Retrieving Resource Availability Insights from Event Logs

Niels Martin, Frank Bax, Benoît Depaire, An Caris
Hasselt University
Agoralaan Building D
3590 Diepenbeek, Belgium
Email: niels.martin@uhasselt.be

***Abstract*—Resources are a critical component of a business process as they execute the activities. These resources, especially human resources, are not permanently available and tend to be involved in multiple processes. However, a company might wish to analyze or model a single process. To this end, insights need to be gathered on the availability of a resource for a particular process. This paper presents a procedure to retrieve daily availability records from an event log, which express a resource's availability for the process under analysis while taking into account (i) the temporal dimension of availability and (ii) intermediate availability interruptions. Both the daily availability records themselves and the resource availability metrics that are introduced allow managers and employees to gain understanding in resource allocation to a process. The outlined procedure and metrics are applied to a real-life call center log, showing the need to post-process the daily availability records. Post-processing increases their comprehensiveness and is required to obtain meaningful values for particular metrics.**

I. INTRODUCTION

Business processes are composed of a series of connected activities which are executed by resources such as staff members or machines. Resources, especially human resources, are not permanently available [1] and tend to be involved in multiple processes [2]. However, a company might wish to analyze or model a single business process such as the inbound logistics process or the billing process. Consequently, the availability of resources for the process under consideration needs to be determined, which is the topic of this paper. Note that, in this work, resource availability refers to time periods during which the resource is allocated to the process under analysis. In [2], van der Aalst et al. claim that companies only have limited insights in the availability of their human resources, underlining the importance of efforts on this matter.

One might argue that resource schedules can be provided by e.g. the human resource department. However, even when they are present, it might not be possible to deduce the availability for a specific process from these generic schedules, especially when the work organization is less rigid. Consequently, alternative information sources are required. Interviews with staff members and their managers can be considered, but are both time-consuming and do not guarantee accurate results as employees might not be aware of the time dedicated to a specific process. Hence, investigating the use of more readily available information sources is valuable. To this end, process-aware information systems such as ERP or CRM systems make available event logs, which are data files containing process execution information. Extracting knowledge from event logs belongs to the process mining field [3]. In particular, the resource perspective of this field is targeted, which is in line with the research recommendations in [4].

This paper focuses on retrieving resource availability insights from an event log. To this end, all information embedded in the event log is systematically used to obtain daily availability records, which express the resource's availability for the process under consideration. The mined daily availability records are innovative as they are the first to take into account (i) the temporal dimension of availability, i.e. at which time of day a resource is available, and (ii) intermediate availability interruptions caused by, e.g., a break or the allocation of a resource to another process. Given these characteristics, they can be used to support resource schedule modeling, which is marked as a research gap in [5]. Besides the insights conveyed by the daily availability records themselves, another contribution of this paper is that an extensible set of metrics is presented which allow managers and employees to gather an aggregated view on a resource's allocation to the process.

The remainder of this paper is structured as follows. The importance of taking into account the temporal dimension of availability and intermediate interruptions, which is the case in the daily availability records, is illustrated in Section II. Section III presents a procedure to retrieve daily availability records from an event log and proposes metrics to gain insights on a more aggregated level. The introduced procedure and developed metrics are applied to a real-life call center event log in Section IV. Sections V and VI respectively discuss related work and present a conclusion.

II. PROBLEM STATEMENT

This section illustrates the importance of taking into account the temporal dimension of availability and intermediate availability interruptions, which are the distinguishing characteristics of the daily availability records. To this end, the influence of simplified resource availability definitions on simulated process performance measures is determined. The process model in Figure 1 is used, which is annotated with all assumed parameters. New cases arrive on a daily basis between 08:00:00 and 17:00:00, following an exponential interarrival time distribution with a mean of eight minutes.

978-1-4673-9885-5/16/$31.00 ©2016 IEEE

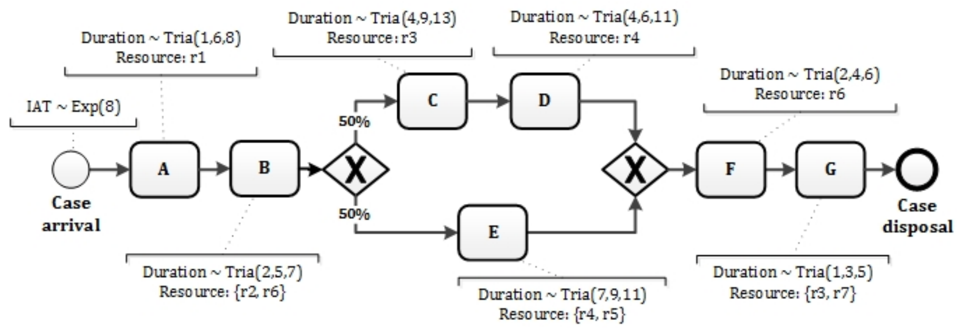

Fig. 1. Running example

Activity durations are expressed in minutes by a triangular distribution. Furthermore, suppose that the seven resources are available according to the following schedules in reality:

- *r1*, *r4*, *r5*, *r6*: 08:00:00 - 12:00:00, 13:00:00 - 17:00:00
- *r2*: 09:00:00 - 12:00:00, 13:00:00 - 18:00:00
- *r3*: 07:30:00 - 12:00:00, 13:30:00 - 17:00:00
- *r7*: 13:30:00 - 17:30:00

Consequently, all resources, besides *r7*, are available eight hours a day for the process. However, the distribution of this working time over the day is not the same for all resources. Resource *r7* is only assigned to the process four hours a day, i.e. *r7* is only available in the afternoon.

The above availability specification takes into account the temporal dimension of availability and intermediate interruptions and, hence, reflects the characteristics of the daily availability records. However, resource availability can also be modeled in a simplified manner by, e.g., merely focusing on the number of hours that a resource is available. To illustrate the consequences on process performance metrics of including such simplified specifications or ignoring this matter altogether, four resource availability definitions are considered:

- Benchmark scenario: Resources are available according to the schedules outlined above.
- Scenario 1: Resource availability is not taken into account, i.e. resources are assumed to be permanently available.
- Scenario 2: Resource availability is known in terms of the number of hours, but it is unknown how these are spread over the day. To this end, it is assumed that resources become available when cases start arriving at the process. Intermediate availability interruptions are not taken into account.
- Scenario 3: Resource availability is known in terms of the number of hours, but it is unknown how these are spread over the day. However, the start of the resource's availability for the process is assumed to be known. Intermediate availability interruptions are not taken into account.

The process in Figure 1 is simulated for each of these scenarios and the mean and standard deviation (sd) of the flow time and waiting time are calculated. The use of common random numbers ensures mutual comparability between reality and the alternative scenarios as model components such as case arrival time, activity durations and case routing at the decision point are the same in each specification [6]. Consequently, the observed differences in performance measures can be attributed to differences in the resource availability definition.

TABLE I
EFFECT OF VARIOUS RESOURCE AVAILABILITY SPECIFICATIONS

	Flow time		**Waiting time**	
	mean	**sd**	**mean**	**sd**
Benchm.	241.70	323.65	213.00	323.46
Scen. 1	47.77 (-80%)	18.24	19.07 (-91%)	17.40
Scen. 2	495.70 (+105%)	507.89	467.00 (+192%)	507.59
Scen. 3	332.20 (+37%)	416.74	303.50 (+42%)	416.81

After simulating 500 cases under each specification, Table I summarizes the aforementioned process performance measures. The numbers between brackets reflect percentage deviations from the benchmark values.

Table I shows large differences in reported process performance depending on the specification. While it is intuitively clear that ignoring resource availability, as is the case in Scenario 1, is an undue simplification of reality, Scenarios 2 and 3 also lead to important deviations compared to a model that correctly includes the resource schedules observed in reality. This stresses the importance of taking into account the temporal dimension of availability and integrate intermediate interruptions when modeling resource availability. This paper takes these observations into account when mining daily availability records from an event log.

III. Daily availability record retrieval and analysis

This section describes a procedure to retrieve daily availability records from an event log and introduces resource availability metrics based on these records. The event log requirements and necessary log transformation are outlined in in Sections III-A and III-B, respectively. Using this transformed event log, daily availability records are constructed following the procedure described in Section III-C. As this procedure might lead to fragmented daily availability records, post-processing can be required as will be discussed in Sec-

TABLE II
ILLUSTRATION OF EVENT LOG STRUCTURE

Case id	Timestamp	Activity	Event type	Resource
...	...	...	...	...
72	06/04/2016 09:02:01	C	start	r3
56	06/04/2016 09:05:51	D	start	r4
72	06/04/2016 09:08:42	C	complete	r3
38	06/04/2016 09:08:42	G	start	r3
56	06/04/2016 09:11:17	D	complete	r4
38	06/04/2016 09:11:52	G	complete	r3
...	...	...	...	...

TABLE III
ILLUSTRATION OF RESOURCE LOG STRUCTURE FOR RESOURCE R3

Case id	Timestamp	Activity	Event type	Resource
...	...	...	...	...
72	06/04/2016 09:02:01	C	start	r3
72	06/04/2016 09:08:42	C	complete	r3
38	06/04/2016 09:08:42	G	start	r3
38	06/04/2016 09:11:52	G	complete	r3
...	...	...	...	...

tion III-D. Based on the daily availability records, Section III-E proposes some resource availability metrics.

A. Event log requirements

An event log consists of ordered events related to a particular case and activity. In this paper, the timestamp, resource and event type needs to be registered for each event. Two event types have to be recorded: start and complete, which are both transitions of the XES lifecycle extension[1]. Each start event should have an accompanying complete event with the same resource being associated to both events. The requirement of matching start and complete events is in accordance with the notion of consistent traces, as defined in [7].

When considering the process visualized in Figure 1 as an illustration, Table II exemplifies the event log structure.

B. Event log transformation

As daily availability records are defined at a resource level, the event log is divided in resource logs containing all events associated to a particular resource. An excerpt of the *r3* resource log from the running example is shown in Table III.

Initially, resource logs are composed of atomic events. To retrieve daily availability records, the activity instance concept, i.e. the execution of a particular activity on a particular case, is introduced. To this end, each start event is mapped to its corresponding complete event, where the latter is the complete event associated to the same case and activity in the resource log. When multiple start and complete events for a particular case-activity combination are present, the first occurring unmapped start event will iteratively be mapped to the first occurring unmapped complete event. The activity instance mapping for the resource log excerpt in Table III is shown in Table IV.

[1]http://www.xes-standard.org/_media/xes/xesstandarddefinition-2.0.pdf

TABLE IV
ILLUSTRATION OF ACTIVITY INSTANCE MAPPING IN RESOURCE LOG R3

Case id	Activity	τ_{start}	τ_{complete}
...	...	...	...
72	C	06/04/2016 09:02:01	06/04/2016 09:08:42
38	G	06/04/2016 09:08:42	06/04/2016 09:11:52
...	...	...	...

TABLE V
EXCERPT FROM ACTIVITY INSTANCE MAPPING IN RESOURCE LOG R3

Case id	Activity	τ_{start}	τ_{complete}
...	...	...	...
104	C	07/04/2016 11:32:07	07/04/2016 11:43:27
82	G	07/04/2016 11:43:27	07/04/2016 11:46:04
86	G	07/04/2016 12:31:09	07/04/2016 12:33:02
107	C	07/04/2016 12:33:02	07/04/2016 12:42:46
108	C	07/04/2016 13:08:17	07/04/2016 13:19:42
111	C	07/04/2016 13:19:42	07/04/2016 13:25:06
112	C	07/04/2016 13:48:22	07/04/2016 13:59:22
...	...	...	...

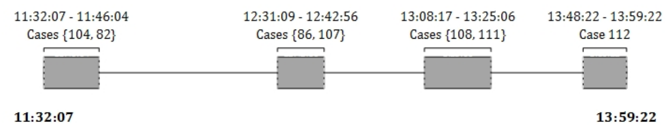

Fig. 2. Merge activity periods - illustration

C. Event log analysis

The resource logs with activity instance mappings are the input for the remainder of the analysis. The goal is to construct daily availability records by maximally exploiting explicit and implicit information in this log to mark time periods during which a resource is either available or unavailable for the process under consideration. The event log analysis is performed in several steps, which are outlined in de remainder of this subsection. To illustrate these steps, consider the activity instance mappings for *r3* in Table V.

1) Merge active periods: Activity execution marks an active period, indicating that the resource was certainly available for the process during that time frame. When multiple activity executions by a particular resource (quasi-)immediately follow each other, they are merged to reflect a period of availability. In Table V, this holds for the instances associated to cases {104,82}, {86,107} and {108,111}. For example: the instances for cases 104 and 82 are merged in a single available period from 11:32:07 until 11:46:06. This is visualized in Figure 2, where gray bars represent periods of availability.

It is indicated that activity instances that quasi-immediately follow each other can also be merged. This is consistent with the intuition that, e.g., some set-up time might be required to start handling a new case when the current one is processed. Consequently, a time tolerance between consecutive activity instances can be specified to accommodate set-up time and avoid undue fragmentation of available periods. Domain knowledge will be valuable to specify a tolerance as an appropriate value will be context-dependent. However, a log-based recommendation can also be useful. As the set-up time to start a new case is likely to depend upon the duration of the activity that needs to be executed, a percentage of the median duration of this particular activity can be used. In order to keep the tolerance value small, a default value of 5% is proposed. Regardless of how the time tolerance is specified, activity executions of a resource will be merged in a single available period when the time difference between their completion and the start of the following one is lower than the tolerance value.

2) Identify boundary pending cases: While the active periods, used to mark resource availability, are explicitly recorded in the event log, implicit event log insights can also identify periods in which the resource is unavailable for the process. A resource is stated to be unavailable during a time frame when cases are present on which the resource could perform a particular activity, but no resource activity is recorded. The latter builds upon the assumption that a resource will perform work when work is present and he/she is assigned to the process. Moreover, it is consistent with [8], stating that a resource is not available in a time period when the number of activities awaiting execution in this period is larger than zero. This intuition is operationalized through the notion of pending cases. A pending case for a resource is a case that requested service from that resource, but for which the request is not yet fulfilled at a particular point in time.

The identification of pending cases requires knowledge on the time at which a case requests a resource. When a Q-log [9] is present, which is an event log containing events expressing a case's entrance in a queue, obtaining insights on this matter is trivial. However, hypothesizing the existence of a Q-log is a rather strong assumption. Consequently, a proxy for the resource request time based on an event log as shown in Table II is warranted. When a case is assumed to request a resource at the current activity when the case's previous activity is completed, this comes down to identifying this prior activity. The latter can require a process notion, which can be retrieved using domain knowledge or through the application of a control-flow discovery algorithm on the event log, of which an overview is presented in e.g. [3] and [10]. Given the large research body on control-flow discovery, the operationalization of the prior activity identification is not treated in this work. The complete timestamp of this previous activity for the case under consideration is used as a proxy for the resource request time at the current activity.

When the resource request time is determined for each activity instance, pending cases can be detected. The presence of pending cases is verified first at the end of each merged active period such as 11:46:06, 12:42:56 and 13:25:06 in Figure 2. When so called boundary pending cases, i.e. pending cases at the end of a merged active period, are present, the resource is unavailable until the start of the following merged active period. If, for instance, case 86 issued a resource request for *r3* at 11:44:04, case 86 was waiting for service from *r3* when this resource' first active period ended, but no resource's activity is recorded until 12:31:09. Consequently, *r3* is stated to be unavailable between the first two active periods in Figure 2. This is visualized in Figure 3, where a period of

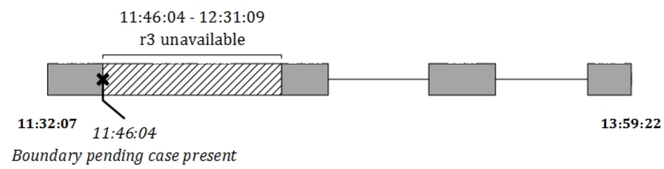

Fig. 3. Boundary pending cases - illustration

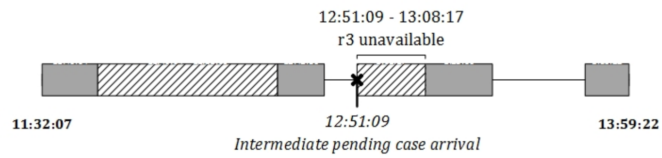

Fig. 4. Intermediate pending cases - illustration of request fulfilled by analyzed resource

unavailability is marked using a dashed box.

The example of the request of case 86 focuses on a request that is eventually fulfilled by the resource under analysis, resource *r3*. However, when multiple resources can execute a particular activity and a single queue is present, e.g. one waiting room with multiple counters, more insights can be retrieved from the event log. Under these conditions, pending cases which are eventually serviced by another resource also qualify as boundary pending cases for the resource under analysis. Consequently, the resource under analysis is unavailable until another resource starts servicing the last of these boundary pending cases when this precedes the start of the following merged active period. An illustration on this matter will be provided in the discussion on intermediate pending cases. Note that a situation where multiple resources can execute an activity, but each resource has a dedicated queue, is conceptually similar to a scenario where only one resource performs an activity.

3) Identify intermediate pending cases: When no boundary pending cases are present, the analysis proceeds by considering the first request for the resource under analysis that is issued after the end of a particular merged active period. If this request is not fulfilled immediately while no other resource activity is recorded, the case under consideration is an intermediate pending case. When such a case is detected, the resource is unavailable from the time the resource request is issued until the start of the next merged active period. Suppose, for instance, that case 108 requested *r3* at 12:51:09, while the next active period of *r3* only starts at 13:08:17. From this, it follows that *r3* is unavailable from 12:51:09 until 13:08:17, as visualized in Figure 4.

As with boundary pending cases, intermediate pending cases can also originate from resource requests that are eventually fulfilled by other resources in case of a single queue for a particular activity. Consider for instance activity G, which can be carried out by either *r3* or *r7*. Suppose that case 93 issues a request for activity G at 13:28:08 and *r7* fulfills this request at 13:42:51. As no activity is recorded for *r3* within this period,

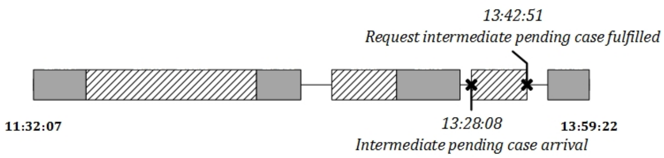

Fig. 5. Intermediate pending cases - illustration of request fulfilled by other resource

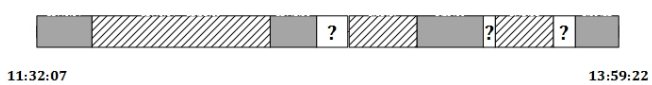

Fig. 6. Daily availability record - illustration

while an intermediate pending case is present, *r3* is unavailable within this time window. This is depicted in Figure 5.

4) Outcome: The previous steps analyzed the event log to identify periods during which the resource is available or unavailable for the process. The remaining time periods are unassigned periods, which are time frames during which no resource activity is recorded an no work is present for the resource under consideration. Consequently, the event log does not provide any evidence to assign these periods as either available or unavailable. Figure 6 visualizes the obtained daily availability record, where unassigned periods are represented by a question mark.

D. Daily availability record post-processing

Depending on the characteristics of the process such as the typical workload and the typical processing time, the daily availability records obtained following the procedure in Section III-C can be very fragmented. The latter will also become apparent for the real-life event log in Section IV. When this is the case, post-processing efforts can increase the comprehensibility of the daily availability records and improve the expressiveness of certain resource availability metrics. Several post-processing steps can be taken into consideration, as briefly outlined below.

1) Working day specification: A first post-processing step involves the specification of the start and the end of a resource's working day. This enables the simplification of the daily availability records as a resource is unavailable for the process outside the working day. Moreover, a resource is assumed to be unavailable for the process on days where no resource activity has been recorded.

To operationalize the latter, a definition of the start and end of a working day is required. These can be proxied by the start of the first activity instance of a resource on a particular day and the end of the last instance of this resource on that day, respectively, which is consistent with [11]. It should be noted that, in this paper, a working day is approached from the perspective of the process under analysis. Consequently, its start and end does not correspond by definition to their equivalents from the organizational perspective when resources are involved in multiple processes.

2) Unassigned period imputation: The daily availability records created using the procedure in Section III-C can contain unassigned periods. During post-processing, these periods can be assigned as an available period, an unavailable period or a mixture of both. This imputation reduces the fragmentation of the daily availability records and seems warranted, especially when the unassigned periods are relatively short.

An elementary method to designate unassigned period is a single imputation approach contingent upon the type of periods immediately before and after an unassigned period. When both the prior and the next period are of the same type, i.e. the resource is available or unavailable during both time frames, the unassigned period is also assumed to be of that type. In contrast, when the resource is available in the prior period and unavailable in the next period, or vice versa, the unassigned period is split up. The position of the split is proportionate to the relative length of the adjacent periods. Suppose that an unassigned period is flanked by an available period of one hour and an unavailable period of half an hour. In this case, two thirds of the unassigned timespan is added to the available period and the remainder to the unavailable period. This approach will maintain the proportionality between the adjacent available and unavailable period.

3) Unavailability period threshold: A final post-processing step involves the specification of an unavailability period threshold. Only unavailability periods that are longer than this boundary value will be maintained. Setting such a threshold is warranted as unavailability is defined rather strictly in Section III-C. A resource is deemed to be unavailable when he can potentially process a case, but refrains from doing so. This assumes that a resource is eager to immediately process an arriving case whenever this is possible and does not take into account that, e.g., particular cases might be deferred to a more experienced colleague. This is rather rigid as unavailability conceptually reflects a time period during which a resource is not available for the process under consideration due to, e.g., a break or the resource's assignment to another process. As a consequence, short time frames that are marked as unavailable following the specification in Section III-C might not represent genuine unavailability period. Introducing a threshold can both reduce the fragmentation of the daily availability records and better align the remaining unavailable periods with their conceptual meaning. An appropriate boundary value will depend upon the process under consideration.

E. Resource availability metrics

The daily availability records, potentially after post-processing, allow managers to gain insight in resource allocation to a particular process, the variation of this availability over several days and weeks, etc. However, depending on the time horizon of the analysis period, a large number of daily availability records might be obtained. To describe resource availability on a more aggregated level, an extensible set of resource availability metrics is introduced. These can convey additional insights to employees and managers or can focus their attention when analyzing the daily availability records.

Table VI presents an overview of the resource availability metrics. All metrics are defined on a working day level which is, as mentioned in Section III-D1, approached from the perspective of the process under analysis. The vast majority of the metrics are centered around available, unavailable and idle periods. For each of them, a distinction is made between their total length, the number of periods of this type and the length of individual periods. The latter two are especially useful as they provide insight in the fragmentation of the total length of, for instance, the available periods over a working day.

Besides a description, a recommended calculation base is indicated for each of the metrics. The daily availability records before post-processing and after each post-processing step form potential calculation bases. The only exception is the resource active time, reflecting the time during which a resource is actually performing activities for the process. For this metric, the event log should be used as available periods in the daily availability records before post-processing also include, e.g., set-up time when a tolerance is specified while merging active periods. This time tolerance is not resource active time under its strict definition. Conversely, when for instance set-up time is considered to be part of resource active time, the daily availability records before post-processing can also be used as a calculation base. A detailed discussion on why a particular calculation base is recommended is beyond the scope of this paper. However, the selected calculation base will influence the interpretation of the metrics. Consider that, e.g., daily availability records after the second post-processing step are used as a calculation base for the available period metrics. In that case, these metrics present an approximation of real-life availability as unassigned periods have been imputed.

IV. Application to Call Center Log

The procedure outlined in Section III is applied to a real-life event log, based on data of a bank's call center made available by the Technion Service Enterprise Engineering Center[2]. To this end, the procedure and metrics from Section III are implemented in the R programming environment given the wide range of packages that are available to perform data manipulations and analyses.

A. Process description

When a call enters the call center, it is directed to a voice response unit (VRU) where automated voice information guides the caller. Some features of the VRU allow callers to service themselves, without interference of a call center agent. The other calls are redirected to a queue, after which they are connected to an agent [12]. As the VRU is permanently available and has ample capacity, this section will focus on retrieving insights on the availability of the call center agents. Consequently, a resource is a call center agent in this section.

B. Data preparation

To convert the dataset to the analysis format of Section III-A, a unique case identifier is added to each incoming call. The start and complete timestamp of an activity instance, i.e. handling a particular call, is a combination of the *date* field with *ser_start* and *ser_exit*, respectively. Case arrival is, in principle, equated to *vru_exit* as this reflects the end of the prior activity. However, for 117819 of the 347380 calls involving resource intervention, the data indicates that service started before the VRU procedure is finished. This can probably be largely attributed to synchronization issues between system clocks as the observed differences between both timestamps are 1 or 2 seconds for 117055, or over 99%, of these calls. For calls where service starts before the VRU procedure ends, case arrival is equated to the start of service.

C. Daily availability records retrieval and post-processing

After data preparation, the event log analysis outlined in Section III-C can be conducted. For illustrative purposes, the analysis is executed for resources *AVNI* and *MORIAH* for January. The tolerance percentage used during active period merging is set to 10% to obtain a reasonable tolerance as calls tend to be rather short. For the identification of pending cases, a single queue is assumed as it seems warranted that incoming calls can be handled by each resource.

Table VII shows some key figures on the resulting daily availability records, both before post-processing and after each post-processing step. From the second and third column, representing the situation before post-processing, it follows that the daily availability records are highly fragmented. For instance, the median number of unavailable periods for *AVNI* equals 106, but these periods are very short as the median duration equals 0.63 minutes. The fragmented output can be attributed to the characteristics of the call center process. The duration of calls is relatively short, with the median call duration of 3.11 minutes for *AVNI* and 2.08 minutes for *MORIAH*, leading to scattered periods during which the resource is certainly available. Moreover, waiting times are limited. For all calls connected to an agent in January, the median and mean waiting time equalled 4 and 35.21 seconds, respectively, with a standard deviation of 62.92 seconds. Consequently, the notion of pending cases can only mark relatively short periods as unavailable in the call center process. Both of these process traits contribute to the fragmented character of the obtained daily availability records and underline the need for post-processing, as described in Section III-D.

The first post-processing step limits the analysis to time periods between the start of the first activity instance associated to the resource and the end of the resource's last activity instance on a particular day. As the time outside the working day is composed of a mixture of unavailable and unknown periods, the fourth and fifth column of Table VII shows a reduction in the prevalence of these period types. For example, the median number of unavailable periods per working day for *AVNI* reduces from 106 to 22.

The unassigned periods after the first post-processing steps are rather short with a median duration of 1.37 and 1.13 minutes for *AVNI* and *MORIAH*, respectively. This observation also holds when the unassigned periods are split up according

[2]http://ie.technion.ac.il/Labs/Serveng

TABLE VI
OVERVIEW OF RESOURCE AVAILABILITY METRICS

Metric	Description	Recommended calculation base*
Available period metrics		
total available time	summary statistics related to the total time during which the resource is available for the process during a working day***	APP2, APP3
number of available periods	summary statistics on the number of available periods of the resource within the process during a working day	APP2, APP3
available period length	summary statistics related to the length of individual time periods during which the resource is available for the process during a working day***	APP2, APP3
Unavailable period metrics		
total unavailable time	summary statistics related to the total time during which the resource is unavailable for the process during a working day***	APP2, APP3
number of unavailable periods	summary statistics on the number of unavailable periods of the resource within the process during a working day	APP2, APP3
unavailable period length	summary statistics related to the length of individual time periods during which the resource is unavailable for the process during a working day***	APP2, APP3
Idle period metrics**		
total idle time	summary statistics related to the total time during which no work is present for the resource within the process during a working day***	APP1
number of idle periods	summary statistics on the number of periods during which no work is present for the resource within the process during a working day	APP1
idle period length	summary statistics related to the length of individual time periods during which no work is present for the resource within the process during a working day***	APP1
Other metrics		
working day length	summary statistics on the length of a working day	BPP, APP1, APP2, APP3
active time	summary statistics related to the time during which the resource is performing activities for the process during a working day***	event log, APP1

** BPP: daily availability records before post-processing, APP1: daily availability records after post-processing step 1, APP2: daily availability records after post-processing step 2, APP3: daily availability records after post-processing step 3*
*** These metrics cannot be computed for APP2 and APP3 as unassigned periods, expressing periods during which no work is present, are no longer included in the daily availability records.*
**** can be expressed both in absolute terms as relative to the length of that working day*

to the adjacent period type, as shown in Table VIII. Consequently, the imputation approach outlined in Section III-D2 is applied as a second post-processing step. The corresponding column of Table VII illustrates that imputation causes a reduction in the number of the number of available and unavailable periods and an increase of their length . This is due to the fact that, e.g., an unassigned period with two adjacent available periods are merged into a single longer available period.

After the imputation of unassigned periods, the unavailability periods are still quite short with a median duration of 1.67 and 0.72 minutes for *AVNI* and *MORIAH*, respectively. These periods represent time frames during which a resource could have taken a call, but refrained from doing so. Given the shortness of the unavailability periods, it seems warranted to specify an unavailability period threshold in a third post-processing step. Suppose that call center management states that unavailability periods of less than five minutes do not express genuine unavailability, which can still be seen as a rather conservative value. Applying this threshold leads to a median number of unavailable periods of three for *AVNI* and six for *MORIAH*. The associated median duration of these periods equals 9.66 and 8.55 minutes respectively, which can express legitimate unavailability due to, e.g., a break.

The aforementioned post-processing steps improve the comprehensibility of the daily availability records. This enables managers or employees to gain insight in resource allocation in a particular process. The daily availability records for the

TABLE VIII
UNASSIGNED PERIODS AFTER FIRST POST-PROCESSING STEP

Prior period	Next period	N	Unassigned period duration mean*	sd*	median*
AVNI					
available	available	551	2.95	4.11	1.65
unavailable	unavailable	294	1.58	6.16	0.55
available	unavailable	172	3.03	4.36	1.59
unavailable	available	93	4.45	5.84	2.40
MORIAH					
available	available	283	4.36	5.05	2.73
unavailable	unavailable	416	1.29	1.88	0.55
available	unavailable	280	1.70	2.62	0.73
unavailable	available	155	3.02	4.44	1.70

** expressed in minutes*

first four days of January are visualized in Figures 7 and 8 for *AVNI* and *MORIAH*, respectively. These figures show the differences in resource availability during consecutive days for the same resource and between the two resources on the same day, stressing the importance of considering the temporal dimension of availability. Moreover, with *AVNI* on January 4th as an exception, the length of availability interruptions seems to be rather limited. The fact that multiple shorter unavailability periods are still included originates from the fairly conservative unavailability threshold of five minutes used during post-processing.

TABLE VII
EFFECT OF POST-PROCESSING IN REAL-LIFE EVENT LOG

	Before post-processing		**After post-processing**					
			Step 1		**Step 2**		**Step 3**	
Resource	*AVNI*	*MORIAH*	*AVNI*	*MORIAH*	*AVNI*	*MORIAH*	*AVNI*	*MORIAH*
Available periods								
mean number per day	33.52	53.16	33.52	53.16	11.48	38.26	4.04	6.42
sd number per day	9.50	13.94	9.50	13.94	4.86	16.77	1.97	1.82
median number per day	36.0	51.0	36.0	51.0	12.0	42.0	4.0	7.0
mean duration*	8.41	4.56	8.41	4.56	32.78	8.99	95.68	58.50
sd duration*	10.82	5.44	10.82	5.44	36.76	12.68	88.07	58.81
median duration*	5.00	2.70	5.00	2.70	19.32	4.63	66.63	40.68
Unavailable periods								
mean number per day	98.17	139.25	23.17	59.16	10.92	37.26	3.30	5.72
sd number per day	65.09	82.37	11.59	23.75	4.43	16.77	1.82	2.49
median number per day	106.0	163.0	22.0	57.0	11.5	41.0	3.0	6.0
mean duration*	1.49	1.16	1.10	0.82	4.89	2.42	13.46	10.80
sd duration*	3.14	2.19	2.00	1.06	10.88	4.25	17.38	5.98
median duration*	0.63	0.60	0.48	0.57	1.67	0.72	9.66	8.55
Unassigned periods								
mean number per day	110.61	130.04	46.25	59.68	-	-	-	-
sd number per day	68.11	81.69	10.83	11.92	-	-	-	-
median number per day	129.0	146.5	46.5	60.0	-	-	-	-
mean duration*	9.46	8.27	2.73	2.39	-	-	-	-
sd duration*	69.79	64.93	4.98	3.63	-	-	-	-
median duration*	1.09	0.98	1.37	1.13	-	-	-	-

* *expressed in minutes*

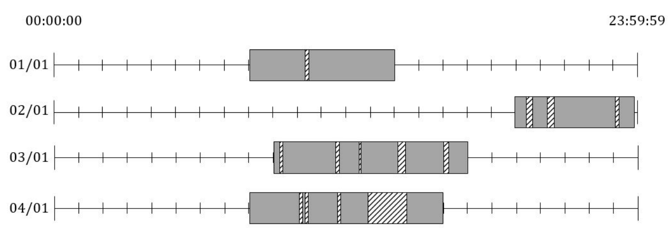

Fig. 7. Daily availability records resource *AVNI* - illustration

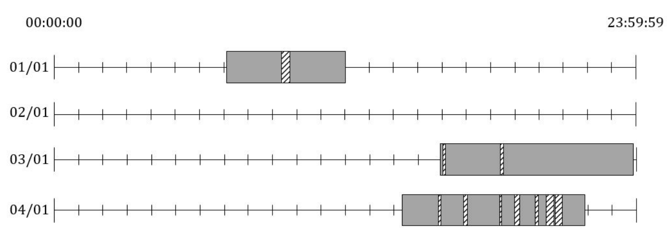

Fig. 8. Daily availability records resource *MORIAH* - illustration

D. Resource availability metrics

The resource availability metrics from Section III-E generate a more aggregated perspective on resource availability. The obtained results for resource *AVNI* and *MORIAH* are shown in Table IX. Not all metrics proposed in Section III-E are included in this table as the other metrics are already included in Table VII. The number of available and unavailable periods and their length are reported in the last two columns of Table VII, while the remaining idle period metrics correspond to the unassigned period values in the fourth and fifth column. This is due to the fact that unassigned periods in the daily availability records are conceptually equivalent to idle periods, i.e. periods during which no work is present for the resource.

Several insights can be gained from the resource availability metrics. For example, the idle time is rather high with a median proportion of 30% of the working day for *AVNI* and 34% for *MORIAH*. However, the idle time is highly fragmented as the median number of idle periods equals 46.5 for *AVNI* and 60 for *MORIAH*. This is due to the specificities of the call center process as the incoming call volume is unpredictable and customer service requirements impedes the collection of calls in a backlog to avoid idle periods. Moreover, alternative activities that management might allocate to agents to fill idle periods should be easily interruptible when new calls arrive.

Table IX presents the values aggregated over all working days. However, a more refined analysis can be conducted by, e.g., combining working days by the day of the week. This does not influence the calculation method of the metrics, but changes the way the results of individual working days need to be combined. Instead of combining all of them, they can be grouped by, e.g. the day of the week they represent.

For illustrative purposes, only the mean, standard deviation and median are included as summary statistics in Table IX. However, other summary statistics such as the minimum, maximum, the interquartile range and e.g. a 95% percentile can also be helpful and, hence, be reported to a business user.

V. RELATED WORK

Resource availability has several meanings in literature. [8] and [13] use the term to express whether a resource can accept new tasks during a time period, i.e. when no running cases or cases that need to be dealt with are present. Consequently, it entails a situation in which no work is available for a resource while it is assigned to the process. Conversely, this paper focuses on periods during which the resource is allocated to

TABLE IX
RESOURCE AVAILABILITY METRICS IN THE CALL CENTER LOG - ILLUSTRATION

Metric	**Calculation base***	**Values for each resource** *AVNI*		*MORIAH*	
		absolute	*relative*	*absolute*	*relative*
Total available time**	APP3				
mean		6.44	0.91	6.26	0.87
sd		1.70	0.08	0.95	0.06
median		6.93	0.94	6.55	0.86
Total unavailable time**	APP3				
mean		0.74	0.10	1.03	0.14
sd		0.60	0.07	0.40	0.05
median		0.54	0.07	1.08	0.14
Total idle time**	APP1				
mean		2.10	0.30	2.38	0.34
sd		0.87	0.14	0.87	0.14
median		1.96	0.25	2.49	0.34
Working day length**	BPP				
mean		7.12	-	7.24	-
sd		1.86	-	1.12	-
median		7.95	-	7.78	-
Active time**	event log				
mean		4.67	0.66	4.03	0.55
sd		1.67	0.14	1.13	0.10
median		5.20	0.66	4.38	0.56

** BPP: daily availability records before post-processing, APP1: daily availability records after post-processing step 1, APP2: daily availability records after post-processing step 2, APP3: daily availability records after post-processing step 3*
*** absolute time values expressed in hours*

the process, regardless of whether the resource can accept new tasks. This approach is related to [1] and [2], where resource availability expresses the percentage of time during which a resource can perform activities. However, the availability concept in this paper also takes into account the temporal dimension of availability and intermediate interruptions.

Significant work is done on resource assignment to activities. It is studied from several perspectives, including resource allocation taking into account operational issues such as the task deadline [14], under emergency circumstances [15], with a focus on resource cooperation [16], [17] and with a link to business process improvement [18]. Other authors aim to identify rules underlying resource assignment. To this end, techniques such as decision tree analysis [19], [20], association rule mining [21], naive Bayes classifiers [20], support vector machines [20] and declarative modeling [22] are used. In the majority of these references, no explicit attention is attributed to resource availability [16], [18]–[22]. Other authors briefly mention resource availability. [17] assume that resources are permanently available, while [15] state that resource availability is not taken into account as resources work in shifts. [14] posit that availability is established in a subjective way based on planned and unplanned resource absence and its workload. [20] indicate that a supervisor must decide upon the feasibility of the suggested allocation based on, e.g., the resources availability. In general, resource availability places additional constraints on resource assignment. The work presented in this paper can contribute to shaping these constraints by retrieving objective resource availability insights from event logs.

Existing work on event logs usage to investigate resource availability is scarce as only [1] and [11] touch upon this matter. Even though [11] focuses on activity duration, the authors also retrieve the start and end of a working day from an event log. As mentioned in Section III-D1, the start of the first activity instance recorded for that resource on a particular day and the end of the last instance are used as proxies [11].

In [1], a resource's available period corresponds to the elapsed time between the start of the first activity execution of that resource and the end of the last one in a particular time horizon such as a working day. This implies that a resource is permanently available within this period as intermediate unavailability periods such as breaks are not considered. The length of the availability period is used to calculate an availability measure, which is obtained by dividing the cumulative length of the active periods by the length of the available period [1]. However, phrasing this as resource availability measure can be misleading as it conceptually represents the active time rather than the available time for the process. This paper, more in particular the resource availability metrics, clearly treats resource active time as a autonomous concept.

When specifying daily availability records, this paper considers two resource states, i.e. available and unavailable. In contrast, [1] and [2] consider inactive, busy and ready as possible states. While the inactive state reflects the resource's unavailability for the process, both busy and ready can be mapped to the available state. While busy reflects the execution of activities by the resource, ready corresponds to a situation in which the resource is assigned to the proces, but waiting for work [1], [2]. However, both situations are consistent with the notion of availability in this work.

Using daily availability records as an input, an extensible set of resource availability metrics is defined. These can be positioned in the utilization category in the framework of [23] as they relate to the assignment and activeness of a resource within a process. Moreover, conceptual similarities can be found between the specified metrics and literature on the use of resources, for instance between the idle period metrics and the resource utilization measure on this topic in [24].

VI. Conclusion and future work

This paper focused on the retrieval of resource availability insights from an event log. Daily availability records are constructed, which express a resource's allocation to a particular process and take into account (i) the temporal dimension of availability and (ii) intermediate availability interruptions. To this end, the explicit and implicit information in the event log on resource availability is maximally exploited. Both the daily availability records and the introduced resource availability metrics, conveying insights on a more aggregated level, allow managers and employees to gain understanding in resource allocation to the process. This is especially useful in less structured business environments where resources are involved in multiple processes. However, as is also shown in the real-life event log of a call center, log analysis can result in fragmented daily availability records. Consequently, post-processing might be required to increase their comprehensiveness and obtain meaningful values for particular metrics.

Even though this paper provides valuable starting points related to the use of event logs to model and analyze resource availability, several areas for future work can be distinguished. Firstly, the developed procedure can be tested on artificial data to determine its ability to rediscover the (un)availability of resources in a controlled environment. Moreover, the influence of alternative modeling decisions such as the use of an alternative imputation approach in the second post-processing step can be explored. Besides evaluation on artificial data, additional real-life event logs can be studied to show the procedure's applicability in a variety of business contexts.

Secondly, both the daily availability records and the resource availability metrics can be subject to further aggregation. The daily availability records can be aggregated into aggregated resource schedules, expressing stable availability patterns for the process. These can, e.g., be used as an input in a business process simulation model. The resource availability metrics, currently defined on the level of an individual resource, can be aggregated over all resources or groups of similar resources. Moreover, the evolution of these metrics over time can also provide valuable insights.

Finally, resource availability records can be linked to case attributes or variables expressing the system state. While resource availability is currently approached from a time perspective, it can also depend upon the workload for instance. Future work can focus on discovering rules underlying resource's availability and investigate deviations from stable availability patterns due to, e.g., an approaching deadline.

References

[1] J. Nakatumba, "Resource-aware business process management analysis and support," PhD thesis, Eindhoven University of Technology, 2013.

[2] W. van der Aalst, J. Nakatumba, A. Rozinat, and N. Russell, "Business process simulation," in *Handbook on Business Process Management*. Heidelberg: Springer, 2010, pp. 313–338.

[3] W. van der Aalst, *Process mining: discovery, conformance and enhancement of business processes*. Heidelberg: Springer, 2011.

[4] J. Recker and J. Mendling, "The state of the art of business process management research as published in the BPM conference," *Business & Information Systems Engineering*, vol. 58, no. 1, pp. 55–72, 2016.

[5] N. Martin, B. Depaire, and A. Caris, "The use of process mining in business process simulation model construction: structuring the field," *Business & Information Systems Engineering*, vol. 58, no. 1, pp. 73–87, 2016.

[6] A. Law, *Simulation modeling and analysis*. New York: McGraw-Hill, 2007.

[7] S. Leemans, D. Fahland, and W. van der Aalst, "Using life cycle information in process discovery," *Lecture Notes in Business Information Processing*, vol. in press, 2015.

[8] Z. Huang, X. Lu, and H. Duan, "Resource behavior measure and application in business process management," *Expert Systems with Applications*, vol. 39, no. 7, pp. 6458–6468, 2012.

[9] A. Senderovich, M. Weidlich, A. Gal, and A. Mandelbaum, "Queue mining: predicting delays in service processes," *Lecture Notes in Computer Science*, vol. 8484, pp. 42–57, 2014.

[10] J. De Weerdt, M. De Backer, J. Vanthienen, and B. Baesens, "A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs," *Information Systems*, vol. 37, no. 7, pp. 654–676, 2012.

[11] A. Wombacher, M. Iacob, and M. Haitsma, "Towards a performance estimate in semi-structured processes," in *Proceedings of the 2011 IEEE International Conference on Service-Oriented Computing and Applications*, 2011, pp. 1–5.

[12] A. Mandelbaum, A. Sakov, and S. Zeltyn, "Empirical analysis of a call center," Tech. Rep., 2001. [Online]. Available: http://ie.technion.ac.il/Labs/Serveng

[13] W. Zhao, L. Yang, L. Haitao, and R. Wu, "The optimization of resource allocation based on process mining," *Lecture Notes in Computer Science*, vol. 9227, pp. 341–353, 2015.

[14] A. Kumar, W. van der Aalst, and E. Verbeek, "Dynamic work distribution in workflow management systems: how to balance quality and performance," *Journal of Management Information Systems*, vol. 18, no. 3, pp. 157–193, 2002.

[15] H. Reijers, M. Jansen-Vullers, M. zur Muehlen, and W. Appl, "Workflow management systems + swarm intelligence = dynamic task assignment for emergency management applications," *Lecture Notes in Computer Science*, vol. 4714, pp. 125–140, 2007.

[16] H. Yang, C. Wang, Y. Liu, and J. Wang, "An optimal approach for workflow staff assignment based on hidden Markov models," *Lecture Notes in Computer Science*, vol. 5333, pp. 24–26, 2008.

[17] A. Kumar, R. Dijkman, and M. Song, "Optimal resource assignment in workflows for maximizing cooperation," *Lecture Notes in Computer Science*, vol. 8094, pp. 235–250, 2013.

[18] J. Xu, C. Liu, and X. Zhao, "Resource allocation vs. business process improvement: how they impact each other," *Lecture Notes in Computer Science*, vol. 5240, pp. 228–243, 2008.

[19] L. Ly, S. Rinderle, P. Dadam, and M. Reichert, "Mining staff assignment rules from event-based data," *Lecture Notes in Computer Science*, vol. 2812, pp. 177–190, 2006.

[20] Y. Liu, J. Wang, Y. Yang, and J. Sun, "A semi-automatic approach for workflow staff assignment," *Computers in Industry*, vol. 59, pp. 463–467, 2008.

[21] Z. Huang, X. Lu, and H. Duan, "Mining association rules to support resource allocation in business process management," *Expert Systems with Applications*, vol. 28, pp. 9483–9490, 2011.

[22] S. Schönig, C. Cabanillas, S. Jablonski, and J. Mendling, "Mining the organisational perspective in agile business processes," *Lecture Notes in Business Information Processing*, vol. 214, pp. 37–52, 2015.

[23] A. Pika, M. Wynn, C. Fidge, A. ter Hofstede, M. Leyer, and W. van der Aalst, "An extensible framework for analysing resource behavior using event logs," *Lecture Notes in Computer Science*, vol. 8484, pp. 564–579, 2014.

[24] R. Wild, *Operations management*. Oxford: Thomson Learning, 2003.