

## **A HYBRID PROCESS-MINING APPROACH FOR SIMULATION MODELING**

Waleed Abohamad  
Ahmed Ramy  
Amr Arisha

3S Group, College of Business  
Dublin Institute of Technology (DIT)  
Aungier Street, Dublin 2  
Dublin, IRELAND

### **ABSTRACT**

This paper presents a hybrid Modeling and Simulation framework to address business process challenges. The framework has integrated process mining techniques in the conceptual modeling phase to support developing simulation models that are unbiased and close reflection of reality in a timely manner. The hybrid approach overcomes the pitfalls of traditional conceptual modeling by using process mining techniques to discover valuable process knowledge from the analysis of event logs. The proposed hybrid framework has been applied to an Emergency Department (ED) in order to identify performance bottlenecks and explore improvement strategies in an attempt to meet national performance targets. A large number of unique process flows (i.e. patient pathways) within the ED were uncovered and deviations from process guidelines were accurately identified. Results show that unblocking of ED outflows have a significant impact on patients length of stay (over 80% improvement) rather than increasing the ED physical capacity.

### **1 INTRODUCTION**

Rapid advances in Information Technology (IT) and diversity in products and services have a substantial effect on the acceleration of dynamics in business systems (e.g., manufacturing systems, healthcare systems ...etc). Understanding and improving the performance of such systems is a challenging task which is imputable to a high level of uncertainty, conflicting objectives, lack of necessary information, and a large number of constraints, inter-connected components and decision variables.

Addressing these challenges requires a thorough understanding of system constraints, which can be an overwhelming task, given the high levels of uncertainty and interdependence. Simulation modeling has contributed to the understanding of different levels of complexity within business processes (Wagner, Seck, and McKenzie 2016). Discrete-event simulation (DES), agent-based simulation (ABS), Monte Carlo simulation (MCS), and system dynamics (SD) are widely used in modeling complex processes (Katsaliaki and Mustafee 2011, Abo-Hamad and Arisha 2013). Each of these simulation techniques addresses a particular level of complexity within the system. Modelers are often challenged to accurately model the system complexity in order to provide managers with effective results resembling reality (Lynch et al. 2014). Research has indicated that the use of hybrid simulation would improve the capabilities of simulation solutions (Viana 2014, Zulkepli and Eldabi 2015, Arisha and Rashwan 2016). Despite the growing number of hybrid simulation models, there are still a number of challenges that have not been addressed adequately. For instance, the focus of most reported hybrid simulation cases is mainly on model implementation phase (i.e. combining DES and ABS) rather than considering hybridization in other aspects such as conceptual modeling. Although conceptual modeling has been identified as potentially the most significant stage of any simulation study, it is the most underestimated aspect (Law 1991). The hybrid approach should be extended to conceptual modeling phase to allow accurate modeling

of other aspects of the business process (e.g. human behavior and variability) that have always been challenging (Daellenbach 2001). A hybrid Modeling and Simulation (M&S) study is characterized by using of methods from fields such as Operations Research (OR), Computer Science, Systems Engineering, and Information Systems in the wider simulation study, not only in the implementation phase (Powell and Mustafee 2016). The focus of this paper is on the hybridization in the conceptual modeling phase of M&S studies.

Conceptual modeling is the abstraction (i.e. non-software specific description) of a simulation model from the part of the real-life process it is representing (Robinson 2015). Conceptual models are typically created based on insights from information sources such as process documentation, interviews with stakeholders and experts and direct process observations (Rozinat et al. 2009). Developing conceptual models using these traditional sources is time-consuming, and arguably the longest stage of any M&S project. This lengthy process significantly affects the validity and effectiveness of the M&S study recommendations. Furthermore, the obtained information from these traditional sources might present a biased and inaccurate view of the process. Process documentation might deviate from real-life process behavior (Mărușter and van Beest 2009). Interviews with business experts can result in contradictory information as the human perception in general, tends to be biased to a certain extent (Pronin 2007). Observational data, in their turn, might not be reflective of reality as staff members tend to modify aspects of their normal behavior in response to their awareness of being observed (McCarney et al. 2007). Relying only on these traditional sources contribute to a discrepancy between the actual behavior of the process and the developed simulation model. Therefore, valuable insights can be obtained by analyzing data outlining the actual behavior of the process. Business processes nowadays are increasingly supported by information systems that produce event logs. These logs record highly relevant information on the actual behavior of the process under consideration and can be analyzed using process mining techniques (van der Aalst 2011). Process Mining (PM) techniques have emerged to handle the discrepancy between how processes operate in the real world and process guidelines (Turner et al. 2012). Knowledge discovered from the analysis of event logs can support modeling activities in the conceptual modeling phase. Such process knowledge can then be cross-checked against traditional information sources. Moreover, the process mining output can be validated by staff members and experts due to errors or inaccuracies in recording data.

This paper presents a hybrid Modeling and Simulation framework that integrates process mining techniques in the conceptual modeling phase to support developing simulation models that are unbiased and close reflection of reality in a timely manner. The primary objective is to overcome the pitfalls of traditional conceptual models by using process mining to support modeling activities. The proposed framework is demonstrated on a real-world case study of an emergency department of one of the leading hospitals in Ireland.

## **2 LITERATURE REVIEW**

Process mining aims at knowledge acquisition from enterprise information systems for the sake of reconstruction of business process models (Rojas et al. 2016). There are four perspectives of application of PM: control-flow discovery; performance analysis; conformance checking and enhancement; and resource organizational structure (van der Aalst 2011). Control-flow perspective explores the execution order of process activities and automatically discovers the underlying process structure without using any further information (Ferreira 2012). The performance perspective analyzes the execution timing of activities and process bottlenecks (Mans et al. 2008). On the other hand, Conformance process techniques can be used to check the compliance of process guidelines against the actual practices that are recorded as an event log (Rovani et al. 2015). Enhancement process mining can enrich existing process models with conformance-related diagnostics (Centobelli, Converso, and Gallo 2015). Finally, the organizational perspectives focuses on analyzing collaboration between resources (Mans, van der Aalst, and Vanwersch 2015).

The most commonly used PM techniques for control-flow discovery are Heuristics Miner, Fuzzy Miner and Trace Clustering (Rojas et al. 2016). Heuristics Miner has the advantage of dealing with noise in event logs and hence generating robust business process models (Weijters, van der Aalst, and Medeiros 2006). Fuzzy Miner, on the other hand, can be configured to generate multiple process control models at different levels of details suitable for unstructured processes such as healthcare processes (Günther and Aalst 2007). Similarly, Trace Clustering techniques have the capability of discovering simpler structured process models by subdividing into partitions (Song, Günther, and van der Aalst 2009). The sophistication of PM techniques has witnessed a progressive increase in the last few years. Such progression has led to the emergence of commercial PM mining tools such as ProM and DISCO capable of solving existing mining problems such as event log noise and duplicate activities (Turner et al. 2012). A recent extensive literature review (more than 74 articles) conducted by Rojas et al. (2016) on the application of process mining has identified research opportunities within this area. Interestingly, only one paper in their review has attempted to combine process mining with simulation models for an outpatient clinic model (Zhou, Wang, and Li 2014). Although preliminary research efforts on this integration have been done (Augusto et al. 2016), there is no clear frameworks or guidelines on how PM techniques can be used to support the construction of simulation models. In an attempt to bridge this gap, Martin, Depaire, and Caris (2016) suggested a structured overview of business process simulation and how PM can support modeling tasks. However, no practical demonstration or evidence was provided to support these guidelines. Most of the reported studies neglect the internal structure of simulation models and its relationship with the output of PM algorithms. Moreover, current research focuses mostly on a single modeling aspect with little or no considerations given to other aspects of the modeling process. The main contribution of this paper is the integration of process mining techniques in the conceptual modeling phase to support developing simulation models that are unbiased and close reflection of reality in a timely manner. The framework is demonstrated using a real-life healthcare case study of an Emergency department in Dublin – Ireland.

### **3 A HYBRID PROCESS – BASED SIMULATION MODELING FRAMEWORK**

#### **3.1 Proposed Framework**

The proposed hybrid framework consists of three phases; Formulation and Data Analysis, Conceptual Modeling, and Model Development (Figure 1). Typically, each business process has its own database that supports its functions. The input data of PM is an event log with each row represents an event that is executed in the process. Each event includes the following data elements: Case ID, Activity Name, Originator, a timestamp, and other case specific data (i.e. case attributes). Each event refers to a distinct step in the process (i.e. an activity). Case ID associates this activity with a particular case (e.g. customer ID or order ID). Originator refers to human resources performed the activity while the timestamp specifies when this event occurred. More information on the case are sometimes recorded in event logs which are specific data elements referring to case attributes. In simulation context, mining these event logs can significantly support essential modeling tasks associated with the construction of process simulation models. These tasks are modeling flow units (arrival rates and job types), modeling process activities (types, duration, precedence relations and control-flow, activity networks and paths), and resources (roles, allocation, and utilization).

Regarding modeling flow units, each case corresponds to a flow unit or a job while events represent the various activities that are performed by human resources to complete such jobs. Because each event in the log typically carries a timestamp, arrival rates of jobs can be determined from events that refer to initial activities such as system entrance or registration. The analysis of case attributes can reveal valuable knowledge regarding the different types of jobs that are presented to the business process. Process mining can guide activity modeling tasks in several ways. Flow units have to pass through a network of activities in order to be transformed from inputs to outputs.

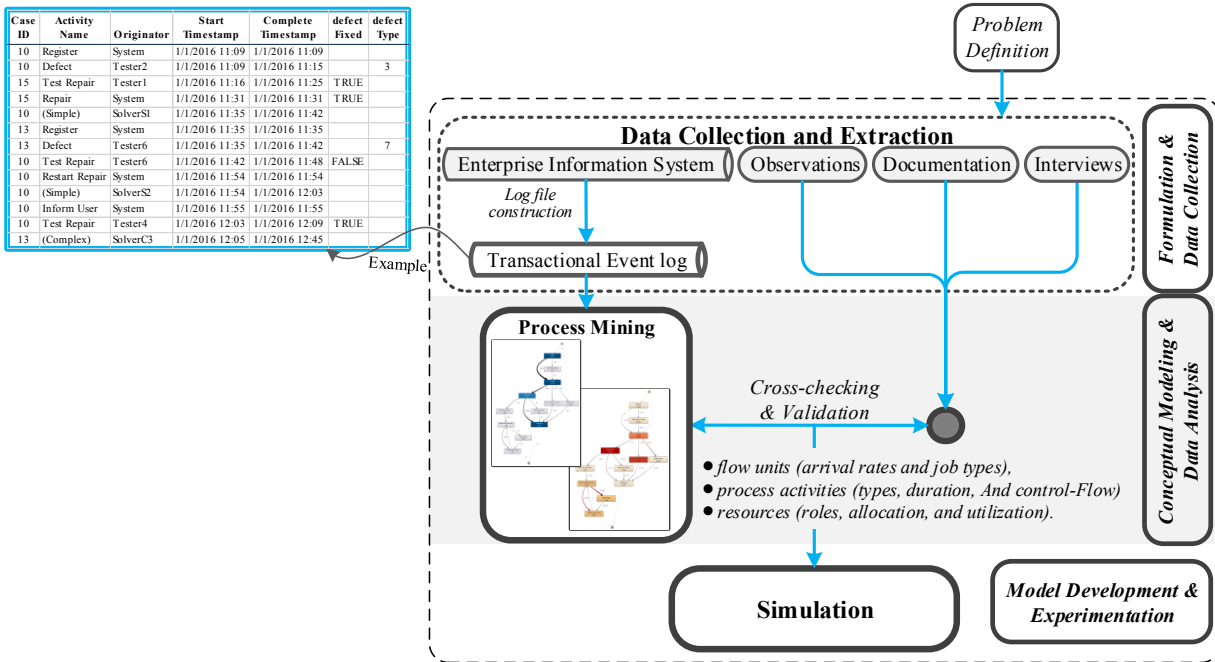


Figure 1: An overview of the proposed process-mining-based modeling & simulation framework.

Activity names and corresponding timestamps in the event log can be analyzed to define all relevant activities and their precedence relationships and hence determine which paths a job can follow while moving throughout the process model. Control-flow discovery techniques, as discussed in the literature, retrieve the process structure by mining process activities and their flow sequence from an event log. Due to the unstructured nature of business processes, especially service operations, the selection of the process mining discovery algorithm is crucial. The Fuzzy miner (Günther and Aalst 2007) is more suitable for unstructured processes and simulation and managerial purposes. Other algorithms such as Alpha-algorithm produces a very complicated process models which present all process flow details without distinguishing what is crucial and what is unnecessary. On the other hand, fuzzy miner algorithm analyze complex processes at different levels of granularity and provides meaningful abstraction and various views of the process. Providing these granularity levels is achieved by applying two fundamental metrics: significance and correlation. The significance metric assesses the relative importance of a precedence relation between two activities, i.e. the more often two activities are observed after one another, the more significant their precedence relationship. The correlation metric indicates how closely two activities are following each other. Therefore, fuzzy mining produces different levels of aggregation and abstraction by adjusting the discovered control-flow using these metrics. As resources tend to be recorded for each event (i.e. originator), process mining algorithms can support tasks related to resource modeling where profound insights regarding resources allocation, roles, and utilization can be acquired.

After cross-checking and validation, the knowledge obtained from mining the process event log are to be used in the model development and experimentation phase. Depending on the model scope and objective, the model can be implemented using a single simulation technique such as Monte Carlo simulation (MCS), discrete event simulation (DES), system dynamics (SD) and agent-based simulation (ABS) or a hybrid between these simulation techniques.

### 3.2 Emergency Department Case Study

Overcrowding in emergency departments (EDs) has become a significant international crisis that negatively affects patient safety, quality of care, and patient satisfaction. This project was a joint effort

involving hospital staff and our Institute research team. The hospital is an acute public hospital in North Dublin. This 570-bed hospital provides a variety of health-care services, with a 24h ED that receives over 55,000 patients annually. The ED could not meet the national target of 6h average length of stay (LOS) for patients. The figures show clear evidence of overcrowding, with an average of 17% of patients choosing to leave before being seen by the ED clinician. The report indicated that the average time from ED registration to discharge was 9.1h and the average duration from registration to acute admission was 18h (i.e., three times higher than the national metric). Apparently, patients who are admitted will usually experience longer LOS times than those who are discharged due to delays between admission referral by an ED doctor, the allocation of bed, and patient transfer time. The proposed framework is demonstrated on the ED to develop a representative simulation model with the aim to identify performance bottlenecks and explore improvement strategies to meet the HSE targets.

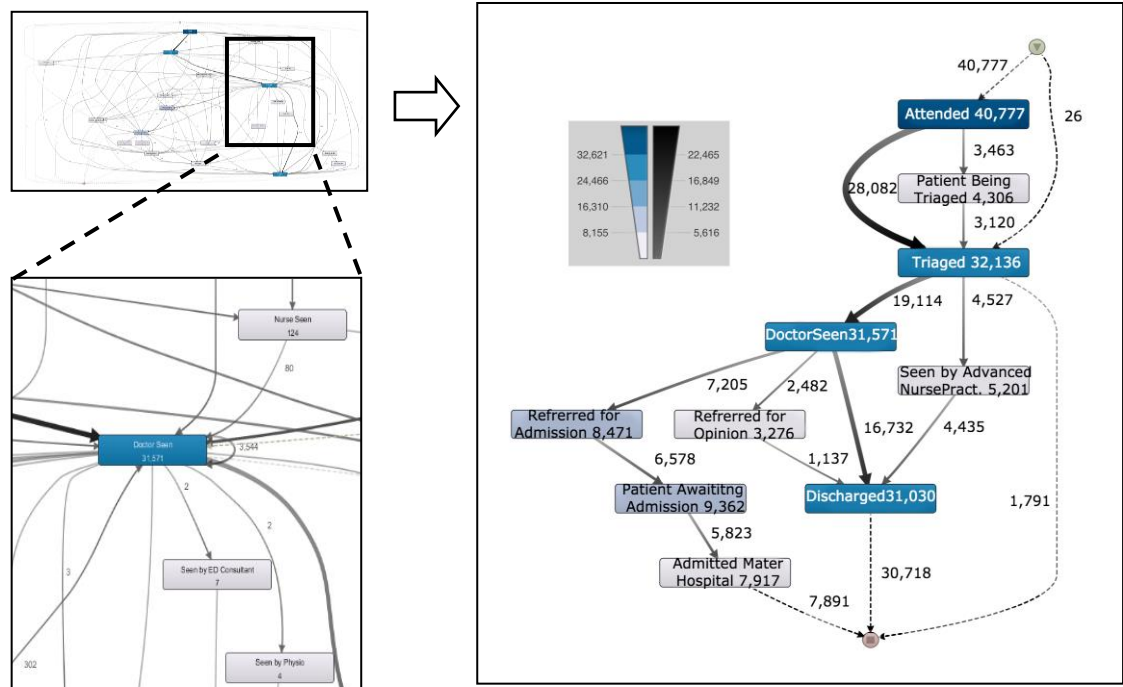
### **3.2.1 Data Collection and Analysis**

A real-time patient tracking information system was used to track the patient's journey within the ED. The hospital managers have provided a one-year historical data with anonymous patients' records. The dataset was provided in an event log structure with a total of 229,971 event logs representing 40,777 patients. Each record in the table represents an event (i.e. one process stage of the patient journey in the ED) with the following attributes (patient ID, Triage Category, Presenting Complaint, Date of Birth, Gender, Event ID, Tracking Step Name, Tracking Step Date Time, Location, Staff). Events with the same name, patient ID and timestamp were removed which resulted in a total of 210,180 records in the ED event log. The following sections represent the findings of applying process mining techniques in the conceptual modeling phase. Plugins in the ProM and Disco software packages were used to derive these findings.

### **3.2.2 Conceptual Modeling**

Upon their arrival, patients are assigned a clinical priority (triage category) according to the Manchester Triage System (MTS) (Cronin 2003). The MTS uses a five-level scale for classifying patients according to their care requirements; immediate, very urgent, urgent, standard, and non-urgent. Immediate and very urgent patients represent 15%, urgent patients (triage category 3) represent the largest group of attendees to the ED annually (59% average), while standard and non-urgent patient 26% of all patients. As advised by ED consultants, the analysis of these patients' groups is critical as each group of patients can have a different journey within the ED and hence a different pathway (i.e. control-flow).

The main building blocks of the process control-flow (patient pathways) are the activities that patients go through in the ED. Twenty-two different activities within the ED were identified from the event log data. The fuzzy miner has then applied on the event log data to discover the first top-level control-flow of the overall ED (Figure 2a). The rectangles in Figure 2 refers to activities with the number inside each rectangle shows how many times an activity has been performed (the larger the number inside the rectangle, the darker the shade of the rectangle color).The arcs represent the dependency between activities and the frequency (i.e. number of patients) of occurrences are reported beside each arc (the thickness of the arc is proportional to its frequency). The resulted control-flow is complicated and hard to interpret due to the high variances in patients' pathways. Such complexity is an indication of great divergence in patients pathways. The fuzzy miner allows observing complex processes at different levels of granularity by applying two fundamental metrics: significance and correlation. Fuzzy miner algorithm has therefore applied to the event log to reduce the process complexity by hiding less frequents process flows and present the only process flows with high frequency (Figure 2b). The count inside each rectangle in Figure 2 shows how many times an activity has been executed (e.g. activity 'Doctor Seen' occurred 31,571 times), while, the count on the arc represents the co-occurrence frequency between any two activities.



a) ED process model with all activities and paths. b) The process model with main activities and paths.

Figure 2: The discovered patient flow model of the emergency department.

For example, the co-occurrence frequency between ‘Doctor Seen’ and ‘Referred for Admission’ is 7,205. Further analysis of the complex process flow of patients revealed that there are 1,984 unique control-flow (i.e. patient pathways) (Table 1).

Table 1: Discovered control-flow of patients within the ED.

Pathway	Cases	Relative frequency	Events	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	...	Step 13
1	9374	23%	4	Attended	Triaged	Doctor Seen	Discharged				
2	3406	8%	4	Attended	Triaged	Seen by Advanced Nurse Practitioner	Discharged				
3	3264	8%	3	Attended	Doctor Seen	Discharged					
4	3005	7%	3	Attended	Triaged	Discharged					
5	1949	5%	6	Attended	Triaged	Doctor Seen	Referred for Admission	Patient Awaiting Admission	Admitted to Hospital		
.	.	.	.	.	.	.	.	.	.	.	.
1984	1	0.002%	6	Attended	Did Not Answer	Triaged	Referred for Opinion	Doctor Seen	Admitted to Hospital		

Over 60% of these control-flows are one-off path and only 31 paths account for 80% of ED patients. Therefore, the remaining low frequent process flows were filtered out in order to model the common behavior of the ED. The fuzzy miner was then applied on the resulted 31,447 patients to derive final top-level process map of the ED (Figure 3a).

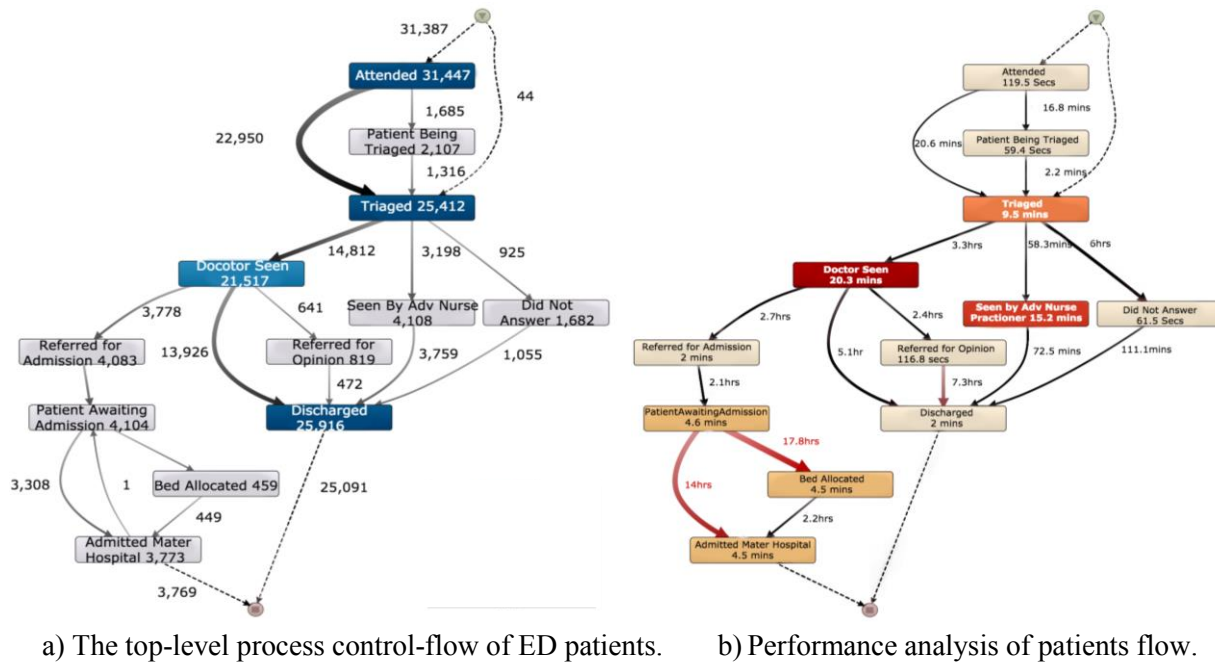
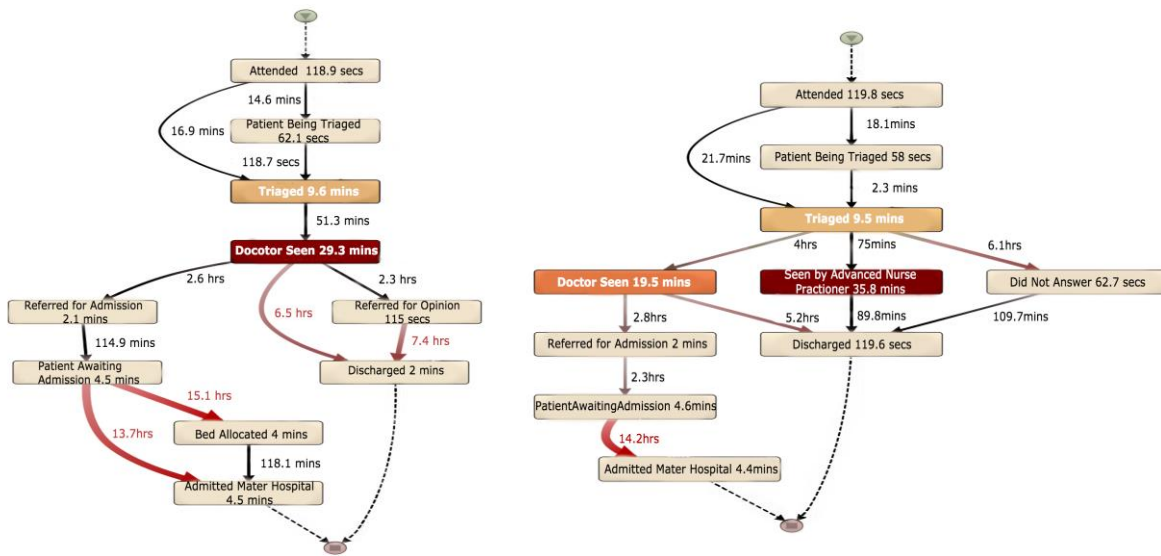


Figure 3: The top-level process control-flow and performance analysis of ED patients.

The analysis of timestamp data associated with activities has uncovered several performance bottlenecks. Numbers inside rectangles in Figure 3b represent the average activity duration while numbers on the arcs represent the waiting time between any two activities. The average LOS for all patients from arrival to departure (whether patients discharged or admitted to the hospital) is 9.1h which is 3 hours above the national target. Patients awaiting admission are experiencing an average LOS of 18h, with an average 14h of waiting for bed allocation. After triaging, patients have to wait 3.3h on average to be seen by a physician and 5.1h to be discharged. The main bottleneck activities in the ED are the “Seen by Doctor” and “Patients awaiting admission”. The process flow of patients was then analyzed at a more fine-grained level to analyze the leading causes of these bottlenecks. The “Triage Category” attribute was used to divide patients into three groups; Immediate and very urgent, non-urgent and standard, and urgent patients. The process flow of each patient type was then derived and analyzed (Figure 4). There are obvious variances in the associated control-flow for patients with different urgency levels. The majority of Immediate and Very Urgent patients are admitted to the hospital with an average waiting time of 13.7h between the request and allocation of inpatient bed (Figure 4a). On the other hand, Standard and Non-Urgent patients have a shorter flow with 5.1h average LOS. “Did not Answer” activity refers to patients who left the ED after being triaged without waiting to be investigated by a physician. 9.6% of patients left the ED after being triaged due to the prolonged waiting times. Urgent patients represent almost 60% of all patients with 10h average LOS. This patient group are presented with a broad range of complaints with 27% are referred for admission and the remaining are discharged with an average waiting time of 5.2h (Figure 4b). Insights from this analysis enabled the ED decision makers to identify the bottlenecks for each patients group.

The department has 12 trolleys for emergency patients, six trolleys for ambulatory care, two rapid assessment triage bays and two triage rooms. The ED has in total eleven nurses during the day and nine nurses at night with a variety of nursing duties (Advanced Nurse Practitioner (ANP), triage nurse, resuscitation nurse, respiratory nurse, majors/minors nurse, and healthcare assistant). Physicians are divided into registrar/specialist registrar and Senior House Officer (SHO). Location and staff members who performed each activity is recorded in the event log of the ED.



a) Immediate & Very Urgent patients.

b) Urgent patients.

Figure 4: Performance analysis for patients with different triage categories.

Analysis of these valuable attributes is summarized in Table 2. The analysis reveals that 68% of the Triage activity took place in the triage room and the activity was mostly performed by RGN (77%). According to the clinical guideline, the Triage process should be held every single time in the Triage room. Therefore, the process does not comply with the guidelines.

Table 2: Resource analysis for the main activities in patient pathways.

		Triaged	Doctor Seen	Seen by ANP	Referred for Admission	Referred for Opinion	Discharged	Admitted to Hospital
Medical Staff	ED Consultant				46%		21%	95%
	ANP			24%		4%	3%	3%
	SPR		15%		11%	10%	10%	1%
	RGN	77%	0%					1%
	ADN			76%		13%	11%	
	SHO		58%		36%	51%	30%	
	Registrar		20%		7%	10%	12%	
	CNM	13%				5%	9%	
Physical	Intern		7%			7%	4%	
	Majors Area (9)	6%	55%		65%	57%	71%	72%
	Resuscitation Room (3)	6%	18%		31%	29%	4%	24%
	Ambulatory Care Unit (6)	3%	11%	57%	3%	14%	23%	2%
	Rapid Assessment Triage (2)	17%	9%					
	Triage Room (2)	68%						

ANP: Advanced Nurse Practitioner  
 SHO: Senior Hospital Officer

SPR: Specialist Registrar  
 CNM: Clinical Nurse Manager

RGN: Registered General Nurse  
 ADN: Associate Degree in Nursing

Similarly, the “Doctor seen” activity is performed by SHOs (58%) and registrars (20%) in the major area in the ED (55%) and in the Resuscitation room (18%). These figures highlight the actual locations



within the ED where physician examine patients. These insights help the ED managers to better understand the actual allocation of staff and resources within the ED.

### 3.2.3 Model Development and Experimentation

Knowledge discovered from the analysis of event logs have been cross-checked against information obtained from interviews and process documentation. Upon validation, this knowledge was used to develop a discrete event simulation model using AnyLogic simulation platform. Simulation results of the baseline represent runs for one year (Table 3). These results have a satisfactory level comparing to the KPIs extracted from the ED historical data. Average LOS is 9h vs 9.1h, average waiting time for triage is 19.9min vs 20.6min. Average waiting for first clinical contact is 190.8min and 57.5min vs 198min and 58.3min respectively. Percentage of patients left without being seen (L.W.B.S) is 10% vs 9.6%. Accordingly, simulation results have a satisfactory level of accuracy which is at the same time an indication of the accuracy and credibility of the process knowledge derived from the ED event logs.

Based on the simulation results and process mining outcomes, the main bottlenecks in the workflow are the “Seen by Doctor” and “Patients waiting admission” activities. Therefore, the simulation scenarios tested were the impact of variation in medical staffing (Scenario1 – an increase of 25%), increasing clinical assessment space (Scenario 2 – a 50% increase), and finally assessing the impact of incorporating a ‘zero-tolerance’ policy (Scenario 3). According to ED managers, the goal of the ‘zero-tolerance’ policy is to assess the performance of the ED if the LOS of all patients complies with the HSE 6 h target. The aim is to identify the underlying factors that contribute to the unacceptable overcrowding status of the ED. Scenario 3 is simulated dismissing patients from the ED model who are waiting to be admitted to a hospital bed, and their LOS exceeds 6h. The simulation results of these scenarios are shown in Table 3.

Table 3: Simulation results of baseline model and scenarios.

	<b>Baseline</b>	<b>Scenario 1</b>	<b>+/- (%)</b>	<b>Scenario 2</b>	<b>+/- (%)</b>	<b>Scenario 3</b>	<b>+/- (%)</b>
<i>Avg. W. T. Triage (min)</i>	19.9	19.6	-1%	19.8	0%	20.4	3%
<i>Avg. W. T. Doctor (min)</i>	<b>190.8</b>	186.9	-2%	2.5	<b>-99%</b>	12.2	<b>-94%</b>
<i>Avg. W. T. ANP (min)</i>	57.5	60.7	6%	65.6	14%	62.5	9%
<i>Avg. # of Pts in W.R.</i>	13.4	13.1	-2%	1.2	<b>-91%</b>	1.7	<b>-87%</b>
<i>% of Pts L.W.B.S</i>	10%	9%	<b>-3%</b>	0.01%	<b>-100%</b>	0.01%	<b>-100%</b>
<i>Avg. LOS Discharged Pts (h)</i>	7.5	7.5	-1%	5.2	-30%	4.1	<b>-45%</b>
<i>Avg. LOS Admitted Pts (h)</i>	<b>17.8</b>	17.7	0%	15.9	-11%	2.8	<b>-84%</b>
<i>Avg. LOS All Pts (h)</i>	9.0	9.0	0%	7.0	-23%	3.9	-56%

Results show that access block have a significant impact on prolonged waiting times for patients. These insights stress the importance of streamlining patient flow across hospital units. Therefore, the hospital management team has suggested three initiatives to implement the recommendations of this study. The first initiative is based on executing an escalation plan that includes placing of additional beds on inpatient wards for moving patients quickly who are waiting in the ED. While the second initiative is contracting cheaper beds in community care for elderly patients – this will shorten the average LOS in across the whole hospital, and consequently, more beds will be available for ED admissions. The third one is suggesting that the hospital executives initiate lean training across all departments in the hospital. The aim of this lean training is to increase the coordination level between hospital units to achieve better utilization of hospital resources. A key limitation of proposed recommendations is the financial implication of any decision. Incorporating costing was not possible in this project due to lack of cost related information to support the analysis phase and also the variability in the cost model used in various public hospital in Ireland.

## **4 DISCUSSION**

Based on the findings of applying process mining techniques on the ED event logs, a large number of unique control-flows were uncovered. Identifying these process flows would have been impossible task using traditional information sources for conceptual modeling. Process flows with very low frequency can give profound insights for analyzing any exceptional behavior of processes. Furthermore, deviations from process guidelines were uncovered during the resource analysis. However, process knowledge discovered by process mining techniques might lead to wrong conclusions due to errors and inaccuracy in recording data. Cross-checking with traditional information sources and validation has to take place prior to any further analysis. The output of applying PM on event logs can be used as an initial draft for the final process model and hence a significant time saving will be achieved compared to the time spent developing conceptual models relying on interviews and observations only. Nevertheless, the developed models are more accurate representation of the real-life process.

Despite these benefits of the application of PM techniques for developing simulation models, there are still significant research challenges are still ahead. Further research efforts are needed to capture the interaction between cases from event logs instead of focusing on a serious of independent cases. Also, new PM techniques are required to extract simulation related information such as queues disciplines, resource schedules, and activity durations with limited timestamp information. Finally, for more sustainable applications of process mining, new methods are to be developed to incorporate domain knowledge into the mining process for optimal discovery of models that accurately describe business processes.

## **5 CONCLUSION**

Conceptual models are typically created based on insights from information sources such as process documentation, interviews, and direct observations. Developing conceptual models using these traditional sources is not only time-consuming but also might present a biased and inaccurate view of the process. Relying only on these traditional sources contribute to a discrepancy between the actual behavior of the process and the developed simulation model. Valuable insights can be obtained by analyzing event logs of the process. These logs record highly relevant information on the actual execution of process activities and can be analyzed using Process Mining (PM) techniques. PM techniques aim to extract and discover hidden knowledge from enterprise information systems for the sake of reconstruction of business process models. This paper presented a hybrid Modeling and Simulation framework that integrates PM techniques in the conceptual modeling phase to support modeling activities. The proposed framework was demonstrated on a case study of an Emergency Department with event logs of 210,000 records. Resulted process model revealed that there are almost 2000 unique patient flows with only 30 flows account for 80% of patients. Several deviations from process guidelines were also uncovered regarding the allocation and tasks of staff members. The experimentation results showed that unblocking of ED outflows by in-patient bed management has the highest impact on patients LOS at every stage, especially among patients who are supposed to be discharged directly after their ED care (45% reduction in length of stay) with 87% reduction in patients waiting times.

It worth mentioning that process knowledge discovered by process mining techniques might lead to wrong conclusions due to errors and inaccuracy in recording data. Discovered models should be cross-checked and validated prior to developing simulation models. Further research efforts are still needed to capture the interaction between cases from event logs, to extract simulation related information (queues disciplines, resource schedules, and activity durations), and to incorporate domain knowledge for optimal discovery of models.

## REFERENCES

- Abo-Hamad, W., and A. Arisha. 2013. "Simulation-Based Framework to Improve Patient Experience in an Emergency Department". *European Journal of Operational Research* 224: 154–66.
- Arisha, A., and W. Rashwan. 2016. "Modeling of Healthcare Systems: Past, Current and Future Trends". In *Proceedings of the 2016 Winter Simulation*, edited by T. Roeder, P. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. Chick, 1523–34. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Augusto, V., X. Xie, M. Prodel, B. Jouaneton, and L. Lamarsalle. 2016. "Evaluation of Discovered Clinical Pathways Using Process Mining and Joint Agent-Based Discrete-Event Simulation". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. Roeder, P. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. Chick, 2135–46. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Centobelli, P., G. Converso, and M. Gallo. 2015. "From Process Mining to Process Design: A Simulation Model to Reduce Conformance Risk". *Engineering Letters* 23: 1–11.
- Cronin, J. 2003. "The Introduction of the Manchester Triage Scale to an Emergency Department in the Republic of Ireland". *Accident and Emergency Nursing* 11: 121–25.
- Daellenbach, H. 2001. "Hard OR, Soft OR, Problem Structuring Methods, Critical Systems Thinking: A Primer". In *Proceedings of the ORSNZ Conference Twenty Naught One*. Christchurch, New Zealand.
- Ferreira, D. 2012. "Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining". *Information Systems* 37: 99–116.
- Günther, C., and W. van der Aalst. 2007. "Fuzzy Mining—adaptive Process Simplification Based on Multi-Perspective Metrics". In *Proceedings of the International Conference on Business Process Management*, edited by G. Alonso, P. Dadam, and M. Rosemann, 328–43. Springer Berlin Heidelberg.
- Katsaliaki, K., and N. Mustafee. 2011. "Applications of Simulation within the Healthcare Context." *Journal of the Operational Research Society* 62: 1431–51.
- Law, A. 1991. "Simulation-Models Level of Detail Determines Effectiveness". *Industrial Engineering* 23: 16–18.
- Lynch, C., J. Padilla, S. Diallo, J. Sokolowski, and C. Banks. 2014. "A Multi-Paradigm Modeling Framework for Modeling and Simulating Problem Situations". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller, 1688–99. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Mans, R., H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglioni, and W. van der Aalst. 2008. "Process Mining Techniques: An Application to Stroke Care". *Studies in Health Technology and Informatics* 136:573–78.
- Mans, R., W. van der Aalst, and R. Vanwersch. 2015. *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*. Springer.
- Martin, N., B. Depaire, and A. Caris. 2016. "The Use of Process Mining in Business Process Simulation Model Construction". *Business & Information Systems Engineering* 58: 73–87.
- Mărușter, L., and N. van Beest. 2009. "Redesigning Business Processes: A Methodology Based on Simulation and Process Mining Techniques." *Knowledge and Information Systems* 21: 267–97.
- McCarney, R., J. Warner, S. Iliffe, R. van Haselen, M. Griffin, and P. Fisher. 2007. "The Hawthorne Effect: A Randomised, Controlled Trial". *BMC Medical Research Methodology* 7:30.
- Powell, J. H., and N. Mustafee. 2016. "Widening Requirements Capture with Soft Methods: An Investigation of Hybrid M&S Studies in Health Care". *Journal of the Operational Research Society*, December, 1–12. doi:10.1057/s41274-016-0147-6.
- Pronin, E. 2007. "Perception and Misperception of Bias in Human Judgment". *Trends in Cognitive Sciences* 11: 37–43.

- Robinson, S. 2015. "A Tutorial on Conceptual Modeling for Simulation". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. Chan, I. Moon, T. Roeder, C. Macal, and M. Rossetti, 1820–34. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Rojas, E., J. Munoz-Gama, M. Sepúlveda, and D. Capurro. 2016. "Process Mining in Healthcare: A Literature Review". *Journal of Biomedical Informatics* 61: 224–36.
- Rovani, M., F. Maggi, M. de Leoni, and W. van der Aalst. 2015. "Declarative Process Mining in Healthcare". *Expert Systems with Applications* 42: 9236–51.
- Rozinat, A., R. Mans, M. Song, and W. van der Aalst. 2009. "Discovering Simulation Models". *Information Systems* 34: 305–27.
- Song, M., C. Günther, and W. van der Aalst. 2009. "Trace Clustering in Process Mining". In *Business Process Management Workshops: BPM 2008 International Workshops*, edited by D. Ardagna, M. Mecella, and J. Yang, 109–20. Springer Berlin Heidelberg.
- Turner, C., A. Tiwari, R. Olaiya, and X. Yuchun. 2012. "Process Mining: From Theory to Practice". *Business Process Management Journal* 18: 493–512.
- van der Aalst, W. 2011. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer Science & Business Media.
- Viana, J. 2014. "Reflections on Two Approaches to Hybrid Simulation in Healthcare". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller, 1585–96. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Wagner, G., M. Seck, and F. McKenzie. 2016. "Process Modeling for Simulation: Observations and Open Issues". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. Roeder, P. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. Chick, 1072–83. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Weijters, A., W. van der Aalst, and A. Medeiros. 2006. "Process Mining with the Heuristics Miner Algorithm". Technical Report No. 166, Technische Universiteit Eindhoven, Netherland.
- Zhou, Z., Y. Wang, and L. Li. 2014. "Process Mining Based Modeling and Analysis of Workflows in Clinical Care – A Case Study in a Chicago Outpatient Clinic". In *proceedings of the 11th IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 590–95. IEEE.
- Zulkepli, J., and T. Eldabi. 2015. "Towards a Framework for Conceptual Model Hybridization in Healthcare". In *Proceedings of the 2015 Winter Simulation Conference*, edited by W. Yilmaz, V. Chan, I. Moon, T. Roeder, C. Maca, and M. Rossetti, 1597–1608. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

## AUTHOR BIOGRAPHIES

**WALEED ABOHAMAD, Ph.D.** is a Senior Lecture in College of Business at Dublin Institute of Technology (DIT) and a Team Leader in 3S Group. His research interests include Modeling and Simulation, Optimization, Computational Intelligence, Machine Learning and Analytics. His email address is [waleed.abohamad@dit.ie](mailto:waleed.abohamad@dit.ie).

**AHMED RAMY** is a PhD Candidate in 3S Group – College of Business (DIT). He completed his BSc and MSc from Arab Academy for Science and Technology. His research interests include Scientometric Analysis, Quality Management, Knowledge Assessment and System Thinking. His email address is [ahmed.ramy@dit.ie](mailto:ahmed.ramy@dit.ie).

**AMR ARISHA, Ph.D.** is a Professor at Dublin Institute of Technology (DIT) and the Director of 3S Group. His research interests include analysis and optimization of complex dynamic systems in manufacturing and service industries. [amr.arisha@dit.ie](mailto:amr.arisha@dit.ie).