Insuring Sensitive Processes through Process Mining

Jorge Munoz-Gama
Universitat Politecnica de Catalunya
Barcelona, Spain
jmunoz@lsi.upc.edu

Isao Echizen
National Institute of Informatics
Tokyo, Japan
iechizen@nii.ac.jp

***Abstract*—Any system, company or complex task is composed by processes, i.e., sequences of actions performed in some established order. Some of these processes are considered sensitive for the role they play within the system, or for the sensitive data they manage. In such cases, the trust in this processes is desirable, requiring their verification, monitoring, auditing, but also the possibility of being insured by a third party. In this approach we propose a schema for insuring sensitive process based on the use of formal models and Process Mining techniques (i.e. process management techniques that allows for the analysis of processes based on event logs). The experimental results presented show that this new approach could be useful in the context of insurance and analysis of processes.**

***Keywords*-process mining; insurance schema; trust model; risk; accountability;**

I. Introduction

Nowadays, the conception of how the activities within companies and organizations should be managed is changing. The idea of a data-oriented system is being replaced by process-aware systems where the main entity is not the data processed any more, but the process itself [1]. From the procedure of handling building permission applications in a municipality [2], to the treatment and diagnosis steps for patients with cancer [3], all them are being designed, controlled and analysed in terms of processes.

Some of these processes are simple and have no repercussion outside the ambit where they are performed. However, other less isolated and more sensitive processes may have legal, economical or personal consequences for the actors involved, especially when the results of executing these processes are not the ones expected. Example of these processes are the ones related with the health, processing of confidential data, bureaucratic procedures, or economic transactions. In such cases, besides the corresponding verifications, testing, controls and maintenance, it should be necessary (even mandatory according to the law) to insure the process, i.e., even taking all the precautions to prevent it, when the process goes wrong, somebody must be accountable for that.

In this work, we illustrate how *Process Mining* [4] (i.e., the process management techniques that allows for the analysis of processes based on event logs) can be an useful tool for the insurance of sensitive processes. In this approach, the following general scenario is assumed (see Fig.1): a service provider company (SP) provides a sensitive service (P) between an user provider (UP) and an user receiver (UR). This service is insured by a third-party insurance company (IC).

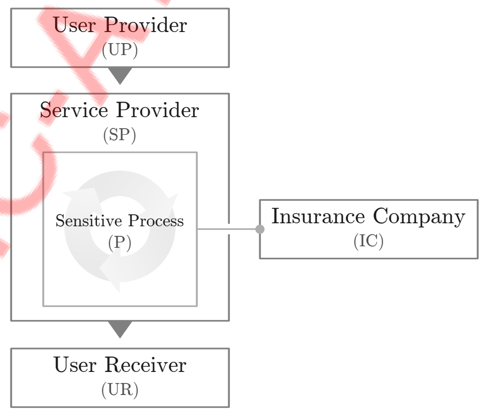


Figure 1. Insuring Sensitive Processes

The paper is structured as follow: in the remaining part of Sec. I, a running example is introduced to illustrate some of the concepts presented in the paper. In Sec. II we introduce the concept of Process Mining, its terminology and approaches, and the similar works in the area. In Sec. III we present the insurance schema proposed, and in Sec. IV the cycle to insure a process based on the Process Mining techniques. Finally in Sec. V and Sec. VI some experimental results, conclusions and future points are addressed.

Illustrative Example

During the remain part of the paper we will use a running example, in order to illustrate the approach proposed in this paper, motivate its utility, and exemplify the techniques presented and its results.

The example reflect a plausible sensitive process in the context of an online car market: an online company that collect and distribute pictures and videos of the cars from the users to help selling the cars to other uses. In particular, the process illustrated represent the *anonymization procedure* over the data: before being distributed, the data collected must be anonymized, i.e., the faces of the people and the license plates of the cars must be blurred, and the audio of the videos must be filtered.

In this process we identify a set of steps or tasks than are performed in a established order. Table I contain all the tasks of the anonymization process. For the sake of clarity, a capital letter has been assigned to each task (e.g., A is content analysis, ...).

A	content analysis	F	license plate detection	K	audio filtering
B	pre-proc. picture	G	license plate anonym.	L	post-proc. picture
C	pre-proc. video	H	audio processing	M	post-proc. video
D	facial anonym.	I	frame processing	N	time-stamping
E	facial detection	J	frame anonym.		

Table I
TASKS OF THE ILLUSTRATIVE EXAMPLE PROCESS

The anonymization process can be divided in two parts depending on the type of data processed: picture or video. In the case of picture, the faces and the license plate must be detected (E,F) and blurred (D,G), until no one remain unanonymized. In the case of video, the audio (H,K) and the image (I,J) are processed independently. Finally, independently of the type of the data, any instance of the process must begin with an analysis of the content (A) and finish with a time-stamping (N). Note that, the formalization and the modelling of the process, as part of the insurance schema, is one of the main goals of the scenario proposed in this paper, and therefore, will be seen in detail in Sec.IV-A

II. PROCESS MINING

Process Mining is a novel area of research that has received a lot of attention in the recent years. One can see a parallelism between *Data Mining* and Process Mining: while the former focuses into analysing massive data in order to find hidden relations, the later focuses into processes governing an information system. Interesting questions regarding an information system are tackled by process mining techniques: What are the processes of my system? Do my current process models conform to the reality? Can I extend the information of my process model with other information? These three questions are faced by *Discovery*, *Conformance* and *Enhancement* disciplines within Process Mining [4], respectively. One can read the *Process Mining Manifesto* [5] to have a general overview of the field and challenges.

Discovery is maybe the most challenging problem in process mining: given a *log* (set of system traces) reflecting a process execution, derive a *formal* process model that includes the behaviour seen in the log. Algorithms for process discovery can be oriented to the *control-flow* part of the process, the data part, or the social dimension (e.g., discover working groups or strong collaborations that are needed to accomplish a process). Accordingly, formal process models range from *Petri nets* [6] to *Event-Process Chains* [7], or *Business Process Model Notation* [8], among others. In the last ten years, there has been several proposals for process discovery algorithms, but also new formal models to better represent a process have been developed [9]. Discovery is therefore an evolving field which will be widely used in the near future.

Conformance checking [10], [11], is the discipline most related with this paper, and tackles a crucial problem: given a process model (which represents the intended behaviour), and a log (which represents the real behaviour), how similar are these two behaviours? In conformance, four quality dimensions are used to estimate the conformance level between a process model and a log:

- *Fitness*: indicates how much of the observed behaviour is captured by (i.e. "fits") the process model.
- *Precision*: refers to overly general models, preferring models with minimal behaviour to represent as closely as possible the log.
- *Generalization*: addresses overly precise models which overfit the given log, thus been possible to generalize.
- *Simplicity*: refers to models minimal in structure which clearly reflect the described behaviour.

Finally, the Enhancement discipline [4] addresses the incorporation of extra information into the model that can be used to enrich it. The use of this discipline is not considered in this paper.

Related Work

The use of Process Mining techniques in the context of auditing is not new. In [12], the authors discuss about idea of how such techniques might be useful for the different types of auditing: auditing using historic data, auditing based on models only (i.e., the models describing the processes are analysed without the use of any kind of data), and auditing using current data (i.e., monitoring processes on-the-fly). In [13] a deeper experimental study about the use of Process Mining for internal auditing is presented, illustrating the differences with other auditing approaches. The work includes the results obtained for a real business context study. Finally, in [14], the main focus resides on security auditing. In that work, the authors present some methodology and a real case study where a set of different security requirements are verified using Conformance techniques. This security requirements are not restricted to control flow, but also include restrictions about usage control and time constrains.

One of the main differences of the approach presented here with respect to other approaches in the literature is the focus, i.e., while other approaches focus on the use of Process Mining techniques for auditing and security verification of the system, the focus of this paper is the insurance of the processes, and how to deal with claims from actors involved in the process. Moreover, while other works discuss about the several aspects of the process, in this paper we focus in detail about the flow that characterise the process. The other main difference of this paper resides in the use of the conformance techniques. Like in [14], the work presented in this paper is specially focused on the Conformance discipline within Process Mining (without skipping the discussion about the utility of the other disciplines, e.g., discovery in Sec.IV-A). However, instead of using conformance replay techniques based on local decisions, we propose the use of alignment conformance techniques which result will include the analysis from a global point of view. Moreover, in contrast with other approaches where only fitness dimension within conformance is considered, we also propose the use of other dimensions in the analysis (e.g., precision in Sec.IV-D).

III. Insurance Schema

The most crucial part and common in all insurance procedures of any kind (e.g., objects, events, ...) is the insurance agreement. A contractual document that include aspects concerning about the conditions of the insurance, who will be accountable in each possible scenario, and of course, the details of the element insured. In the schema for insurance sensitive processes proposed in this paper this is not an exception. However, unlike other contracts where a more legal and textual notation is used, we propose the use of formalization and models instead. In other words, the use of a formal model notation will remove from the agreement any kind of ambiguity, given a framework for formal verifications and analysis in such cases where there are problems with the process insured. Figure 2 illustrate the schema proposed in this work, where the formal model of the process is the corner stone in the insurance agreement. Furthermore, the use of formal models provides the possibility of using automatic techniques for the analysis of the process [4], or even use this model as part of the implementation of a process-aware information systems [1].

Given a process, there is a wide variety of perspectives that can be insured, e.g., the possible access to sensitive data, the satisfaction of some given rules or properties, or the roles of actors within the process, among others. Unlike other approaches, in this work, we will focus on the flow of the process, i.e., the sequence of steps in order a process must follow to be considered correct. It exists a wide range of notations for the formalization of a flow model, each one with different properties and different existing tools for it, e.g., Petri nets, BPMN, or EPC, among others. This model could be extended with annotations concerning the other aspects of the insurance agreement, e.g., who is accountable for each one of the steps in the flow.

IV. Process Mining based Insurance Cycle

Besides the schema introduced in the previous section, this work also propose a cycle of actuation for the insurance of processes. Such cycle is composed by four phases, and can be iterated during the life time of the process within

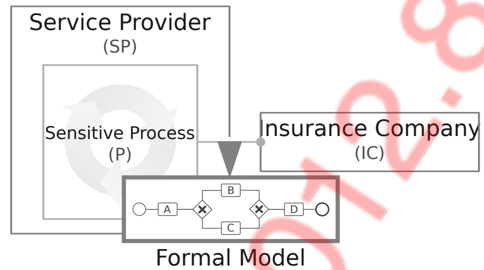


Figure 2. Schema proposed for insuring processes based on Formal Models

the system: the *modelling phase*, the *validation phase*, the *forensic phase*, and the *re-design phase*.

A. Modelling Phase

The initial phase in the Process Mining based Insurance Cycle correspond with the *Modelling Phase*, i.e., the stage where the flow of the process to be insured is modelled, so it can be the cornerstone for the future insurance policy between the service provider and the insurance company.

The procedure of modelling a process is a huge resource-consuming task, e.g., it involves experts in the domain, holding interviews and questioning the several actors that interact with the process, and it may take months to complete it [15]. Moreover, hand-made models tend to be subjective, the perception of the people involved may be biased, and are normally concentrated on the ”expected” behaviour or simplified for the sake of understandability [4]. For that reason, in such cases where it exists event logs reflecting the process to be insured, it is recommended to use some of the Process Mining discovery approaches available [4]. Nowadays, the existence of logs recording the process steps is a really plausible assumption, e.g., the treatment followed by the cancer patients in a hospital [3] or actions performed by a X-ray machine [4] are practical examples that confirm that assumption.

The usage of discovery techniques is the option chosen to illustrate the modelling phase with the running example of Sec.I. An event log containing sequences or traces of the process described has been processed by the ILPMiner discovery algorithm [16], given as a result the model of Fig.3. Note that the result is modelled using Petri Nets[1] [6]. We choose this notation to illustrate the approach given the mathematical bases of Petri nets, its intuitive graphical representation, and the wide range of Process Mining approaches based on that notation (included the bast majority of conformance approach such as [10], [2] or [11]). However, some of these concepts can be extrapolated to other notations [17].

[1]For the reader not familiar with Petri nets, a Petri net is a bipartite graph that contains two types of nodes: places (circles) and transitions (boxes). A place may contain tokens (black dots), and a transition can fire if its predecessor places contain a token. When fired, the transition removes a token from each input place and adds a token to each successor place.

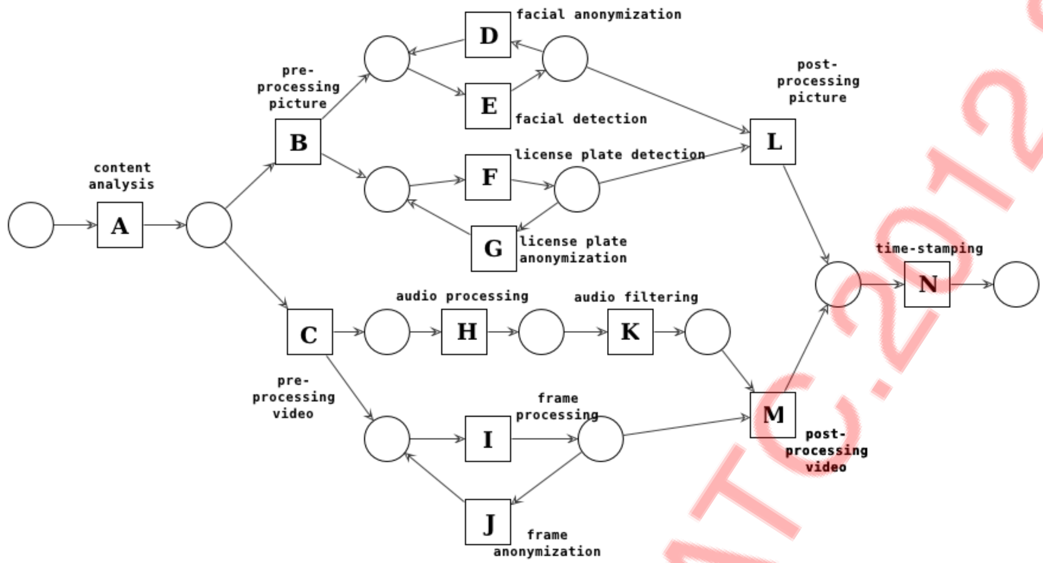


Figure 3. Model discovered for the running example

B. Validation Phase

The second phase on the Process Mining insuring cycle consists on the *Validation Phase*, i.e., the stage where the third party insurance company (IC) analyse the process to be insured. According to the results of the analysis, the conditions of the insurance agreement are set, e.g., the price of the insurance according to the possible risk, which parts of the process is the IC willing to insure, or if it is not willing to insure it at all.

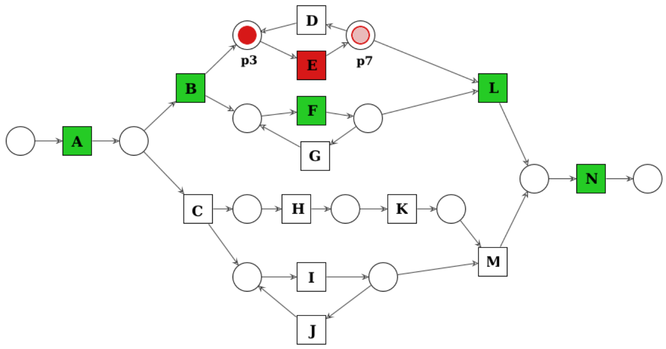
This validation of the correct flow of the process is performed by a series of benchmarks. Each benchmark is composed by some possible well-known input the process is expecting. The input of each benchmark is tested, and the results are compared with the expected correct flow. The number of mismatches, its importance, and specially, in which part of the process they are located, are the elements used to determine the risk of insuring the process. From a IC point of view, it is recommended that this validation procedure is applicable periodically, or each time some aspects of the process changed, e.g., new version of the software or changes in the habits of the service consumers.

Although the comparison between model and the resulting flow may seem straightforward, this is not the case for a huge number of cases, specially such where it exists some kind of discrepancies between flows (the most interesting ones from a insurance point of view). In other words, the resulting flow corresponds with a realistic and not model-driven execution of the process. In such cases, the resulting traces must be mapped over the model in order to be analysed, given where the mismatches exactly are within the model, and therefore, who must be accountable for that. This process of superposing the trace over the model is known as *fitness*, and it is a non trivial task studied in works such as [10] (where they propose a non-blocking replay of trace through the model). In the work presented in this paper, we propose the use of a different fitness analysis technique based on finding the best *alignment* between the flow and the possible model paths, from a global point of view [2]. In other words, given a trace, the well-known A* algorithm is used to find the optimal corresponding model run that minimize the cost. This cost is computed based on a matrix of costs, where it is assigned a penalty to each possible action based on if this action is performed in the right moment or not (according to the model). Unlike other approaches where the solution is local, the use of this alignment technique make possible to find the best optimal solution from a global point of view.

In order to illustrate the validation approach, we present three possible benchmarks cases for the running example. The first case correspond with a picture where the face of a person must be blurred, but not the licence plate because it is not visible in the picture. When the process of anonymization is executed with this picture, the trace resulting is ABFLN. However, this is not a valid run according to the model. This is reflected on Fig.4a, where the results of the alignment conformance technique shows some discrepancies with the model included in the insurance agreement, i.e., the task of *facial detection*(E) has been skipped, leading to a incorrect anonimization process. Figure 4b shows the mapping of the trace over the model graphical representation, i.e., according to Petri Net firing rule, the given trace is not a valid trace of the model. In such case, insure this process may be a risk from the insurance company point of view.

The first case contrast with the second case, where a

(a)



(b)

Figure 4. Alignment result and Petri Net display of first benchmark

license plate of a car must be blurred. In this second case, the resulting trace is ABEFGFLN. When the trace is juxtaposed with the model (Fig.5a), it is corroborated that the path the process has followed has no fitness mismatches, and correspond exactly with the path a picture of this kind must follow.

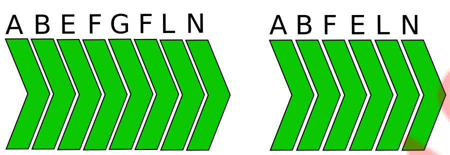


Figure 5. Alignment results for the second and third benchmarks

Finally, the third case involves a video where the audio has to be filtered. The resulting trace in such case is ABEFLN. The conformance technique detect no fitness problem with this trace (Fig.5b), i.e., in the model there is a path that perfectly replay this trace. However, the same techniques make possible to identify the flow of the process through the model. The flow flowed in this case does not correspond with the flow of a video as this must follow, but the flow of a picture. In this particular case, insure this process will be a serious risk for the insurance company.

C. Forensic Phase

A *Forensic analysis* is a analysis performed *a-posteriori*, i.e., when the error, problem or anomaly in the process has already happened. In the insurance scenario proposed in this work, this analysis is performed when some execution of the process was been the object of the claim from part of a external user of the service provider. A forensic analysis is proposed to determine where the error has been made, and who was to be accountable for it.

The same conformance technique presented in the validation phase can be used for the forensic analysis. However, unlike validation phase where the analysis involves several traces, the forensic phase analysis is performed with only one trace, corresponding with the instance of the process involved in the insurance dispute. The conformance technique make it possible to superpose the flow involved in the claim, detect the discrepancies and where are located, and determine according to that who must be accountable for that claim.

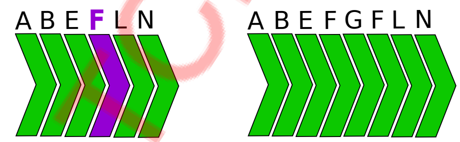


Figure 6. Two possible claims for the running example

Figure 6 show two possible claims for the running example, and how the conformance alignment method is used to identify who must be accountable for that. Both claims are related with some picture of a car where the license plate has not been correctly anonymized. The trace for the first claim is ABELN. The alignment approach presented in the Sec.IV-B is used to detect that licence plate anonymization action (F) has been completely skipped. This correspond with a violation of the model agreed between the service provider and the insurance company in the insurance contract, and therefore, the insurance company will not be responsible for that claim. This contrasts with the second claim, where the trace is ABEFGFLN. In such case, the alignment prove that the model agreed has been respected, and therefore, the insurance company should be accountable for that claim.

D. Re-design Phase

Practical experience shows that the assumption that a process remains in a steady state during the time is unrealistic, i.e., processes evolve adapting to changes in the real world. This is known as *concept drift* [18]. Hence, this should be also considered in the insurance scenario proposed in this work, i.e., the model included in the insurance policy (even the ones obtained by process discovery algorithms) must be *re-designed* during the time. Even when the changes in the process had been minor ones, it may be interesting from both, insurance company and service provider point of view, to re-design the model for reflecting more *precisely* the reality, i.e., if the model allows less possible behaviour, the insurance company takes less risk, and therefore, this is also reflected in the policy price that the service provider company has to pay.

Following with the idea of this paper, Process Mining techniques can be also used to aid in the phase of re-design. In particular, conformance approaches focused on the *precision* dimension instead of the fitness dimension, i.e., not if the model captures all the traces in the log, but how precisely the model captures the behaviour reflected in the log, and not more behaviour.

In [10], a technique is presented to measure the precision of system. This approach is based in comparing the *precedence* and *following* relation between both, the log and the model. This comparison make it possible to visualize where the imprecisions may be produced. However, this approach may require the completely exploration of the behaviour, making the efficiency of this approach for large and real cases a serious drawback.

The authors in [11] propose an approach, *ETConformance*, to efficiently determine the precision between a log and a model based on *escaping points*, i.e., the points where the behaviour reflected in the model escapes (is more generic) than the one really reflected in the log. In [19] this approach is extended to tackle logs with noise and to assess the severity of this escaping points. We propose to use the escaping points detected by ETConformance as a starting point for re-designing the model, improving the its precision.

In the running example scenario, a change in the process is assumed, e.g., the rise of high-speed internet connections makes the use of the video option the almost-only used option, to the detriment of the picture option. In this new context, a re-design is required to adapt to the new reality, decreasing the risk for the insurance company, and consequently the price the service provider pays for the insurance (it may not be worth it to insurance a part of the process that it is barely used). When the model and the new log reflecting this changes in the consumer habits is processed with ETConformance, a main escaping point is detected, i.e., the one concerning B after executing A. In other words, after executing the task A, the possibility of executing B is an option contemplated in the model, but not reflected in the log any more. This escaping point is used to determine which behaviour should be branched to make the model more precise, i.e., the B and all the tasks that can not be reached any more. The resulting model can be seen in Fig.7. This model is more precise according to the precision metrics available in the Process Mining literature (for example, etc_p [19] result is 0.85 instead of the 0.78 of the old model).

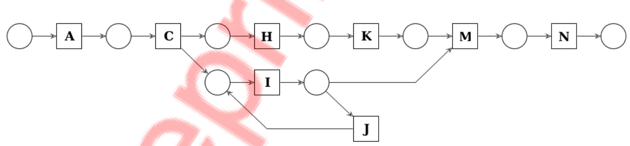


Figure 7. New model for the running example after the Re-design phase

V. Experimental Results

In this section we present the results of a study in order to show experimentally the usefulness of the approach. The study is divided in three parts. In the first one, we illustrate the results of applying the validation phase in possible and different variants of the same process. In the second part, we study a set of possible claims and how are they managed by the approach. And finally, in the third part we illustrate how precision conformance techniques can be used to re-design the model. All the experiments are been performed using the ProM tool. ProM[2] is an open-source extensible framework that supports a wide variety of process mining techniques in the form of plug-ins. For the validation and claim experiments, the *Replayer plug-in* has been used, while for the re-design experiments we used *ETConformance plug-in*. Both plug-ins are publicly available.

In the first part of the experiments, three variants of the same process are being considered, each one corresponding with a well-differentiated scenario. In all three cases, the same model for the insurance agreement is considered. The first case reflect a perfect well-performed scenario, where it is proved (before the validation analysis) that not error is possible. As it was expected, the results obtained by the experiments reflect this situation, i.e., all the set of benchmarks executed correspond with the expected flow within the model. In this case, insuring this process might suppose a low-risk for the insurance company. The second case reflect a more common scenario, where we introduce some noisy and errors in a 5% of the executions of the process, in order to simulate possible problems. In such case, the approach is able to detect such problems and place them within the model. This information is crucial in order to decide or not to insure the process, or the exact condition an responsibilities of such insurance. Finally, the third case reflect a chaotic scenario, where the process is executed in a sloppy way, and the errors are common (only 50% of the process executions are considered correct by a domain expert). Also in this case, the approach is able to the detect the discrepancies between the agreed model and the real process execution. To insure this third scenario will be an inadmissible risk for the insurance company.

In the second set of experiments we show how the approach presented is able to correctly manage claims from possible service users. Two sets of benchmarks has been designed for that experiment. The first benchmark contains claims where the execution of the process assigned to them has been manually modified simulating being slightly non compliant with the model in the insurance agreement. The second benchmark simulates cases where, despite that the correctness of the sequence has been verified by a domain expert, the result is not satisfactory. Both sets of benchmarks

[2] http://www.promtools.org/prom6/

has been juxtaposed with the model using ProM, confirming the results expected, i.e., on one hand, the discrepancies in all the cases of the first benchmark has been detected, and therefore the insurance company is not responsible of the claims due an violation on the model agreed; on the other hand, the forensic analysis on the second benchmark determines that agreed model has been respected in all the cases, and therefore, the insurance company must be accountable.

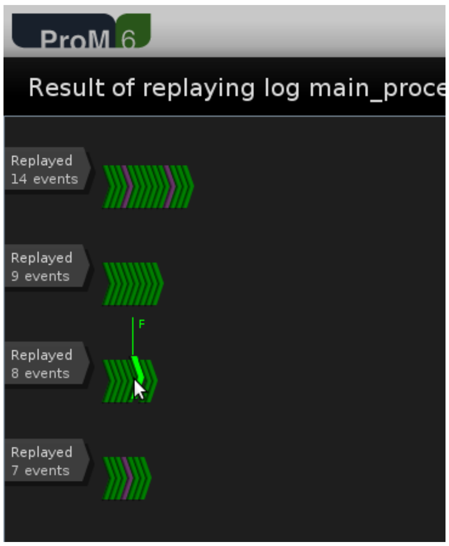


Figure 8. ProM capture

Finally, the last experiments show how precision conformance technique can be used for re-design the agreed model. In that experiment, a change in the underlying model is assumed, i.e., a new slightly different log (with respect to the one used to generate the previous model) is simulated to reflect this new reality. The new log and the agreed model are processed using ETConformance, detecting an imprecision in the model: the concurrency between two task does not exist any more in this new scenario. The model is modified to make this two task sequential, and the resulting model reflects more accurately the real process to be insured. This is confirmed by the results of the different precision metrics available in ProM.

VI. Conclusions and Future Work

In this paper we have presented a methodology for the use of Process Mining in the context of insuring the flow of sensitive processes. This new schema is based on the idea of a formal and unambiguous model as the cornerstone for the insurance agreement between parts. In that sense, the application of Process Mining techniques, and specially the conformance discipline approaches, can be used for validate the process flow, and solve possible claims related with it.

Although this paper is only focused on insuring the flow of sensitive processes, the consideration of other dimensions of the process as part of the insurance agreement is a promising future extension, e.g., the inclusion of a model describing the social relations between actors within the context of the process. In that sense, the use of Process Mining techniques could also become a powerful tool, e.g., the Social Network Miner [20] can be applied to discover a model of the social network.

Acknowledgment

The authors would like to thank Prof. Dr. Josep Carmona and Ramon Xuriguera for their help and useful suggestions. The authors will like also to thank Dr. Sven Wohlgemuth and Prof. Dr. Noboru Sonehara for their comments and discussions. This work has been partially supported by the Spanish Ministerio de Educación (AP2009-4959) and by the NII International Internship Program.

References

[1] M. Dumas, W. M. P. van der Aalst, and A. H. ter Hofstede, *Process-Aware Information Systems: Bridging People and Software Through Process Technology*. Wiley, 2005.

[2] A. Adriansyah, B. F. van Dongen, and W. M. P. van der Aalst, "Conformance checking using cost-based fitness analysis," in *EDOC*. IEEE Computer Society, 2011, pp. 55–64.

[3] R. P. J. C. Bose and W. M. P. van der Aalst, "Analysis of patient treatment procedures," in *Business Process Management Workshops (1)*, ser. Lecture Notes in Business Information Processing, F. Daniel, K. Barkaoui, and S. Dustdar, Eds., vol. 99. Springer, 2011, pp. 165–166.

[4] W. M. P. van der Aalst, *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

[5] W. M. P. van der Aalst, A. Adriansyah, A. K. A. de Medeiros, F. Arcieri, T. Baier, T. Blickle, R. P. J. C. Bose, P. van den Brand, R. Brandtjen, J. C. A. M. Buijs, A. Burattin, J. Carmona, M. Castellanos, J. Claes, J. Cook, N. Costantini, F. Curbera, E. Damiani, M. de Leoni, P. Delias, B. F. van Dongen, M. Dumas, S. Dustdar, D. Fahland, D. R. Ferreira, W. Gaaloul, F. van Geffen, S. Goel, C. W. Günther, A. Guzzo, P. Harmon, A. H. M. ter Hofstede, J. Hoogland, J. E. Ingvaldsen, K. Kato, R. Kuhn, A. Kumar, M. L. Rosa, F. M. Maggi, D. Malerba, R. S. Mans, A. Manuel, M. McCreesh, P. Mello, J. Mendling, M. Montali, H. R. M. Nezhad, M. zur Muehlen, J. Muñoz-Gama, L. Pontieri, J. Ribeiro, A. Rozinat, H. S. Pérez, R. S. Pérez, M. Sepúlveda, J. Sinur, P. Soffer, M. Song, A. Sperduti, G. Stilo, C. Stoel, K. D. Swenson, M. Talamo, W. Tan, C. Turner, J. Vanthienen, G. Varvaressos, E. Verbeek, M. Verdonk, R. Vigo, J. Wang, B. Weber, M. Weidlich, T. Weijters, L. Wen, M. Westergaard, and M. T. Wynn, "Process mining manifesto," in *Business Process Management Workshops (1)*, ser. Lecture Notes in Business Information Processing, F. Daniel, K. Barkaoui, and S. Dustdar, Eds., vol. 99. Springer, 2011, pp. 169–194.

[6] T. Murata, "Petri Nets: Properties, analysis and applications," *Proceedings of the IEEE*, pp. 541–580, Apr. 1989.

[7] G. Keller, M. Nüttgens, and A. Scheer, "Semantische Prozeßmodellierung auf der Basis Ereignisgesteuerter Prozeßketten," *Veröffentlichungen des Instituts für Wirtschaftsinformatik*, vol. 89, 1992.

[8] S. White and D. Miers, *BPMN Modeling and Reference Guide: Understanding and Using BPMN*. Future Strategies Inc., 2008. [Online]. Available: http://www.bpmnreferenceguide.com/

[9] W. van der Aalst, A. Adriansyah, and B. van Dongen, "Causal nets: a modeling language tailored towards process discovery," in *CONCUR*, 2011, pp. 28–42.

[10] A. Rozinat and W. M. P. van der Aalst, "Conformance checking of processes based on monitoring real behavior," *Inf. Syst.*, vol. 33, no. 1, pp. 64–95, 2008.

[11] J. Muñoz-Gama and J. Carmona, "A fresh look at precision in process conformance," in *BPM*, ser. Lecture Notes in Computer Science, R. Hull, J. Mendling, and S. Tai, Eds., vol. 6336. Springer, 2010, pp. 211–226.

[12] W. M. P. van der Aalst, K. M. van Hee, J. M. E. M. van der Werf, and M. Verdonk, "Auditing 2.0: Using process mining to support tomorrow's auditor," *IEEE Computer*, vol. 43, no. 3, pp. 90–93, 2010.

[13] M. Jans, B. Depaire, and K. Vanhoof, "Does process mining add to internal auditing? an experience report," in *BMMDS/EMMSAD*, ser. Lecture Notes in Business Information Processing, T. A. Halpin, S. Nurcan, J. Krogstie, P. Soffer, E. Proper, R. Schmidt, and I. Bider, Eds., vol. 81. Springer, 2011, pp. 31–45.

[14] R. Accorsi and T. Stocker, "On the exploitation of process mining for security audits: The conformance checking case," in *SAC*. ACM, 2012.

[15] W. M. P. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 9, pp. 1128–1142, 2004.

[16] J. M. E. M. van der Werf, B. F. van Dongen, C. A. J. Hurkens, and A. Serebrenik, "Process discovery using integer linear programming," in *Petri Nets*, ser. Lecture Notes in Computer Science, K. M. van Hee and R. Valk, Eds., vol. 5062. Springer, 2008, pp. 368–387.

[17] J. Munoz-Gama and J. Carmona, "A General Framework for Precision Checking," *International Journal of Innovative Computing, Information and Control (IJICIC)*, vol. 8, no. 7, Jul. 2012.

[18] R. P. J. C. Bose, W. M. P. van der Aalst, I. Zliobaite, and M. Pechenizkiy, "Handling concept drift in process mining," in *CAiSE*, ser. Lecture Notes in Computer Science, H. Mouratidis and C. Rolland, Eds., vol. 6741. Springer, 2011, pp. 391–405.

[19] J. Muñoz-Gama and J. Carmona, "Enhancing precision in process conformance: Stability, confidence and severity," in *CIDM*. IEEE, 2011, pp. 184–191.

[20] W. M. P. van der Aalst, H. A. Reijers, and M. Song, "Discovering social networks from event logs," *Computer Supported Cooperative Work*, vol. 14, no. 6, pp. 549–593, 2005.

[21] F. Daniel, K. Barkaoui, and S. Dustdar, Eds., *Business Process Management Workshops - BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I*, ser. Lecture Notes in Business Information Processing, vol. 99. Springer, 2012.