Fraud Detection on Event Logs of Goods and Services Procurement Business Process Using Heuristics Miner Algorithm

Dewi Rahmawati[1], Muhammad Ainul Yaqin[2], Riyanarto Sarno[3]

Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
[1]dewirahmaw4ti@gmail.com , [2]yaqinov@gmail.com, [3]riyanarto@gmail.com

Abstract. **Event logs are history records that contain sequence data for the activity of a case that has been executed by an information system. Event logs can be valuable information with a technique called mining process. With this technique, cheating on the business processes of an enterprise can be detected early on. Thus, the company can commit further examination of business processes, especially the business process of procurement of goods and services to achieve business process is expected.[8] In this study, management data of event log obtained from log data at each event transaction procurement and services. The event log data is then analyzed using a heuristic miner algorithm. Heuristics miner algorithm chosen because it has advantages that are not owned by Alpha++ algorithm that this algorithm can calculate the frequency relation between activities in the log to determine the causal dependency. Heuristic Miner can be used to determine the predominant process of thousands of logs and detect behaviors that are not common in a process.[11] This study aims to detect anomalies on business processes that occur during the process of procurement of goods and services by calculating the fitness value of the event log into the system. Heuristic miner algorithm using the results obtained identification accuracy of 0.88%.**

Keywords : event logs; heuristic miner; process mining; fraud; procurement.

I. INTRODUCTION

Procurement is a series of activities to provide logistics according to the needs, both with regard to the type and specifications, quantity, time and place, price and resources that can be justified. Therefore the purpose of the procurement is to acquire goods or services at a price that can be justified, with the appropriate quantity and quality, as well as the timely completion.[5] Over time, According to data corruption detected by KPK in 2014, 44% of the data in the Commission is the case of procurement. Procurement is one of the biggest sources of corruption in the public finance sector. Every year, KPK and the Commission, reported cases of procurement that contain elements of corruption. However, not a lot that goes into court proceedings. Some cases of procurement were successfully resolved in the courts is precisely to confront the legend that markup "only" 30%.[10]

Corruption is an act of fraud are the most difficult to detect because it deals with the strategy pursued by the other party in the sense of collusion in the form of cooperation between government officials with unscrupulous others illegally (breaking the law) through a violation of the process by covering the financial condition of the truth by doing financial engineering (financial engineering) in the presentation of its financial statements to gain an advantage or perhaps analogous to the term window dressing.[13] Fraud itself is all kinds that can be of men, and pursued by a person, for the benefit of another person with the wrong suggestion or coercion truth, and include all unpredictable ways, full of finesse. Sneaky, hidden, and every dishonest ways that cause others to be fooled. In short it can be said that fraud is cheating (cheating) relating to a sum of money or property.[9]

Fraud can be detected through the mining process, Process Mining is a process that allows management techniques to analyze business processes based on the event log. Process mining uses a special data mining algorithm to extract event log that has been recorded by an information system.[20] The process of mining to detect fraud in the business process by examining the mismatch between event logs from running business processes with business processes compliant companies.

The main assessment in the use of Process Mining is Accuracy and comprehensibility, which focuses on the precision accuracy of visualization of business processes is extracted from the Event Log, while the comprehensibility is easy to understand the business processes that occur on the Event Log. With a data management event log obtained from event data transaction log procurement of goods and services were analyzed using heuristic miner algorithm, is expected to reduce the possibility of misuse of the procurement of goods / services at schools.[2] Due, common practice in many ways resulting in losses significant in the company.[6]

In this study, data management event log obtained from each event log data on procurement transactions. The event log data is then analyzed using a heuristic miner algorithm. Heuristics miner algorithm chosen because it has advantages that are not owned by Alpha++ algorithm that this algorithm can calculate the frequency relation between activities in the

978-1-5090-1381-4/16/$31.00 ©2016 IEEE

log to determine the causal dependency Heuristic Miner can be used to determine the dominant processes of thousands of logs and detect behaviors that are not common in a process. The analysis process is particularly useful for detecting anomalies on business processes that occur during the process of procurement of goods and services.[16]

II. BASIC THEORY

A. *Goods and Service Procurement*

Procurement function is essentially a series of activities to provide logistics according to the needs, both with regard to the type and specifications, quantity, time and place, price and resources that can be justified. Therefore the purpose of the procurement is to acquire goods or services at a price that can be justified, with the appropriate quantity and quality and timely completion.[4]

Some ways procurements for an organization or company is Purchase, Loan, rental, Make Your Own, Redeeming, substitution, Giving / reward and repair / reconditioning. Purchasing is the way procurement is the most common and most easily implemented, but still have to pay attention to the principles of good purchase is the right price, right quantity, the right time, the right place, the right quality and the right source.[3]

B. *Fraud*

Based on the definition of the Institute of Internal Auditors ("IIA"), is a fraud is "An array of irregularities and illegal acts Characterized by intentional deception" which means "set of actions that are not permitted and unlawful characterized by an element of fraud intentional".[14] The Association of Certified Fraud Examiners (ACFE) or the Association of Audit Fraud Certified) divides fraud into three types namely deviations on assets (asset Missappropriation), Statement of Counterfeit (Fraudulent Statement), and corruption (Corruption).[7]

There are three things that contribute to a fraud attempt, the pressure (boost), opportunities (opportunities), and rationalization (rationalization), as illustrated below :

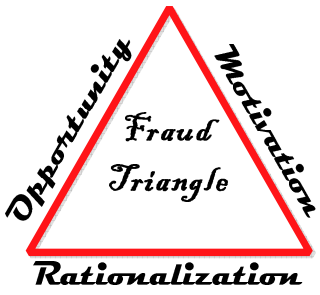

Fig. 1. Fraud Triangle

C. *Event Logs*

Event log is a recording process in the form of transaction history or audit trail in an information system tool. Each system event log information definitely have evidence of ongoing transactions. For example only existing recording of the event log on the ERP (Enterprise Resource Planning). Event Log contains information about activities in the form of a case or a specific task.[17]

Case itself is called the "process instance" is an ongoing activity.[19] For example, the order to the supplier (purchasing), the order by the customer (customer order) and several other events. While the task is activity in the trace, could be stages of activity. So in trace can have many tasks.[15] Event Log consists of several attributes including the Case ID, task, event, user (originator) and time (timestamp).[20][1]

D. *Process Mining*

Process Mining is a process that allows management techniques to analyze business processes based on the event log.[20] The benefits can be obtained with the Mining Process, like to know how the process actually happens. Knowing whether running processes are in accordance with the model previously designed. Knowing which stage deceleration process.

For example, in a paper entitled “Increasing Accuracy of Process-based Fraud Detection Using a Behavior Model”, the paper discuss process mining methods are implemented to analyze both the business processes and the behavior of the originator who performed an event.[18] Process Mining considered quite appropriate because generally log an information system containing data from various cases that executed the organization. Data recorded generally in the form of the start and completion of the work on a part, whoever the culprit, and so forth. Process mining itself consists of three main activities in it based on the presence or absence of previous models, namely discovery, conformance checking, and extension.[17]

E. *Heuristic Miner Algorithm*

Heuristic miner algorithm are algorithm whose application to consider the sequence of an event in a particular case, without regard to the incident that happened in other cases. Heuristic Miner is the second of a process mining algorithm that closely follow the algorithm Alpha. This algorithm was developed by Dr. Ton Weitjers, which uses heuristic approach to solve a lot of problems with the alpha algorithm, so make the algorithm more suitable in practice.[17]

Value Fitness will measure the extent to which the trace log can be associated with the process model valid procedure.[16] For all grades i, $m_i < c_i$and $r_i < p_i$, therefore $fitness = 0 < f < 1$.[12] The following formula is used to obtain the value of fitness [15] :

$$f = \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{k} n_i m_i}{\sum_{i=1}^{k} n_i c_i}\right) + \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{k} n_i r_i}{\sum_{i=1}^{k} n_i p_i}\right)$$

Description :

n_i : number of instances the process of trace i
m_i : number of tokens missing of trace i
c_i : the number of tokens consumed from trace i
r_i : token remaining amount of trace i
p_i : number of tokens produced from trace i

III. DESIGN SYSTEM

A. *Data Source*

The data obtained by the authors is the data from the application of procurement of goods and services of boarding schools that form of data procurement activities undertaken either division head of procurement and procurement staff in the form of data in .sql and .csv format. The data in .csv format consists of several columns containing :

1. Case_ID : contains user id procurement
2. Event_ID: id contains activities that are underway
3. Time : contains the date of the activity
4. Activity : contains the name of the activity
5. Resource : contains the user who perform

B. *Procedure Research*

The procedure used in this study starts from the event log data collection process .sql format. then the data is sorted by CASE_ID to be analyzed appropriately. And event log data obtained .sql format that has sorted. Then use the event log data is formatted .sql first is the data used to calculate the fitness value by using a heuristic miner algorithm. Having obtained the value fitnessnya, after checking the data contained elements of fraud or not. If the fitness value of this data has a value of 1 or more than 1, then the data is not detected symptoms of fraud. and vice versa. If the data has a value of less than 1, then the data is detected symptoms of fraud.

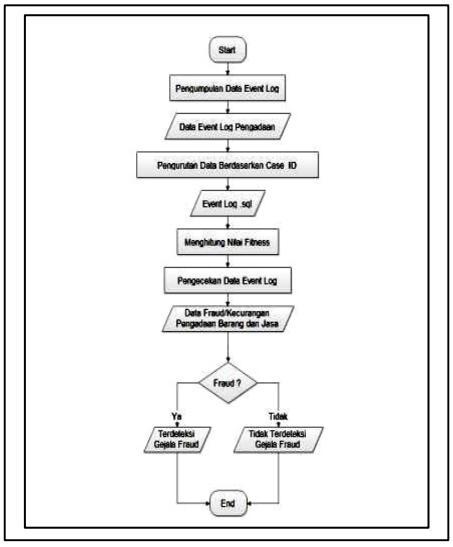

Fig. 2. lines of inquiry

C. *Design Workflow*

Workflow design was originally created using YAWL format editor. YAWL Editor is a tool that is used to create workflow logic that describes working with existing business process flows. The Workflow represents the procurement of goods and services.

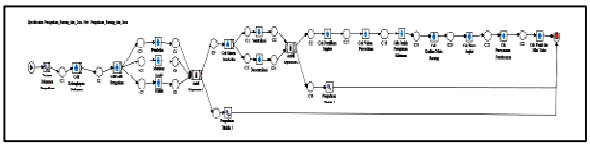

Fig. 3. Design Workflow

Then to get a set of data in the form of Event Log that will be done Miner Heuristic analysis, the model of the process was simulated on YAWL Engine.

TABLE I. TABLE NAMING PROCESS

No. Of Instances	Log Traces
A	Terima Dokumen Pengadaan
B	Cek Kelengkapan Dokumen
C	Cek Cara Pengadaan
D	Pembelian
E	Membuat Sendiri
F	Hadiah
G	Ambil Keputusan 1
H	Cek Sistem Pembelian
I	Pengadaan Ditolak 1
J	Sentralisasi
K	Desentralisasi
L	Ambil Keputusan 2
M	Cek Pemilihan Supplier
N	Pengadaan Ditolak 2
O	Cek Waktu Penyerahan
P	Cek Jumlah Pengiriman Minimum
Q	Cek Kualitas/Mutu Barang
R	Cek Biaya Angkut
S	Cek Persyaratan Pembayaran

With details such as the above, the following is that we make Case Scenario :

TABLE II. TABLE LOG TRACES

No. Of Instances	Log Traces
1	ABCDGHKLMOPQRST
1	ABCDGHJLMOPQRST
1	ABCEGHJLMOPQRST
1	ABCEGHKLMOPQRST
1	ABCFGHJLMOPQRST
1	ABCFGHKLMOPQRST
1	ABCDGHJLPQRST
1	ABCDGHJLMO

IV. RESULTS AND DISCUSSION

A. *Goods and Services Procurement Process*

1. Applicant entering UserID and password that has been owned.
2. Choose the xml data that has been submitted by the applicant to the staff that have been entered into the database of procurement by checking the xml data. After checking the data, then the applicant will be awaiting approval or the approval of the head of the division of procurement for the procurement of applicants are approved or not.

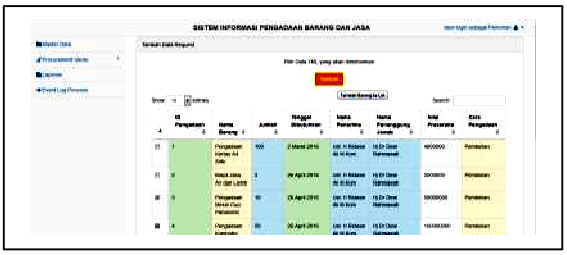

Fig. 4. Display XML Data Table Petitioned

3. After approval of the data by the staff, it will appear in the table applicant's request that the xml data item is approved.
4. The next process is the selection of suppliers by procurement staff proceed with the input goods receipt, invoice verification to the payment to obtain the process log event data such as fig 5. :

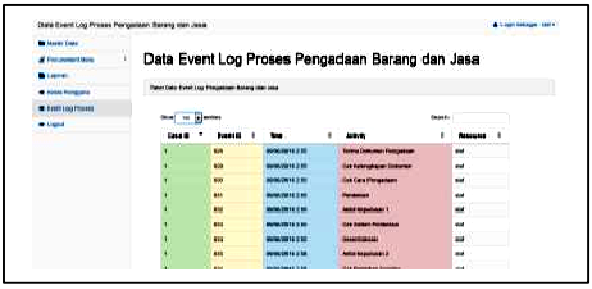

Fig. 5. Data Display Event Log Process

B. *Fraud Detection Process*

In this study, We use 8 Case of the event log for fraud detection procurement of goods and services using heuristic miner. Here is shown the data every case where data is available and display the data calculation fitness values to detect fraud, if the fitness value >= 0.88, the data is not detected symptoms of fraud, and if the value of fitness is at a value < 0.88, the data detected symptoms of fraud. Threshold fitness values are determined (by 0.88) was determined from the data in calculating the value of fitness that we described. Here's the data tally fitness values have been calculated for 1 case 0 missing :

$$f = \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{k} n_i m_i}{\sum_{i=1}^{k} n_i c_i}\right) + \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{k} n_i r_i}{\sum_{i=1}^{k} n_i p_i}\right)$$

$$= \frac{1}{2}\left(1 - \frac{(1\times0)}{(1\times17)}\right) + \frac{1}{2}\left(1 - \frac{(1\times5)}{(1\times21)}\right)$$

$$= \frac{1}{2}\left(1 - \frac{0}{17}\right) + \frac{1}{2}\left(1 - \frac{5}{21}\right)$$

$$= \frac{1}{2}(1-0) + \frac{1}{2}(1-0.238)$$

$$= \frac{1}{2}(1) + \frac{1}{2}(0.762)$$

$$= 0.5 + 0.381$$

$$= 0.88$$

In the same way obtained the results of as on a table III.

TABLE III. TABLE OF FITNESS VALUE IN 8 CASE

No	Case	Missing	Fitness Value
1	1	0	0.88
2	2	0	0.88
3	3	0	0.88
4	4	0	0.88
5	5	0	0.88
6	6	0	0.88
7	7	2	0.76
8	8	5	0.63

From the result that we have done, can be concluded that threshold value of fitness that can be used is 0.88.

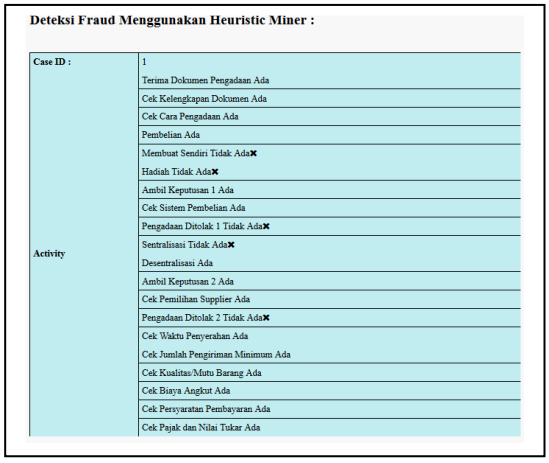

Fig 6. Fraud Identification Page Views

After display page views of all result fraud identification, we display log summary of the fraud identification with showing the log traces values, the missing tokens values, the consumption tokens values, the remaining tokens values, the prduction tokens values, fitness values and fraud checking values of all case. This explanation it can be see on Fig 7.

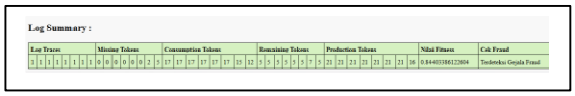

Fig 7. Log Summary Fraud Identification Page View

TABLE IV. RESULT OF FRAUD DETECTION IN 110 CASE

No.	Fitness Value	Amount of Case
1	0.88	75
2	0.76	1
3	0.63	31
4	0.66	1
5	0.65	1
6	0.88	1

In 110 cases the event log that has been detected, there are 6 groups of cases which are divided by the value of fitness resulting as displayed in Table IV, among which:

1. Group Case 1

That produces the value 0.88 which there are 15 activity 0 missing cases and also when the fitness value over a specified threshold based on the training data that is 0.88, then the case was not detected fraud.

2. Group Case 2

For groups of case 2 produces the fitness value of 0.76 that where there are 15 activity 2 missing case that the absence of activity "*cek pemilihan supplier*" and "*cek waktu penyerahan*". Factor because the absence of activity "*cek pemilihan supplier*" in the system event log is not recorded such activities, the user acts to perform these activities do not perform in accordance with the procurement procedures and the user does not select suppliers for the procurement of goods and services. It causes anomalies arise in the procurement system because when there is no activity of supplier selection but why suddenly is the goods and the transaction occurs. And also the activity "*cek waktu penyerahan*" does not exist, this occurs because the activity is not input by the user and also can happen due to the time of delivery is not in accordance with the request of the applicant, for it to check the delivery time abolished because time is not right with the wishes of the user. Therefore, this group of case 2 detected symptoms of fraud. Because of the case of fraud is the absence of a token which indicates that the activity there and also when the fitness value is less than the threshold that is determined based on training data that is 0.88, then the case of fraud detected.

3. Group Case 3

For the case group 3 produces the fitness value of 0.63 that where there are 15 activity 5 missing case, namely the activity "*cek jumlah pengiriman minimum*", "*cek kualitas/mutu barang*", "*cek biaya angkut*", "*cek persyaratan pembayaran*" and "*cek pajak dan nilai tukar*". Occurs missing on "*cek jumlah pengiriman minimum*" due to the absence of such activity in the recording of the event log procurement and also due to the minimum delivery amount is not appropriate because of the number of shipments delivered less than shipped or the number of delivery did not correspond exactly with the desired applicants. For missing happened in "*cek kualitas/mutu barang*" is going missing because the activity is not recorded in the event log procurement and also for the quality / grade goods not in accordance with the willingness of the applicant and not standardized SNI / ISO which is standard stuff. For missing checks on the activities of "*cek biaya angkut*" is going missing due to transport costs do not correspond to the desired applicants and also the activity is not recorded in the event log procurement of goods and services. For missing that occur in the activity "*cek persyaratan pembayaran*" namely the lack of invoices and goods receipt must input procurement staff in the procurement system as proof of payment as well as reporting the finances for the next and can also occur missing because the activity is not recorded in the event log procurement of goods and services. And last missing occurs on "*cek pajak dan nilai tukar*", missing occurs due to the recording of activity does not occur in the event log procurement, there is also because the tax is not in accordance with the tax on goods purchased. Therefore symptoms of fraud is detected. Because of the case of fraud is the absence of a token which indicates that there is activity.

4. Group Case 4

For groups of 4 produces the fitness value of 0.66 with 15 activity 5 missing case, namely the activity "*cek jumlah pengiriman minimum*", "*cek kualitas/mutu barang*", "*cek biaya angkut*", "*cek persyaratan pembayaran*" and "*cek pajak dan nilai tukar*".

5. Group Case 5

For groups of 5 to produce the fitness value of 0.65 with 15 *activity* 5 missing case, namely the activity "*cek jumlah pengiriman minimum*", "*cek kualitas/mutu barang*", "*cek biaya angkut*", "*cek persyaratan pembayaran*" and "*cek pajak dan nilai tukar*".

6. Group Case 6

For groups of 6 produces the fitness value of 0.88 with 15 activity 0 missing cases and also when the fitness value over a specified threshold based on the training data that is 0.88, then the case was not detected fraud.

Definition of activity that is not here is that the activity does not have a standard to be called as such activity. As the activity to make your own, create your own activity means there is not bypassed or done and also the requirements of the process of making itself less fulfilling. It applies to all activity performed.

For the discussion are the procurement of goods and services by naked eye or by a procedural no case of fraud, but in my application fraud is exactly what happened with the missing process / activity or called fraud, namely the activity "*cek cara pengadaan*", in case of "*sentralisasi*" or "*desentralisasi*", the applicant must choose one of the procurement procedure. When choosing his two, then the

procurement system will reject and will not happen procurement of goods and services in the case.

V. CONCLUSION

1. Identification of fraud in the event log procurement of goods and services using heuristic miner algorithm by identification 110 data where 76 case with missing log traces and 34 Case with no missing log traces.
2. The accuracy of the program by using the heuristic miner by 88% with a total of 120 data test and error as much as 13 data.
3. Error occurred on event log groups of 3 to 5 missing case is the absence of "cek jumlah pengiriman minimum", the absence of "cek kualitas/mutu barang", the absence of "tidak adanya cek biaya angkut", the absence of "cek persyaratan pembayaran", the absence of "cek pajak dan nilai tukar".
4. Value Fitness resulting from 110 cases tested is at 0.876 where the status is detected symptoms of fraud.

REFERENCES

[1]. A.J.M.M.Weijters,W.M.P.Van DerAalst,andA.K.Alves De Medeiros, "Process Mining with the Heuristics Miner Algorithm", Netherlands, Eindhoven University of Technology, 2006.

[2]. R. Sarno, Y. A. Effendi, F. Haryadita, "Modified Time-Based Heuristics Miner", International Review on Computers and Software (IRECOS), vol. 11, no. 3, 2016.

[3]. Chary,Production and Operations Management, NewDelhi, Tata McGraw-HillEducation, 2009.

[4]. R.Sarno, Hari Ginardi, Endang Wahyu Pamungkas, Dwi Sunaryono, "Clustering of ERP Business Process Fragments. Proceeding of International Seminar Recent Challenges in Computer, Control and Informatics", http://ic3ina.informatika.lipi.go.id/, published in http://ieeexplore.ieee.org, November 2013.

[5]. R. Sarno, Putu Linda Indita Sari, Dwi Sunaryono, "Decision Mining for Multi Choice Workflow Patterns". Proceeding of International Seminar Recent Challenges in Computer, Control and Informatics (IC3INA), http://ieeexplore.ieee.org, November 2013.

[6]. R. Sarno, D.R. Dewandono, T. Ahmad, M.F. Naufal and F. Sinaga, "Hybrid Association Rule Learning and Process Mining for Fraud Detection," IAENG International Jounal of Computer Science, Vol 42, Issue 2, 2015.

[7]. "Fraud Tree". Retrieved 26 Juni 2016, http://www.acfe.com/fraud-tree.aspx.

[8]. Goeij Yong Sun, "Design and Development of MXML Generator from Event Logs with Txt and Log Extension", Surabaya, ITS Library, 2012.

[9]. Henry Campbell Black,Black' Law Dictionary 4th Edition, ST.Paul, Minn : West Publishing CO, 1968.

[10]. Solichul Huda, R. Sarno, Tohari Ahmad, Heru Agus Santoso. "Identification of Process-based Fraud Patterns in Credit Application", Proceedings International conference on information and communication technology (ICoICT), http://ieeexplore.ieee.org, 2014.

[11]. Nanang Yudi Setiawan, R. Sarno, "Multi-Criteria Decision Making for Selecting Semantic Web Service Considering Variability and Complexity Trade-Off", Journal of Theoretical and Applied Information Technology, vol. 86, issue 2,2016.

[12]. R. Sarno, Kelly R. Sungkono, "Hidden Markov Model for Process Mining of Parallel Business Processes", International Review on Computers and Software (IRECOS), vol. 11, issue 4, pp. 290-300, 2016.

[13]. Solichul Huda, Riyanarto Sarno and Tohari Ahmad,"Fuzzy MADM Approach for Rating of Process-Based Fraud", Journal of ICT Research and Applications, vol. 9, issue 2, pp. 111-128, 2015.

[14]. Olowolaju Monisola, "Effect of Internal Audit On Prevention of Frauds, Errors And Irregularities In Corporate Organisation",Nigeria, IISTE, 2013.

[15]. H. Ritchi, "Identifikasi Pengendalian Aplikasi Dalam Analisis Proses Bisnis", Pustaka UNPAD, 2009.

[16]. A. Rozinat, W.M.P.Van DerAalst,"Conformance Checking of Processes Based on Monitoring Real Behavior", Netherlands, Eindhoven University of Technology, 2007.

[17]. R. Sarno,"Petri Net Model of ERP Business Process Variations for Small and Medium Enterprises", 10th ed, Journal of Theoretical and Applied Information Technology, vol. 54 no.1, August 2013, pp.31-38.

[18]. Solichul Huda, R. Sarno and Tohari Ahmad,"Increasing Accuracy of Process-based Fraud Detection Using a Behavior Model", Surabaya, International Journal of Software Engineering and Its Applications, August 2013.

[19]. P. Weber, "A Framework for The Comparison of Process Mining Algorithms",School of Computer Science University of Brimingham pp.1, 2009.

[20]. W.M.P. Van Der Aalst, Process Mining:Discovery, Conformance and Enhancement of Business Processes, Netherlands, Springer, 2011.