

PROCESS MINING

Lab Session 7

Hands-on Session on Business Process Simulation

Content

Laboratory Se	ession	2
	the BPMN Model	
2. Discover	the Simulation Parameters	
Case Arriva	ıl Rate	4
Resource a	llocation	∠
Timetable		5
Activity dur	ations and distributions	6
Branching	probabilities	8
3. Creation	of the Simulation Model	g
4. Simulatio	on	13
5. Final Exe	rcises	16
Solution of	Exercise 1	16
Solution of	Exercise 2	18



Laboratory Session

Business process simulation investigates process behaviors based on a simulation model (e.g., BPMN model) extended with additional information to define the different simulation parameters such as case arrival rate, resource allocation, timetable, activity durations and branching probabilities.

In this laboratory session, we build a simulation model and run it via the online BIMP simulator¹. The simulation model is constructed by annotating a BPMN model with the different simulation parameters.

The simulation model focuses on the same purchase order process (purchasingExample.xes) we used in the earlier laboratory sessions on PM4PY and Disco.

Figure 1 presents the main steps of our simulation study. We first convert the discovered Petri net into a BPMN model. Then we extract the simulation values using the tools ProM and PM4Py². Third, we construct the simulation model using the BIMP simulator. And finally, we analyze the results.

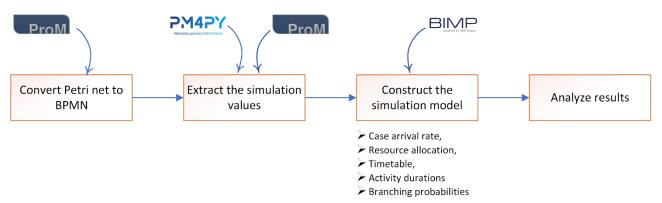


Figure 1. Process simulation main steps

Online BIMP Simulator: https://bimp.cs.ut.ee/simulator (For the documentation https://bimp.cs.ut.ee/simulator/qet-started)

² The PM4Py documentation is available on https://pm4py.fit.fraunhofer.de/docs





1. Discover the BPMN Model

The construction of the simulation model starts from a BPMN model, which can be drawn by hand³ or mined. In the remainder, we will mine the model. In the laboratory session on 'Hands-on Session on Log Filtering and Process Discovery in ProM', we have seen how to discover the Petri net using the ProM. It generates a Petri net, but it is possible to convert the Petri net model into a BPMN model.

Import the event log in ProM, filter start events and apply the ProM plug-in *Mine with inductive visual miner*. Export the model as Petri net with default values. The Petri net can be seen in the ProM workspace. Select the Petri net model on the workspace and find *Convert petri net to BPMN diagram* to convert it into a BPMN model (see Figure 2). The resulting BPMN is as in Figure 3. Go to the ProM workspace and click on *Export to disk* to export and save the BPMN file.

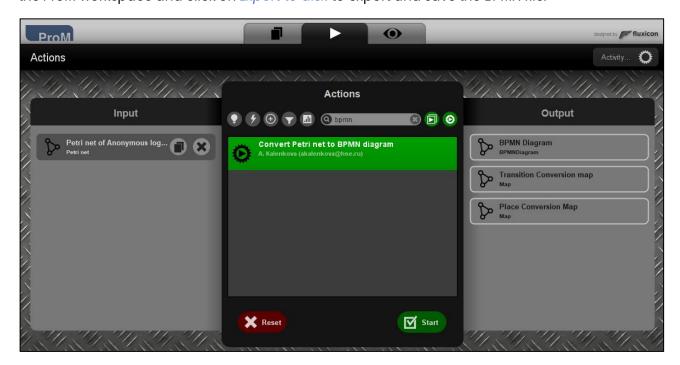


Figure 2. ProM plug-in Convert petri net to BPMN diagram

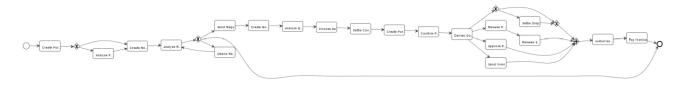


Figure 3. BPMN model converted from Petri Net

³ A simple and functional editor for BPMN model is https://bpmn.io/.



2. Discover the Simulation Parameters

The construction of the simulation model requires defining the different simulation parameters related to resources, time, and control perspectives. In this laboratory session, we will focus on discovering the following simulation aspects:

- > Case arrival rate (Scenario specification)
- > Resource allocation
- Timetable (Working days and hours)
- Activity durations (with distribution)
- Branching probabilities (Gateways)

Case Arrival Rate

We can compute the mean of the daily case arrival rate, i.e., the interval between two started process instances. using the code in Figure 4 (see more details in the laboratory on *Hands-on Session on Pm4py*).

```
#Daily case arrivals
import numpy as np

case_arrival = []
for i in range(1, len(log)):
    arrive1 = log[i-1][0]['startTimestamp']
    arrive2 = log[i][0]['StartTimestamp']
    if arrive1.date() == arrive2.date():
        case_arrival.append(arrive2-arrive1)
print(np.mean(case_arrival))
```

Figure 4. Case arrival rates

Resource allocation

In the next step, we retrieve, again using PM4py, the resource role, and the respective allocation through function <code>roles_discovery.apply()</code>. All the resulting roles are in Table 1 (see more details in the laboratory on <code>Hands-on Session on Pm4py</code>). In PM4Py, roles are represented through a set of sets of activities. Each set of activities includes those that are executed by similar resources. The role detection algorithm is more elaborated than seen in the class and allows the same resource to belong to two or more different roles. For example, the first role includes the executors of <code>Amend Purchase Requisition</code>. It consists of eight persons: <code>Penn Osterwalder</code>, <code>Nico Ojenbeer</code>, <code>Clement Duchot</code>, <code>Kim Passa</code>, <code>Miu Hanwan</code>, <code>Elvira Lores</code>, <code>Immanuel Karagianni</code>, and <code>Christian Francois</code>.

Table 1. Roles in the event log

Role	Activities	Quantity
ROLE 1	'Amend purchase requisition'	8
	'Amend Request for Quotation',	
ROLE 2	'Analyze Quotation Comparison Map',	17
	'Choose best option',	



Role	Activities	Quantity		
	'Create Purchase Requisition',			
	'Release Purchase Order'			
ROLE 3	'Analyze Purchase Requisition'	3		
	'Analyze Request for Quotation',			
	'Approve Purchase Order for payment',			
ROLE 4	'Create Purchase Order',	.3		
ROLE 4	'Create Quotation comparison Map',	J		
	'Send Request for Quotation to Supplier',			
	'Settle Conditions With Supplier'			
	"Authorize Supplier's Invoice payment",			
ROLE 5	'Pay Invoice',	2		
	"Release Supplier's Invoice"			
	'Confirm Purchase Order',			
ROLE 6	'Deliver Goods Services',	5		
	'Send Invoice'			
ROLE 7	'Create Request for Quotation'	17		
ROLE 8	'Settle dispute with supplier'	5		

Timetable

The next step is to find each role's timetables (the work schedule). Firstly, it is necessary to detect the workdays to discover the timetable. Follow the instruction in the *Hands-on Session on Pm4py* lab session. For example, Figure 5 and Figure 6 illustrate the working day and working hours for ROLE 1, respectively. The timetable for all resources is set 24/7.

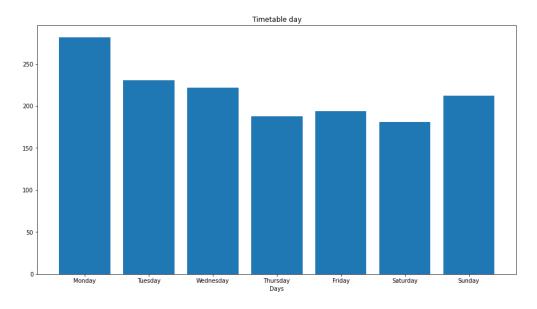


Figure 5. Histogram of working time for the ROLE 1



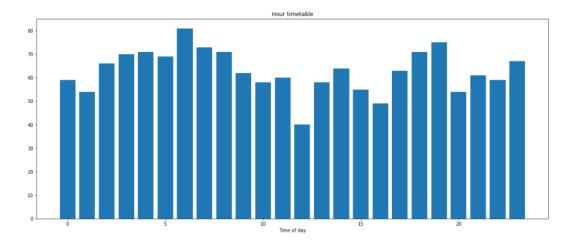


Figure 6. Histogram of working hours for the ROLE 1

Activity durations and distributions

Now we want to extract the duration of each activity of the model. PM4py allows us to compute the activity duration quickly. Table 2 shows the results of the function $soj_time_get.apply()$ in the PM4Py library by calculating the average time between the start and completion times for the activity's events in seconds.

Table 2. Activity durations

Activity	Duration (sec.)	Duration (min)
Amend Purchase Requisition	1641.82	27.36
Amend Request for Quotation	639.43	10.66
Analyze Purchase Requisition	394.87	6.58
Analyze Quotation Comparison Map	1209.30	20.15
Analyze Request for Quotation	1382.01	23.03
Approve Purchase Order for payment	60.00	1.00
Authorize Supplier's Invoice payment	0.00	0.00
Choose best option	0.00	0.00
Confirm Purchase Order	1188.52	19.81
Create Purchase Order	573.12	9.55
Create Purchase Requisition	1843.03	30.72
Create Quotation comparison Map	12155.01	202.58
Create Request for Quotation	321.84	5.36
Deliver Goods Services	91075.50	1517.92
Pay Invoice	561.79	9.36
Release Purchase Order	60.00	1.00
Release Supplier's Invoice	269.03	4.48
Send Invoice	0.00	0.00
Send Request for Quotation to Supplier	1415.74	23.60
Settle Conditions with Supplier	32389.10	539.82
Settle Dispute with Supplier	9221.32	153.69



We also need to determine, by the rule of thumb, the distribution of activities for simulation. We introduced the *PlotWaitingDistribution* function to draw a graph for looking at distributions in Figure 7.

```
# Somtimes, working on a dataframe is easier than xes files.
data_table = pm4py.convert_to_dataframe(log)
# Activity distribution
import matplotlib.pyplot as plt
import numpy as np
def PlotWaitingDistribution(log, activity):
   waitings = []
    for trace in log:
        for event in trace:
            if event['concept:name'] == activity:
                end = event['time:timestamp']
                start = event['StartTimestamp']
                time = end-start
                t = time.seconds
                t /= 60
                waitings.append(t)
   plt.figure(figsize=(10,5))
    plt.title(activity+' (Avg '+ str(np.mean(waitings).round(2)) + ', \\ Std ' + str(np.std(waitings).round(2)) + ')')
   plt.hist(waitings, bins = 50)
    plt.axvline(np.mean(waitings), c='red', label='avg')
   plt.legend()
    plt.xlabel('minutes')
    plt.ylabel('frequece')
   plt.show()
activities = list(sorted(set(data_table['concept:name'])))
for activity in activities:
   PlotWaitingDistribution(log, activity)
```

Figure 7. Activity Distribution Graphs

Figure 8 depicts two examples of uniform and normal distributions. We can say the activity of *Settle Conditions with Supplier* looks uniformly distributed with a minimum of 225 and a maximum of 875, whereas the activity of *Analyze Request for Quotation* has a normal distribution with a mean of 23.03 and a standard deviation of 8.52.

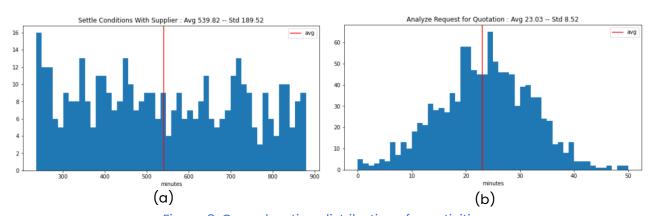


Figure 8. Case duration distributions for activities



Table 3 gives the accepted distributions of activities and the corresponding parameter values.

Table 3. Case duration distributions of activities

Activity	Distribution	Parameter values (min.)
Amend Purchase Requisition	Uniform	Min: 22, Max: 32
Amend Request for Quotation	Uniform	Min: 5, Max: 25
Analyze Purchase Requisition	Exponential	Mean: 6.58
Analyze Quotation Comparison Map	Uniform	Min:6, Max: 34
Analyze Request for Quotation	Normal	Mean: 23.03, Std. Dev.: 8.52
Approve Purchase Order for payment	Fixed	0.001
Authorize Supplier's Invoice payment	Fixed	0.001
Choose best option	Fixed	0.001
Confirm Purchase Order	Uniform	Min: 5, Max: 35
Create Purchase Order	Uniform	Min: 6, Max: 13
Create Purchase Requisition	Uniform	Min: 5, Max: 60
Create Quotation comparison Map	Uniform	Min: 40, Max: 375
Create Request for Quotation	Uniform	Min: 1, Max: 16
Deliver Goods Services	Uniform	Min: 0, Max: 1450
Pay Invoice	Uniform	Min: 4, Max: 15
Release Purchase Order	Fixed	1
Release Supplier's Invoice	Uniform	Min: 2, Max: 7
Send Invoice	Fixed	0.001
Send Request for Quotation to Supplier	Normal	Mean: 23.6, Std. Dev.:9.24
Settle Conditions with Supplier	Uniform	Min: 225, Max: 875
Settle Dispute with Supplier	Uniform	Min: 10, Max: 600

Branching probabilities

The branching probabilities at the gateways can be computed by the *Multi-perspective process* explorer introduced in the lab *Hands-on Session on Conformance Checking in Prom.* When this plug-in is used to discover the branching probabilities for the log, the result is as in Figure 9. In particular, we can identify three *XOR gateways*. The outgoing branching at XOR 1 is *Analyze purchase requisition* (performed 374 out of 608 times, 61.5%) and an invisible activity (performed 234 times, 38.5%). Compared with the BPMN model, the invisible transition corresponds to the branch of the first XOR that directly goes *Create Request for Quotation*. The remaining gateways can also be interpreted similarly.



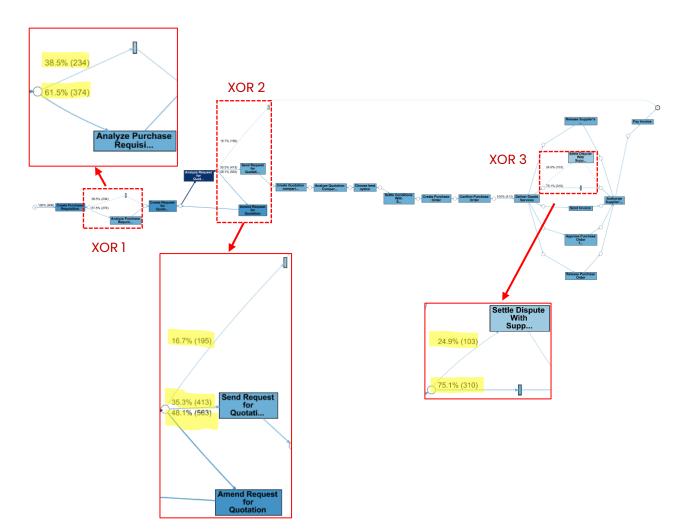
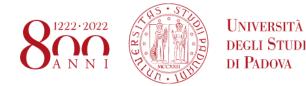


Figure 9. Gateways and probabilities

3. Creation of the Simulation Model

Once we have obtained the BPMN model and discovered the different simulation parameters, we can build the simulation model, which we will run using the BIMP simulator. BIMP Online Simulator can be accessed directly from a web browser at this link https://bimp.cs.ut.ee/. Then go to the Online Simulator window and click on the Academic version and upload the BPMN file that you have already created (see Figure 10).



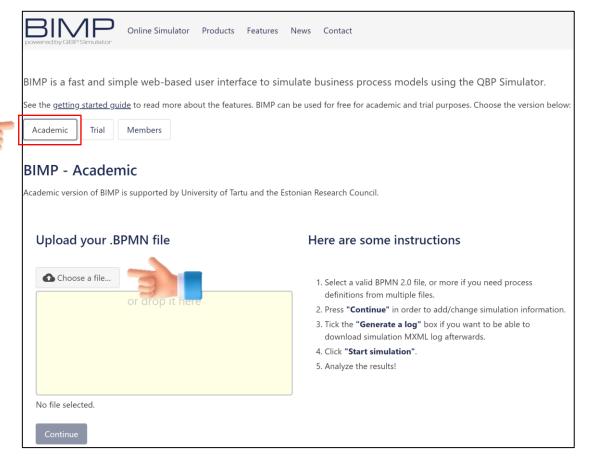


Figure 10. BIMP online simulator

You can visualize the BPMN model by clicking on *View BPMN Diagram*. Figure 11 shows how to define the number of instances to simulate, the inter-arrival time, the warm-up period, and the start timestamp of the scenario.

Since we have yet to determine what distribution best fits the inter-arrival time, we assume an exponential distribution, which is often seen as realistic, with an average of 257 minutes (see Figure 4 for '0 days 4 hours, 17 minutes and 34 seconds'), namely what we obtained from the event log analysis.

We also want to assume some warm-up period, which we assume is 10% of the traces here. We also aim to base the statistics on the same number of traces as in the original log, i.e., 608 traces. Considering that the warm-up and cool-down periods are equal to 5% each, we need to simulate 669 traces (i.e., 110% of the traces of the original event log).

After adding all these simulation parameters, save the simulation scenario as a BPMN file for later use so that you do not have to re-enter simulation parameters again the next time.



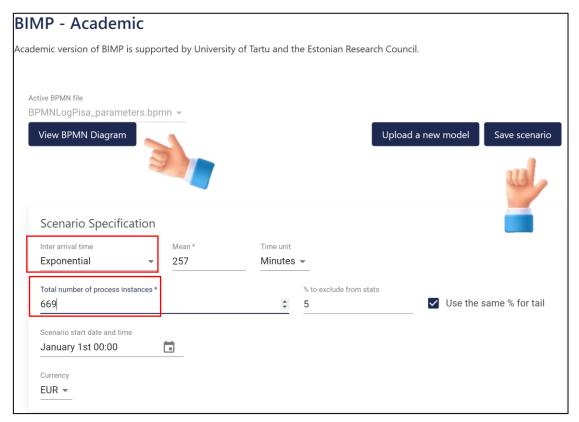


Figure 11. Scenario specifications

Figure 12 presents the roles discovered before and given in Table 1. We assume all resources in each role work 24/7. We ignore the cost input. The timetable created for this exercise is given in Figure 13.

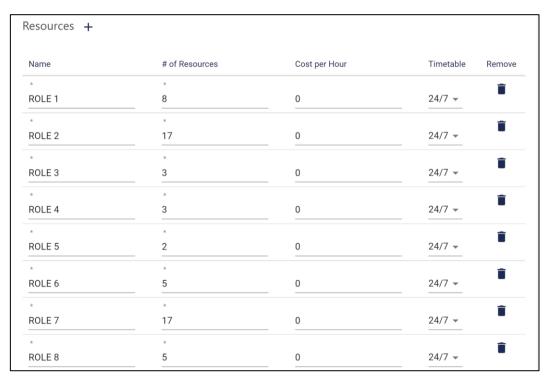


Figure 12. Defining roles



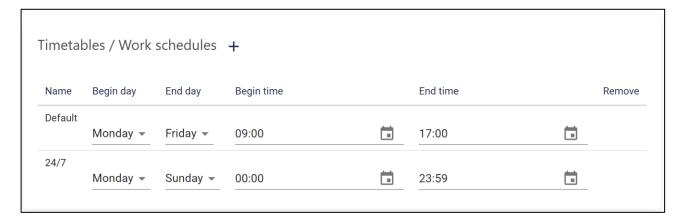


Figure 13. Defining timetables

Tasks part of the simulation requires to define the role, distribution and parameters. For instance, Figure 14 shows the configuration for the activity *Amend Request for Quotation*. In Table 2, we have shown that the activity of *Amend request for Quotation* is performed by ROLE 2 and in Table 3, it is distributed uniformly with a minimum 5 and a maximum 25 seconds.



Figure 14. Defining tasks considering roles and distributions

Because Amend Request for Quotation is infrequent activity (performed only 11 times, 0.12% in all cases), the model discovered by Mine with inductive visual miner ignored it while creating petri net. Therefore, it is absent in the BPMN.

Then, we add the branching probabilities for each XOR split in the model. Figure 15 shows how to set the branching for the BPMN's XOR split between *Create Request for Quotation* and *Analyze Purchase Requisition*. It corresponds to the XOR in Figure 9, which also shows the branching probabilities.



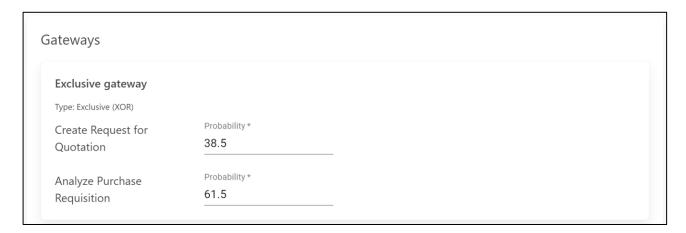


Figure 15. Defining gateway probabilities

4. Simulation

Once the scenario is set up, simulations can be carried out. The simulation results are summarized through specific statistics, but it is also possible to observe the single executions by generating the event-log file of the simulation. To do so, you have to click on *Generate an MXML log* to export the simulated event log in the MXML format (see Figure 16). MXML is an older standard to store event logs, which has later been replaced by XES, but it is still supported by such tools as ProM and DISCO for further analyses.



Figure 16. Starting a simulation by generating an MXML log in BIMP

Figure 17 shows the options after you complete the simulation. BPMN Diagram with results heat map enables us to visualize the process model with respect to the waiting time, counts, costs, and durations. Save results saves the simulation results so that you can reach them later. Save scenario saves the simulation scenario (parameters) so that you can use them later without entering all inputs in each running. Download MXML and Download CSV buttons are used to download the event log generated by the simulation in the format of MXML and CSV, respectively. After you obtain the simulation results, you can click on Back to edit data if you want to edit some parameters.



Figure 17. Options after simulation





After running your simulation scenario, the BIMP simulator will provide you a dashboard with diagrams in Figure 18 and data table in Figure 19. The dashboard gives you information about the simulated process. Because we use the 7/24 timetable, process cycle times, including and excluding off-timetable hours, are the same in Figure 18. If there were not enough resources to perform the activities, then tasks would be queued, and this would cause high waiting and cycle times in process instances. In our example, total waiting time of the process is mainly less than 60 minutes. The process cost is not applicable because we have not entered any input for the cost. Resource utilization indicates how resources are utilized during the process. We expect to see a higher ratio for each resource. ROLE 4 has the most considerable resource utilization rate, but it is still relatively low: the ROLE 6 utilization is lower than 20%.



Figure 18. Simulation charts and statistics





Figure 19 zooms into the dashboard shown in Figure 18 for each role in the simulation scenario. The waiting time column is related to the process waiting times in Figure 18. You can see the minimum, maximum and average waiting durations for the corresponding activity. The Duration column is related to process cycle times including and excluding off-timetable hours in Figure 18. Durations are produced by the distribution of the tasks on BIMP. The activity of *Deliver Goods Services* is the most time-consuming activity with a 11.7-hour duration. *Settle Conditions with Supplier* is the second high-duration activity. The difference between the maximum and minimum durations is higher at *Approve Purchase Order for payment*.

Activity Durations, Costs, Waiting times, Deviations from Thresholds

Name		Wait	ing time		Duratio	n		Durati	on over t	hreshold	Cost			Cost	over thr	eshold
	Count	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Amend Request for Quotation	579	0 s	0 s	0 s	5.1 m	15.1 m	24.9 m	0 s	0 s	0 s	0	0	0	0	0	0
Analyze Purchase Requisition	366	0 s	0 s	0 s	2.1 s	6.8 m	44.2 m	0 s	0 s	0 s	0	0	0	0	0	0
Analyze Quotation Comparison Map	412	0 s	0 s	0 s	6 m	19.6 m	34 m	0 s	0 s	0 s	0	0	0	0	0	0
Analyze Request for Quotation	1181	0 s	1.7 m	4.8 h	0.1 s	25 m	5.2 h	0 s	0 s	0 s	0	0	0	0	0	0
Approve Purchase Order for payment	412	0 s	7.9 m	8.5 h	0 s	7.9 m	8.5 h	0 s	0 s	0 s	0	0	0	0	0	0
Authorize Supplier's Invoice payment	412	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0	0	0	0	0	0
Choose best option	412	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0	0	0	0	0	0
Confirm Purchase Order	412	0 s	7.1 s	48.8 m	5 m	19.8 m	1.2 h	0 s	0 s	0 s	0	0	0	0	0	0
Create Purchase Order	412	0 s	5.8 m	5.7 h	6 m	15.2 m	5.9 h	0 s	0 s	0 s	0	0	0	0	0	0
Create Purchase Requisition	602	0 s	0 s	0 s	5 m	32.1 m	59.9 m	0 s	0 s	0 s	0	0	0	0	0	0
Create Quotation comparison Map	412	0 s	3.5 m	5.5 h	40.6 m	3.6 h	11 h	0 s	0 s	0 s	0	0	0	0	0	0
Create Request for Quotation	602	0 s	0 s	0 s	1 m	8.5 m	16 m	0 s	0 s	0 s	0	0	0	0	0	0
Deliver Goods Services	412	0 s	0 s	0 s	6.1 m	11.7 h	1 d	0 s	0 s	0 s	0	0	0	0	0	0
Pay Invoice	412	0 s	0 s	0 s	4 m	9.5 m	14.9 m	0 s	0 s	0 s	0	0	0	0	0	0
Release Purchase Order	412	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0 s	0	0	0	0	0	0
Release Supplier's Invoice	412	0 s	0 s	0 s	2 m	4.4 m	7 m	0 s	0 s	0 s	0	0	0	0	0	0
Send Invoice	412	0 s	3.3 s	22.8 m	0 s	3.3 s	22.8 m	0 s	0 s	0 s	0	0	0	0	0	0
Send Request for Quotation to Supplier	412	0 s	3.1 m	4.5 h	4.1 m	27.1 m	4.9 h	0 s	0 s	0 s	0	0	0	0	0	0
Settle Conditions With Supplier	412	0 s	7.2 m	5.2 h	3.8 h	9.4 h	17.6 h	0 s	0 s	0 s	0	0	0	0	0	0
Settle Dispute With Supplier	82	0 s	0 s	0 s	26.7 m	5.6 h	9.8 h	0 s	0 s	0 s	0	0	0	0	0	0

Figure 19. Activity statistics for the simulation scenario

Press the *Download MXML* button to download the event log created by BIMP Simulation if you want to discover the process model by a process mining tool.



5. Final Exercises

1. Analysis of the event log produced by the simulation

Download the event log generated by the simulation at the end of Section 4 in the MXML format. Load the event log in DISCO and compare the activity durations and waiting times.

2. What-is scenarios

We assume the company wants to expand and manage more orders (case arrival rate parametrized as an exponential distribution with a mean of 16 minutes). In which roles does the management need to add resources? and how many resources does he have to add to manage the demand?

- ❖ First, create a new simulation scenario using the new mean of the case arrival rate and simulate it through the BIMP simulator. Look at the simulation results, including animating the event log in DISCO. Which role has got the higher resource utilization?
- ❖ Looking at the previous results, we want to add the right amount of resources for each previously detected role. How many resources should the manager add to have the average process instance cycle times under one day? Calculate the average value over ten simulation runs to have more reliable results.

Solution of Exercise 1

Figure 20 depicts a part of the process models belonging to the actual log (purchaseExample.xes) and simulated log. Activity durations are almost similar. The slight differences result from distributions used in the simulation. Table 4 gives all durations of activities. However, the waiting times between activities are considerably different. It is because we defined the resources work 7/24. Disco defines the timetable as 08.00 – 18.00 during working days. If you change it to 7/24 using TimeWarp (right to the Filter button), you can see the waiting times are also similar.

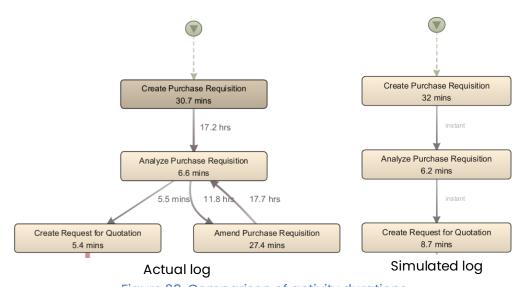


Figure 20. Comparison of activity durations



Table 4. Average activity durations of actual and simulated log

Activity	Actual	Simulation	
Amend Purchase Requisition	27 mins, 21 secs	14 mins, 39 secs	
Amend Request for Quotation	10 mins, 39 secs	(not in the model)	
Analyze Purchase Requisition	6 mins, 34 secs	6 mins, 11 secs	
Analyze Quotation Comparison Map	20 mins, 9 secs	19 mins, 12 secs	
Analyze Request for Quotation	23 mins, 2 secs	23 mins, 14 secs	
Approve Purchase Order for payment	1 min	10 millis	
Authorize Supplier's Invoice payment	0 millis	10 millis	
Choose best option	0 millis	10 millis	
Confirm Purchase Order	19 mins, 48 secs	20 mins, 35 secs	
Create Purchase Order	9 mins, 33 secs	9 mins, 32 secs	
Create Purchase Requisition	30 mins, 43 secs	32 mins	
Create Quotation comparison Map	3 hours, 22 mins	3 hours, 17 mins	
Create Request for Quotation	5 mins, 21 secs	8 mins, 41 secs	
Deliver Goods Services	1 day, 1 hour	12 hours, 3 mins	
Pay Invoice	9 mins, 21 secs	9 mins, 21 secs	
Release Purchase Order	1 min	10 millis	
Release Supplier's Invoice	4 mins, 29 secs	4 mins, 28 secs	
Send Invoice	0 millis	10 millis	
Send Request for Quotation to Supplier	23 mins, 35 secs	24 mins, 32 secs	
Settle Conditions with Supplier	8 hours, 59 mins	8 hours, 57 mins	
Settle Dispute with Supplier	2 hours, 33 mins	5 hours, 11 mins	



Solution of Exercise 2

You can use the previous simulation file to change the case arrival rate to 16 minutes as given in Figure 21.

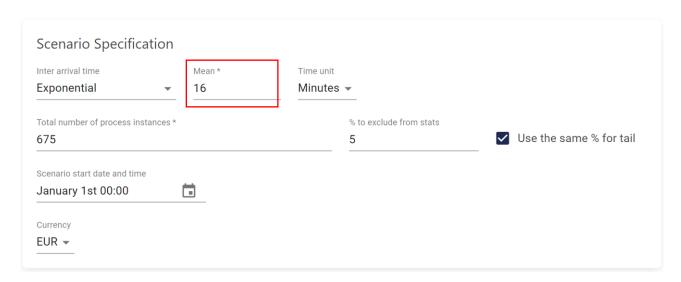


Figure 21. Changing mean of the case arrival rate

The simulation results in Figure 22 indicate that Role 4 covers the most utilized resources. It was less than 20% in Figure 18. A 16-fold decrease (257/16=16) in average case arrival rate caused resource utilization in Role 4 to increase five-fold.

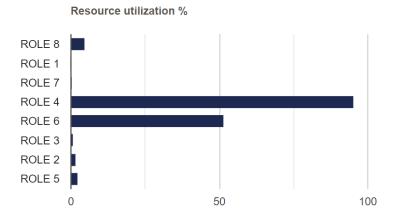


Figure 22. Resource utilization

Figure 23 obtained from Disco also supports this result. When we look at the animation on Disco, we see some activities are performed more frequently than others. These activities are *Analyze Request for Quotation, Send Request for Quotation to Supplier, Settle Conditions with Supplier, Create Purchase Order,* and *Approve Purchase Order for payment*. Table 1 presents all these activities are executed by Role 4.



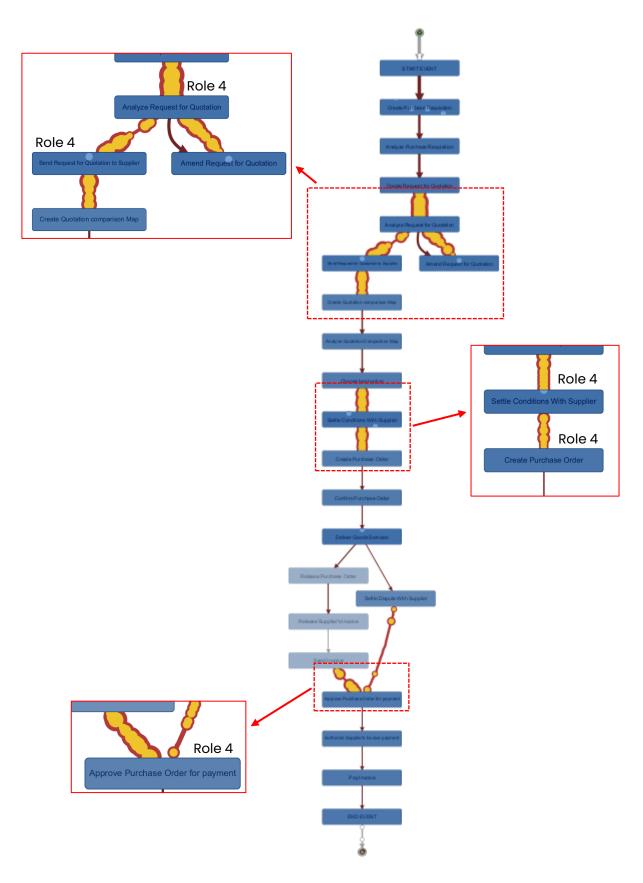


Figure 23. What-If analysis by changing the case arrival rate



Under these circumstances, average process instance cycle times is 6.1 weeks, as given in Figure 24.

Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	5.5 hours	12.2 weeks	6.1 weeks
Process instance cycle times excluding off-timetable hours	5.5 hours	12.2 weeks	6.1 weeks
Process instance costs	0 EUR	0 EUR	0 EUR

Figure 24. Process instance cycle times after changing case arrival rate

Figure 22 shows that although Role 4 covers the most utilized resources, Role 6 is the second role utilizing the resources most. It indicates if we increase the number of resources in Role 4, the resources in Role 6 will be utilized most after a certain point. If you try cautiously increasing the number of resources both in Role 4 and Role 6, you can reach fewer cycle times. The manager should add 12 more resources (15 in total) for Role 4 and 9 more resources (14 in total) for Role 6 to have the average process instance cycle times under one day. Table 5 gives the average process instance cycle times for 11 simulation runs.

Table 5. Simulation runs

Run	Cycle time (hour)
1	23.2
2	24
3	24
4	26.4 (1.1 d)
5	24
6	24
7	21
8	22.3
9	22.7
10	23.7
11	23.2
Average	23.21