Colored Petri Nets for Integrating the Data Perspective in Process Audits*

Michael Werner

University of Hamburg, Germany
michael.werner@wiso.uni-hamburg.de

Abstract. The complexity of business processes and the data volume of processed transactions increase with the ongoing integration of information systems. Process mining can be used as an innovative approach to derive information about business processes by analyzing recorded data from the source information systems. Although process mining offers novel opportunities to analyze and inspect business processes it is rarely used for audit purposes. The application of process mining has the potential to significantly improve process audits if requirements from the application domain are considered adequately. A common requirement for process audits is the integration of the data perspective. We introduce a specification of Colored Petri Nets that enables the modeling of the data perspective for a specific application domain. Its application demonstrates how information from the application domain can be used to create process models that integrate the data perspective for the purpose of process audits.

Keywords: Business Process Audits, Petri Nets, Business Process Modeling, Process Mining, Business Intelligence, Business Process Intelligence.

1 Introduction

The integration of information systems for supporting and automating the operation of business processes in organizations opens up new ways for data analysis. Business intelligence is an academic field that investigates how data can be used for analysis purposes. It provides a rich set of analysis methods and tools that are well accepted and applied in a variety of application domains, but it is rarely used for auditing purposes.

This article deals with the application of Colored Petri Net models that combine the control flow and data perspective for process mining in the context of process audits. We refer to the example application domain of financial audits for illustration purposes. The benefit of this application domain is the fact that the event data which is necessary for the application of process mining displays structural characteristics that are particularly suitable to be used for the integration of a data perspective. These characteristics relate to the structure of financial accounting and are independent from

* The research results presented in this paper were developed in the research project EMOTEC sponsored by the German Federal Ministry of Education and Research (grant number 01FL10023). The authors are responsible for the content of this publication.

W. Ng, V.C. Storey, and J. Trujillo (Eds.): ER 2013, LNCS 8217, pp. 387–394, 2013.

the used source system. The objective of this article is to illustrate an approach for the integration of the data perspective into process models in the context of process mining and process audits. This approach is not restricted to the illustrated example application domain but can be applied in a variety of application scenarios where information about the involved data is available and valuable for process mining purposes.

2 Related Research

Process Mining is a research area that emerged in the late 1990s. Tiwari et al. provide a good overview of the state-of-the-art in process mining until 2008 [1] and Van der Aalst [2] provides a comprehensive summary of basic and advanced mining concepts that have been researched during the last decade. Jans et al. investigate the application of process mining for auditing purposes [3–5]. They provide an interesting case study [6] for compliance checking by using the Fuzzy Miner [7] implemented in the ProM software framework [8]. The study shows how significant information can be derived from using process mining methods for internal audits. The research results are derived from analyzing the control flow of discovered process models and the interaction of users. We are not aware of any implementations or case studies that consider the data perspective in the context of process mining for process audits. One of the reasons may be the fact that the data perspective in process mining has generally not been investigated extensively yet in the academic community [9, 10]. Exceptions are the research results published by Accorsi and Wonnemann [11] and de Leoni and van der Aalst [10]. The research presented by Accorsi and Wonnemann is motivated by finding control mechanisms to identify information leaks in process models. The authors introduce information flow nets (IFnets) as a meta-model based on Colored Petri Nets that are able to model information flows. Instead of using tokens exclusively for the modeling of the control flow colored tokens are used to represent data items that are manipulated during the process execution. De Leoni and van der Aalst use a different approach. Their intention is to incorporate a data perspective for analyzing why a certain path in a process model is taken for an individual case. The modeled data objects influence the course of routing. They introduce Petri Nets with data (DPN-nets) that base on Petri Nets but that are extended by a set of data variables that are modeled as graphical components in the DPN-nets.

3 Application Domain and Requirements

Financial information is published to inform stakeholders about the financial performance of a company. The published information is prepared based on data that is recorded by information systems in the course of transaction processing. Public accountants audit financial statements for ensuring that the financial information is prepared according to relevant rules and regulations. The understanding of business processes plays a significant role in financial audits [12]. The rationale of considering business processes is the assumption that well controlled business processes lead to complete and correct recording of entries to the financial accounts. Auditors traditionally collect information about business processes manually by performing interviews and inspecting available process documentation. These procedures are extremely time-consuming and error-prone [13]. Process mining allows an effective and efficient reconstruction of reliable process models. Its application would significantly improve the efficiency and effectiveness of financial process audits [14].

Most mining algorithms focus on the reconstruction of control flows in process models that determine the relation and sequence of process activities [9, 10]. Information about control flows is important for auditors to understand the structure of a business process. But this information alone is not sufficient from an audit perspective because an auditor additionally needs to understand how the business processes relate to the entries in the financial accounts [15, 16]. It is therefore necessary to receive information on how the execution of activities in a business process relate to recorded financial entries. This can be achieved by incorporating the data perspective. The relationship between transactions, journal entries and financial accounts is illustrated in Figure 1.

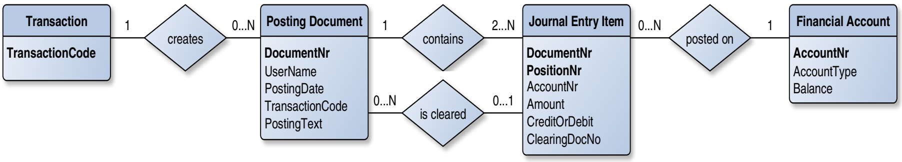


Fig. 1. Accounting Structure Entity-Relationship-Model

A second important concept in financial process audits is materiality [15, 16]. Auditors just inspect those transactions that could have a material effect on the financial statements. To be able to identify which business processes are material information is needed about the amounts that were posted on the different accounts. It is therefore also necessary to model the value of the posted journal entries in the produced process models.

4 Integrating the Data Perspective

The relevant data objects in financial audits are journal entries. They are created during the execution of a business process but their values do not influence the course of routing. They can be interpreted as passive information objects and we therefore refer to the approach used by Accorsi and Wonnemann [11] for integrating them into the process models. Petri Net places and tokens are normally used in process mining to model the control flow. The general approach for integrating the data perspective is to model data objects as colored tokens that are stored in specific places.

The integration is illustrated in the following specification. A Colored Petri Net can formally be expressed as a tuple CPN = $(T, P, A, \Sigma, V, C, G, E, I)$ [17]. Table 1 presents the formal definition of each tuple element, the used net components, and their meaning when applied in the context of this paper. For ease of reference we refer to this type of nets as Financial Petri Nets (FPN) for the remainder of this paper.

Table 1. Specification of Colored Petri Nets for Process Mining in Financial Audits

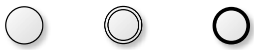
T is a finite set of transitions

The transitions represent the activities that were executed in the process. They display the name of the activity. Further information, for example the transaction code name, can be added.

P is a finite set of places

Places in the FPN represent financial accounts and control places.

Control Places

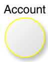


Source Place | Sequence Place | Sink Place

Control places determine the control flow in a process model. For every process model one source place is modeled that connects to the start transactions. The sink place marks the termination of the process. The control places between the start and end transition determine the execution sequence of the process model. A control place belongs to the set of control places CP.

Account Places

credit side of a balance sheet account



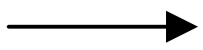
debit side of balance sheet account

credit side of a profit and loss account

debit side of a profit and loss account

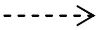
The account places represent financial accounts that are affected by the execution of activities in a process. The symbol color indicates the meaning of an account. Account places belong to the set of account places AP.

$A \in P \times T \cup T \times P$ is a set of arcs also called flow relation



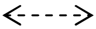
Control Arc

Control arcs connect control places with transitions. They model the control flow in the model.



Posting Arc

Posting arcs illustrate the relationship between activities represented as transitions in the model and financial accounts that are modeled as account places.



Clearing Arc

Clearing arcs are used to model that an activity cleared an entry on the corresponding account. Clearing arcs are double-headed arcs and are used as an syntactical abbreviation for two arcs (p,t) and (t,p).

Σ is a set of non-empty color sets

colset *VALUES* = double	The color set contains the possible values that are posted or cleared.
colset *ACCOUNTS* = string	The color set contains all account numbers.
colset *ACOUNTTYPE* = boolean	The color set contains {1,0} indicating if the represented account is a balance sheet or a profit and loss account.
colset *CREDorDEB* = boolean	The color contains {1,0} indicating if the account place is a representation of the debit or credit side of an account.
colset *EXECUTIONS* = int	The color contains {1,…,n} indicating how often a path was chosen in the FPN.
colset *ACCOUNTPLACES*	*ACCOUNTPLACES* is the color set as a product of *VALUES* * *ACCOUNTS* * *ACCOUNTTYPE* * *CREDorDEB*

Table 1. (*Continued*)

V is a finite set of typed variables such that Type$[v] \in \Sigma$ for all variables $v \in V$.
In FPN models arc inscriptions are modeled as constants. Variables are therefore not necessary and $V = \{\}$.
$C: P \to \Sigma$ is a color set function that assigns a color set to each place.
The color set function in FPN assigns different color sets to places depending if they belong to the group of control or account places: $C(p)\begin{cases} ACCOUNTPLACES\ if\ p \in AP \\ EXECUTIONS\ if\ p \in CP \end{cases}$
$G: T \to EXPR_V$ is a guard function that assigns a guard to each transition t such that $Type\ [G(t)] = boolean.$
$EXPR$ is the set of expressions that is provided by the used inscription language. FPN do not explicitly include guards because they do not model dynamic behavior of transitions that depends on specific input but illustrate the processing of already executed processes. The guard function for FPN is therefore defined as $G(t) = true\ for\ all\ t \in T$.
$E: A \to EXPR_V$ is an arc expression function that assigns an arc expression to each arc a such that $Type\ [E(a)] = C(p)_{MS}$, where p is the place connected to the arc a.
The arc expressions in a FPN are constants. The arc expression function assigns to each posting and clearing arc a set of constants that denote the posted or cleared value, the account number, account type and an indicator if it is a credit or debit posting. For each control flow arc the number of execution times is assigned indicating how often this path was chosen in the process model. $E(a)\begin{cases} \{val \in VALUES, acc \in ACCOUNTS, acctyp \in ACCOUNTTYPE, cord \in CREDorDE \\ with\ Type\ [E(a)] = C(p)_{MS} = ACCOUNTPLACES\ if\ p \in AP \\ \{ex \in EXECUTIONS\ with\ Type\ [E(a)] = C(p)_{MS} = EXECUTIONS\ if\ p \in CP \end{cases}$
$I: P \to EXPR_\emptyset$ is an initialization function that assigns an initialization expression to each place p such that $Type\ [I(p)] = C(p)_{MS}$
The initialization function of a FPN assigns initialization expressions to each place as follows: $I(p)\begin{cases} n'ex \in EXECUTIONS\ with\ Type\ [I(p)] = C(p)_{MS} = EXECUTIONS\ if\ p = source \\ \emptyset_{MS}\ otherwise \end{cases}$ Only the source place is initialized in a FPN. The initialization expression for $p = source$ generates n tokens in the initial marking $M_0(p)$, one for each connected start transition. The inscription of each token is a member of the set $EXECUTIONS$.

Figure 2 illustrates a FPN model for a purchasing process. The model includes:

Transitions:	$T = \{MB01, MIRO, F110\}$
Places:	$P = \{Source, S1, S2, Sink, 100\_D, 200\_D, 200\_C, 300\_D, 300\_C, 400\_D\}$
	$CP \subseteq T = \{Source, S1, S2, Sink\}$
	$AP \subseteq T = \{100\_D, 200\_D, 200\_C, 300\_D, 300\_C, 400\_D\}$.
Color sets:	$VALUES = \{50{,}000\}$, $ACCOUNTS\ \{100, 200, 300, 400\}$, $ACOUNTTYPE = \{0{,}1\}$, $CREDorDEB = \{0{,}1\}$; $EXECUTIONS = \{1\}$; $ACCOUNTPLACES = VALUES * ACCOUNTS * ACCOUNTTYPE * CREDorDEB$.
Initialization:	$I(p) = \{1'1\ for\ p = source\ and\ \emptyset_{MS}\ for\ p \neq source$

The model shows that the processing of received goods created journal entries with the amount of 50,000 on the raw materials and the goods receipt / invoices receipt (GR/IR) account. The receipt of the corresponding invoice for the purchased goods was processed with activity MIRO which led to journal entries on the GR/IR account and a creditor account. It also cleared the open debit item on the GR/IR account that was posted by MB01. The received invoice was finally paid by executing the activity F110 which posted a clearing item on the creditor account and a debit entry on the bank account.

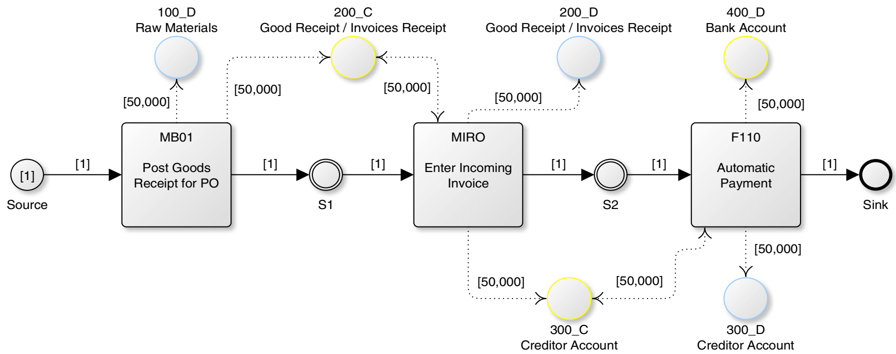


Fig. 2. Simple FPN Example of a Purchase Process

The example in Figure 2 demonstrates how the used Petri Net specification can be used to model the control flow and the data perspective simultaneously in a single model. The transitions in the model create colored tokens when they fire. They store information on the values that are posted on the connected account places. The illustrated model does not only mimic the execution behavior of the involved activities but also the creation of journal entries on the financial accounts. The model shows the execution sequence and the value flows that are produced.

5 Implementation and Experimental Evaluation

We applied the described FPN specification on real world data to evaluate if the theoretical constructs can actually be used in real world settings. The data base for the evaluation included about one million cases of process executions from a company operating in the manufacturing industry. The raw data was extracted from a SAP ERP system and checked for first and second order data defects [18]. We used an adjusted implementation of the Financial Process Mining (FPM) algorithm [19] that is able to produce FPN models. The mining was limited to 100,000 process instances that affected a specific raw materials account. The mining resulted in 113 process variants. They were analyzed in the evaluation phase by observation using the yEd Graph Editor [20] for verifying if the process models presented the desired information adequately. Selected models were further tested using the Renew software [21] for evaluating if correct FPN were created by simulating the execution of the models. The evaluation demonstrated that FPN can be created correctly by using an adapted FPM algorithm. The produced FPN are able to adequately model the control and data flow based on the used real world data.

The same modeling procedure was used in a different scenario to analyze technical customer service processes which is not described in this paper due to place restrictions.

6 Summary and Conclusion

Process Mining is an innovative approach for analyzing business processes but it is rarely used in the context of process audits. The successful application of process mining requires the consideration of domain specific requirements. A common requirement is the incorporation of a data perspective which has not been addressed extensively yet in the academic community. We have introduced a specification of Colored Petri Nets that allows the modeling of the control flow and data perspective simultaneously. We referred to the application domain of financial audits as a representative example to demonstrate how the data perspective can be included by referring to relevant application domain requirements.

The evaluation of mined process models shows the suitability of the presented specification in real world settings. The evaluation included the data from a SAP system of a single company. It can therefore not be concluded that the results also hold true for other companies or ERP systems. But the presented specification bases on the structure of financial accounting and is therefore independent from any proprietary ERP software implementation or industry. Evaluation results from further current research indicate that the presented results are also applicable in other settings.

References

1. Tiwari, A., Turner, C.J., Majeed, B.: A review of business process mining: state-of-the-art and future trends. Business Process Management Journal 14, 5–22 (2008)
2. Van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Heidelberg (2011)
3. Jans, M., Alles, M., Vasarhelyi, M.: Process mining of event logs in auditing: opportunities and challenges. Working Paper. Hasselt University, Belgium (2010)
4. Jans, M.J.: Process Mining in Auditing: From Current Limitations to Future Challenges. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM Workshops 2011, Part II. LNBIP, vol. 100, pp. 394–397. Springer, Heidelberg (2012)
5. Jans, M., van der Werf, J.M., Lybaert, N., Vanhoof, K.: A business process mining application for internal transaction fraud mitigation. Expert Systems with Applications 38, 13351–13359 (2011)
6. Jans, M., Alles, M., Vasarhelyi, M.: Process Mining of Event Logs in Internal Auditing: A Case Study. In: 2nd International Symposium on Accounting Information Systems (2011)
7. Günther, C.W., van der Aalst, W.M.P.: Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 328–343. Springer, Heidelberg (2007)

8. Process Mining Group: ProM, http://www.processmining.org/prom/start
9. Stocker, T.: Data flow-oriented process mining to support security audits. In: Pallis, G., et al. (eds.) ICSOC 2011 Workshops. LNCS, vol. 7221, pp. 171–176. Springer, Heidelberg (2012)
10. De Leoni, M., van der Aalst, W.M.: Data-Aware Process Mining: Discovering Decisions in Processes Using Alignments (2013)
11. Accorsi, R., Wonnemann, C.: InDico: Information Flow Analysis of Business Processes for Confidentiality Requirements. In: Cuellar, J., Lopez, J., Barthe, G., Pretschner, A. (eds.) STM 2010. LNCS, vol. 6710, pp. 194–209. Springer, Heidelberg (2011)
12. International Federation of Accountants: ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment (2012)
13. Werner, M.: Einsatzmöglichkeiten von Process Mining für die Analyse von Geschäftsprozessen im Rahmen der Jahresabschlussprüfung. In: Plate, G. (ed.) Forschung für die Wirtschaft, pp. 199–214. Cuvillier Verlag, Göttingen (2012)
14. Werner, M., Gehrke, N., Nüttgens, M.: Business Process Mining and Reconstruction for Financial Audits. In: Hawaii International Conference on System Sciences, Maui, pp. 5350–5359 (2012)
15. Schultz, M., Müller-Wickop, N., Nüttgens, M.: Key Information Requirements for Process Audits – an Expert Perspective.pdf. In: Proceedings of the 5th International Workshop on Enterprise Modelling and Information Systems Architectures, Vienna (2012)
16. Müller-Wickop, N., Schultz, M., Peris, M.: Towards Key Concepts for Process Audits – A Multi-Method Research Approach. In: Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics, Utrecht (2013)
17. Jensen, K., Kristensen, L.M.: Coloured petri nets. Springer (2009)
18. Kemper, H.-G., Mehanna, W., Baars, H.: Business intelligence - Grundlagen und praktische Anwendungen: eine Einführung in die IT-basierte Managementunterstützung. Vieweg + Teubner, Wiesbaden (2010)
19. Gehrke, N., Müller-Wickop, N.: Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems. In: Proceedings of the 16th Americas Conference on Information Systems, Lima, Peru (2010)
20. yWorks GmbH: yEd - Graph Editor, http://www.yworks.com/de/products_yed_about.html
21. University of Hamburg: Renew - The Reference Net Workshop, http://www.renew.de/