmit the data to the user by using SMTP server address if the proper user is not access the data it's a malicious user or data leaker.

VII. Feature Work

Event logs are support to determine drift in process mining. Feature sets and techniques to effectively detect the changes in event logs and identify the regions of change in a process.

We can transmit the drift data to registered user through their mail address and can notify the Data leakers. Heterogeneity of cases arising because of process changes can be effectively deal with by detecting concept drifts.

V.BIBILOGRAPHY:

[1] (2010). *All-in-one Permit for Physical Aspects: (Omgevingsvergunning) in a Nutshell* [Online]. Available: http://www.answersforbusiness.nl/regulation/all-in-one-permit-physical-aspects

[2] United States Code. (2002, Jul.). *Sarbanes-Oxley Act of 2002, PL 107-204, 116 Stat 745* [Online]. Available: http://files.findlaw.com/news.findlaw.com/cnn/docs/gwbush/sarbanesoxley072302.pdf

[3] W. M. P. van der Aalst, M. Rosemann, and M. Dumas, "Deadline-based escalation in process-aware information systems," *Decision Support Syst.*, vol. 43, no. 2, pp. 492–511, 2011.

[4] M. Dumas, W. M. P. van der Aalst, and A. H. M. Ter Hofstede, *Process- Aware Information Systems: Bridging People and Software Through Process Technology*. New York, NY, USA: Wiley, 2005.

[5] W. M. P. van der Aalst and K. M. van Hee, *Workflow Management: Models, Methods, and Systems*. Cambridge, MA, USA: MIT Press, 2004.

[6] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. New York, NY, USA: Springer-Verlag, 2011.

[7] B. F. van Dongen and W. M. P. van der Aalst, "A meta model for process mining data," in *Proc. CAiSE Workshops (EMOI-INTEROP Workshop)*, vol. 2. 2005, pp. 309–320.

[8] C. W. Günther, (2009). *XES Standard Definition* [Onlilne]. Available: http://www.xes-standard.org

[9] F. Daniel, S. Dustdar, and K. Barkaoui, "Process mining manifesto," in *BPM 2011 Workshops*, vol. 99. New York, NY, USA: Springer-Verlag, 2011, pp. 169–194.

[10] R. P. J. C. Bose, W. M. P. van der Aalst, I. Žliobait˙e, and M. Pechenizkiy, "Handling concept drift in process mining," in *Proc. Int. CAiSE*, 2011, pp. 391–405.

[11] J. Carmona and R. Gavaldà, "Online techniques for dealing with concept drift in process mining," in *Proc. Int. Conf. IDA*, 2012, pp. 90–102.

[12] J. C. A. M. Buijs, B. F. van Dongen, and W. M. P. van der Aalst,"Towards cross-organizational process mining in collections of process models and their executions," in *Proc. Int. Workshop PMC*, 2011, pp. 1–14.

[13] J. J. C. L. Vogelaar, H. M. W. Verbeek, and W. M. P. van der Aalst, "Comparing business processes to determine the feasibility of configurable models: A case study," in *Proc. Int. Workshop PMC*, 2011, pp. 1–12.

[14] 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.

Shaik.Rasheeda received the B.Tech degree from Vasavi Institute of Engineering and Technology Under the JNTUK in Computer Science and Engineering in the year 2014, and pursuing M.Tech degree from Gudlavalleru Engineering College(Autonomous); in specialization with Computer Science and Engineering.

Mr. K. Bhaskar did the Post graduation, M.Tech in Computer Science and Engineering. He has been working in the Gudlavalleru Engineering College, affiliated to Jawaharlal Nehru Technological University. Currently working as an Assistant Professor in the Department of Computer Science and Engineering.