Simulation to analyze the impact of a Schedule-Aware Workflow Management System

R.S. Mans[1,2], N.C. Russell[1], W.M.P. van der Aalst[1], P.J.M. Bakker[2], A.J. Moleman[2]

[1] Department of Information Systems, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands.
{r.s.mans,n.c.russell,w.m.p.v.d.aalst}@tue.nl

[2] Academic Medical Center, University of Amsterdam, Department of Quality Assurance and Process Innovation, Amsterdam, The Netherlands.
{p.j.bakker,a.j.moleman}@amc.uva.nl

Abstract. Today's workflow management systems (WfMSs) offer workitems to users through specific work-lists. Users select the workitems they will perform without having a schedule in mind. However, in many environments work needs to be scheduled and performed at particular times. For example, in hospitals many workitems are linked to appointments, e.g., a doctor cannot perform surgery without reserving an operating theater and making sure that the patient is present and ready. One of the problems when applying workflow technology in such domains is the *lack of calendar-based scheduling support*.

In collaboration with the Academic Medical Center (AMC), a large hospital in the Netherlands, we developed a schedule-aware WfMS that supports the seamless integration of unscheduled (flow) and scheduled (schedule) tasks. However, before deployment of the resultant system in the hospital, a seamless integration with AMC's running healthcare processes needs to be guaranteed. Therefore, for a large and complex healthcare process, we apply *computer simulation* to validate and to investigate, for different configurations of the system, the operational performance for a selected healthcare process when supported by the schedule-aware workflow management system. One of the important characteristics of our approach is the tight coupling between the simulation model and the actual implemented system. While performing simulation experiments, parts of the system may be simulated using CPN Tools while connected to the actual system components. Our simulation experiments demonstrate that the developed schedule-aware WfMS can be safely applied in the AMC hospital.

Keywords: workflow management, healthcare, scheduling, simulation

1 Introduction

Hospitals in many countries are facing increasing pressure to increase productivity and reduce costs [17]. This is due to an increase in medical complexity with an increasing need for high quality healthcare services. In this way, high patient service levels are becoming increasingly important. It is believed that healthcare process measures and standards should focus on three aspects: effectiveness, efficiency, and quality [31]. An important patient related measure in this context is the average waiting time for an appointment in a hospital, e.g. to see a doctor or to undertake a diagnostic test. The effective execution of these kinds of tasks is often tied to the availability of specific resources. In order to guarantee the efficient utilization of resources and to ensure low waiting times, an appointment-based approach is typically utilized for scheduling the tasks performed by these resources. Often, the scheduling of these appointments is undertaken on a manual basis without having a clear view or being unaware of the necessary prerequisite tasks.

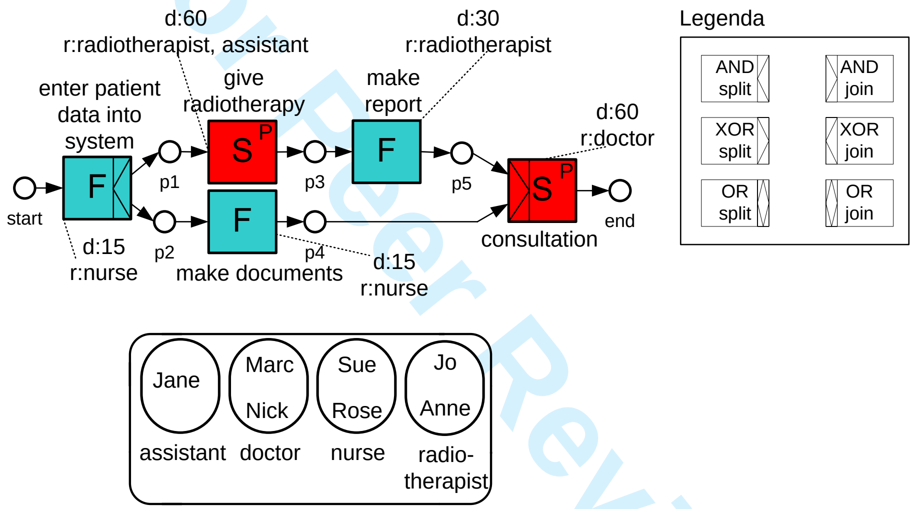


Fig. 1. Running example (at the left-side) showing *schedule* (S) and *flow* (F) tasks. The organizational model at the bottom models the roles played by people in the organization. The prefix "d:" indicates in minutes the average time needed for performing the task and prefix "r:" indicates which roles are necessary to perform the task. From each associated role, exactly one person needs to be assigned to the task. For both schedule tasks, indicated by the character "P" in the top right corner of the task, the patient is also required to be present.

To illustrate this problem, consider the hospital process for the treatment of a patient as shown in Figure 1. First a nurse enters the patient data into the system. Next a radiotherapy session takes place which is undertaken by a radiotherapist and a nurse. After the radiotherapy session, a radiotherapist needs to make a report about the radiotherapy that was given (task "make report"). In parallel, a nurse prepares the documents for the patient (task "make documents"). When these tasks have been completed, a doctor evaluates the results of the radiotherapy session (task "consultation"). Figure 1 also shows the corresponding organizational model which specifies the roles being played by various people in the organization.

From the example, it becomes clear that two kinds of tasks exist. The tasks indicated with an "F" in the figure can be performed at *an arbitrary point in time when a resource becomes available* once they are allowed to be executed. These tasks are called *flow tasks*. The tasks indicated with an "S" in the figure are called *schedule* tasks as they are performed *by one or more resources at a specified time*. For example, the "give radiotherapy" task can only be performed when a radiotherapy device is reserved and a radiotherapist and an assistant are present. Moreover, the patient also needs to be present.

For a consultation task (see last schedule task in Figure 1), it may be the case that a doctor finds out during the appointment that the results of some earlier diagnostic tests or treatments are missing. As a consequence, a new appointment needs to be made. So, for the most effective scheduling of appointments it is important that the whole workflow is taken into account thereby guaranteeing that preceding tasks can be performed on time. By doing so, unproductive time for resources, due to a canceled appointment, is avoided.

In this context, Workflow Management Systems (WfMSs) are an interesting vehicle as, based on a corresponding process definition, it supports processes by managing the flow of work such that individual workitems are performed at the right time by the proper person. Contemporary WfMSs offer workitems via worklists which show available workitems. At an arbitrary point in time, a workitem can be selected by a person from this list in order to perform the associated task.

However, an important limitation of existing WfMSs is revealed when applying this technology in a healthcare context. The effective execution of healthcare processes relies heavily on the efficient scheduling of appointments of scarce resources (e.g. doctors). This suggests that there is a *need to integrate WfMS with scheduling facilities* so that both schedule and flow tasks are supported, called a schedule-aware WfMS. In addition to the classical work-list functionality, the concept of a calendar is also used for allocating work to healthcare practitioners. Therefore, in this paper, we present a *comprehensive conceptual model capable of serving both as a specification and simulation model for the application domain*. So, the conceptual model can be used as a *specification* for the subsequent realization of the system.

The modeling and development of the schedule-aware WfMS was undertaken in conjunction with the Academic Medical Center (AMC), a large hospital in the Netherlands. In the AMC, there are many patient-centered, critical processes for which continuous operation must be guaranteed. Therefore, the introduction of a new technology system requires a seamless integration with the running operational processes of the hospital. For the AMC, it is of the utmost importance that the operational performance of a healthcare process, when supported by the schedule-aware WfMS, needs to be at least as good as the operational performance of the process executed in real-life. Moreover, instead of validation only, it is important to asses the *impact* on the operational performance for different configurations of the system. For example, the hospital might want to change the way appointments are scheduled or add more capacity for seeing patients.

However, an additional important goal is the deployment of the realized system in the AMC. So, this means that we are looking after the *validation of the operational performance of the realized system and its impacts on the operational performance for different configurations of the system.*

To address these kind of issues, we focus on the use of *computer simulation* to gain insight into the operational performance of the developed and implemented schedule-aware WfMS. The strength of simulation is that it can be used to evaluate systems, but also to evaluate alternative configurations of a system. In this way, it allows for a "what-if" analysis.

In order to be able to perform simulation experiments, the following approach will be pursued. Initially, to understand how a WfMS can be extended with scheduling facilities, we have developed a conceptual model, in terms of a CPN model [18], which is a *complete* and *formal* specification of such a system. As the conceptual model is realized in CPN Tools, it can be executed, simulated, and analyzed. In order to realize the functionality contained in the conceptual model, we *incrementally* mapped it to an operational system based on off-the-shelf software components and applications available. Using both the CPN conceptual model and the prototype implementation, we can "easily" replace one or more components in the CPN by their realizations. In order to do this, the simulated components in CPN Tools are connected to the corresponding components in the actual system. As CPN models can also be used for simulation-based performance analysis, parts of the system can be simulated while being connected to the actual system components. This approach allows us to investigate the impact of a WfMS augmented with scheduling facilities while being connected with the actual system components.

The remainder of this paper is organized as follows: in Section 2 we introduce our conceptual model and explain how a WfMS can be augmented with scheduling facilities. In Section 3 we outline the simulation experiments that will be performed and explain how they are performed using the conceptual model and the prototype implementation. Section 4 discusses related work. In Section 5 we provide a discussion on the simulation model and the performed experiments. Section 6 concludes the paper.

2 Workflow Management Systems and Scheduling

In this section, we elaborate on how a WfMS can be augmented with scheduling facilities. First, some concepts need to be introduced. Then, we describe the conceptual model and present the design of a WfMS augmented with scheduling facilities. Note that the concepts that will be introduced in the next section are needed in order to augment WfMSs with scheduling facilities. However, in Section 3, in which the simulation model will be discussed that is going to be used for the simulation experiments, additional empirical notions together with the relevant data are introduced. Thee notions are needed for performing simulation experiments.

2.1 Concepts

We assume that the reader is familiar with basic workflow management concepts, such as case, role, and so on [2]. Using the process model shown in Figure 1, we will show how a workflow language can be integrated with scheduling functionality.

As already indicated, a distinction is made between two kinds of tasks. A flow task is performed at an arbitrary point in time when a required resource becomes available (typically as soon as possible). As only one resource is needed, it is sufficient to define a *single* role for each of them[3]. These tasks can be presented to the user using the standard *work-list* facility in a WfMS. For example, for the flow task "make documents" the work may either be performed by "Sue" or "Rose".

In contrast, schedule tasks are performed by one or more resources at a specified time. As multiple resources can be involved, with different capabilities, it is necessary to specify which kinds of resources are allowed to undertake the task. Multiple resources may be defined for a schedule task where for each role specified, only *one* resource is involved in the actual performance of the task. For example, in Figure 1, the schedule task "give radiotherapy" may be performed by "Jane" and "Rose", but not by "Sue" and "Rose". Note that a resource involved in the performance of a schedule task may also be a physical resource such as medical equipment or a room. In this way, for them a separate role may be specified. Furthermore, for a schedule task, the patient may also be involved which means that the patient is also considered to be a necessary resource for these tasks. Note that the patient is not involved in the actual execution of the task but is a passive resource who needs to be present whilst it is completed. For this reason, the patient is not added to any of the roles for the task, nor is the patient defined in terms of a separate role. Instead, it is sufficient to identify for which schedule tasks the patient needs to be present so that the availability of the patient can be taken into account when making appointments.

For presenting the appointments made for schedule tasks to users, the concept of a *calendar* will be used. More specifically, each resource will have its own calendar in which appointments can be booked. Note that each patient, who is considered a passive resource, also has its own calendar. An appointment either refers to a schedule task which needs to be performed for a specific case or to an activity which is not workflow related (e.g. playing basketball). An appointment appears in the calendars of all resources that are involved in the actual performance of the associated task. Note that an appointment for a schedule task for which a workitem does not yet exist, can be booked into the calendar of a resource. However, when the workitem becomes available it has already

[3] There also exist approaches for which more roles may be defined for a flow task, but this is not the focus of our work.

been determined when it will be performed and by whom. Note that sometimes workitems need to be rescheduled because of anticipated delays in preceding tasks.

In order to be able to determine at runtime the earliest time that a schedule task can be started, information about the duration of every task needs to be known. For example, in Figure 1, for each task the average duration is indicated by prefix "d:". For example, the task "physical examination" takes 60 minutes on average. Note that the duration of a specific task can be described by a probability distribution.

2.2 System Design

The conceptual model which defines the exact behavior of a WfMS augmented with scheduling facilities is defined in terms of a CPN model which can be executed in CPN Tools [18]. CPNs provide a well-established and well-proven language suitable for describing the behavior of systems exhibiting characteristics such as concurrency, resource sharing, and synchronization.

Formalizing a system using CPNs offers several benefits. First of all, building such a net allows for *experimentation*. So, the model or parts of it can be executed, simulated and analyzed which leads to insights about the design and implementation of the system. Second, a complete model of the system allows for *testing* parts of the system that are implemented. Given that a CPN consists of several components, we can "replace" one or more components in the CPN by the concrete implementation of these components by making connections between the CPN model and software which realizes the component. As the CPN is an executable model this allows us to test numerous scenarios which in turn facilitates the discovery of potential flaws in both the architecture and the implementation. Moreover, this approach also allows us to perform simulation experiments. This will be discussed in more detail in Section 3.

Figure 2 shows part of the topmost net in the CP Net model and gives an idea of the main components in the system and the interfaces between them. As can be seen in the figure, there are four substitution transitions. They represent the major functional units in the system which are explained in detail below.

- The **workflow engine** is the most important component of the workflow system as it is the heart of the system. Based on the business process definition, the engine routes cases through the organization and ensures that the tasks of which they are comprised are carried out in the right order and by the right people. In order to do this, the engine takes care of offering workitems to users, once they become available for execution.
- The **workflow client application** communicates the distributed workitems to the users so that they can select and perform them. Workitems that correspond to flow tasks are advertised via the worktray. The appointments that are created for schedule tasks are advertised via a calendar. Once a workitem becomes available for such an appointment, the work can be performed. However, where appointments have been made, users can express

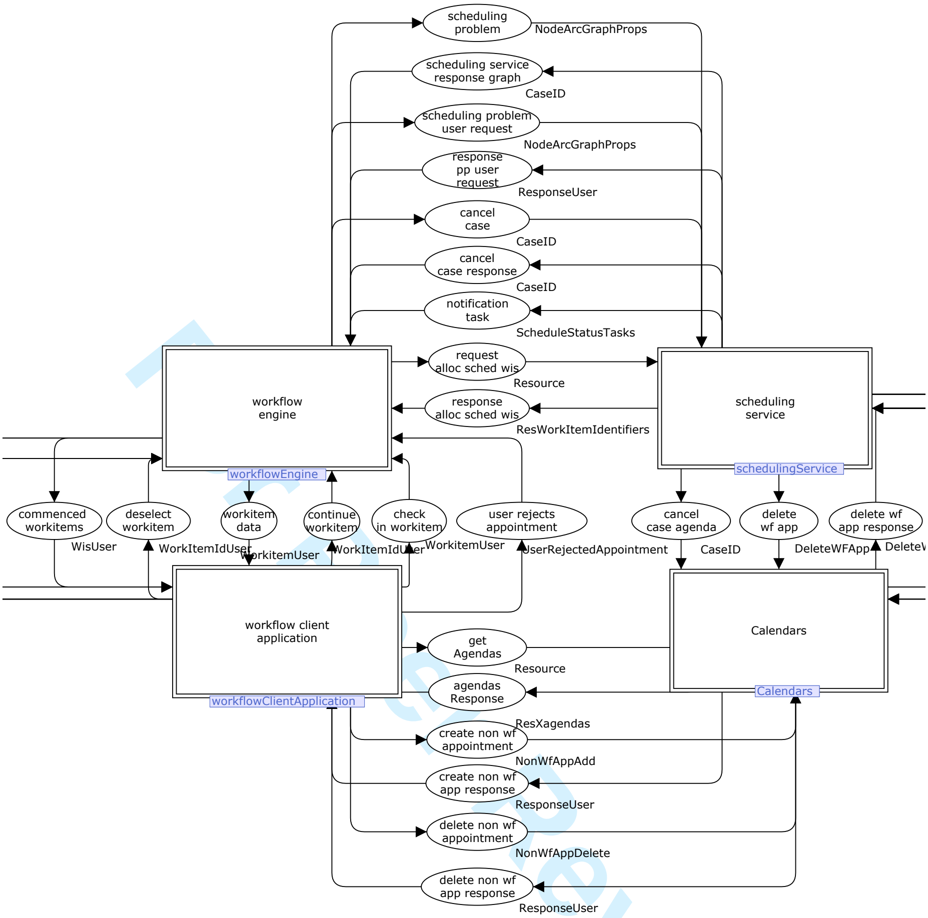


Fig. 2. The topmost model of the conceptual model realized in terms of CP Nets.

their dissatisfaction with the nominated scheduling by requesting: (1) the rescheduling of the appointment, (2) the rescheduling of the appointment to a specified date and time, or (3) the reassignment of the appointment to another employee. The workflow client is also responsible for indicating whether limited time is left in which to undertake workitems related to preceding tasks for an upcoming appointment.

– The **scheduling service** component provides the scheduling capabilities needed by the system. The scheduling service behaves in a passive way and its operation must be explicitly triggered. Scheduling is done sequentially on a case-by-case base. Once a scheduling problem is received for a case, the scheduling service needs to determine whether some of the schedule tasks need to be (re)scheduled. A scheduling problem is represented as a graph which contains all the scheduling constraints for a case which are imposed by

the engine (e.g. the ordering of tasks in the corresponding process definition for the case and the current state of the case).
Several distinct issues need to be addressed during the scheduling process. First of all, the scheduling of tasks needs to occur in the same order as the sequence of schedule tasks in the accompanying process definition for the case and there should be sufficient time between two scheduled tasks. When rescheduling appointments, these constraints also need to be satisfied. For example, in Figure 1, first the "give radiotherapy" task should be scheduled, followed by the "consultation" which needs to occur at least 30 minutes later. Second, for the actual scheduling of an appointment multiple roles can be specified for a schedule task. For each role specified a resource needs to be selected, i.e., the number of roles determines the number of resources involved in the actual performance of the task. If the patient for which the case is performed also needs to be present at an appointment, then this also needs to be taken into account. The scheduling service only books an appointment in the calendars of those resources which need to be present at the performance of the task, being the performers of the task and the patient (if needed).
Based on these requirements several scheduling strategies are possible. The ones used during our simulation experiments will be discussed in Section 3.

– The **calendar** component is responsible for providing a view on the calendars of resources and for manipulating their contents. It is possible for resources to create / delete appointments or to retrieve information about the appointments that have been made. Note that for each patient for which a case is performed also has its own calendar of which the content may be manipulated.

3 Simulation

For the uptake of the system by the hospital, discussed in the previous section, a seamless integration with the running operational processes is of the utmost importance. Additionally, for the AMC, it is vital that the operational performance of a healthcare process, when supported by the schedule-aware WfMS, needs to be at least as good as the operational performance of the process executed in real-life. In addition, different configurations of the system might be relevant for investigation, like additional capacity for seeing patients, or to change the way appointments are scheduled (e.g. a different scheduling strategy).

Therefore, in this section, we elaborate on the use of *computer simulation* to investigate the operational performance of a schedule-aware WfMS before its deployment. In general, simulation focusses on steady-state behavior in order to allow for decision-making at the strategic (e.g. market share) or tactical level (e.g. amount of slack for a certain surgery). However, as our goal is to deploy our system in a healthcare environment, we want to investigate how the system would fit into the current organization, cf. *the current working processes*. In other words, we want to investigate the effects of the system in the current situation instead of studying the long-term effects of a strategic decision. For example, does the system have a negative effect on the average waiting time for an appointment. Note that there is no “steady state” in AMC’s dynamic environment. Moreover, we are particularly interested in the differences between the current situation and a new situation using our schedule-aware WfMS.

Below, we first discuss the healthcare process that is used for the simulation experiments in Section 3.1. The simulation model will be discussed in more detail in Section 3.2. In the simulation model some components of the realized system have been included. This will be discussed in more detail in Section 3.3. The validation of the simulation model and the simulation experiments performed are elaborated in Sections 3.4 and 3.5, respectively.

3.1 Gynecological oncology healthcare process

For the simulation experiments, we take an existing healthcare process from a hospital and investigate the impact of our system on several performance indicators specified for it. The healthcare process that will be studied is the diagnostic process of patients visiting the gynecological oncology outpatient clinic at the AMC hospital. The gynecological oncology department is concerned with the diagnosis and treatments of patients suffering from cancer. The process deals with the diagnostic process that is followed by a patient who is referred to the AMC hospital for treatment, up to the point where the patient is diagnosed. For the simulation we focus on the regular path followed by non-acute patients as shown in Figure 3[4].

At the beginning of the process, a doctor in a referring hospital calls a nurse or doctor at the AMC hospital (tasks “write down data patient and make decision” and “ask information from doctor referring hospital”) resulting in an appointment being made for the first visit of the patient (task “make conclusion”). Before the first visit of the patient several administrative tasks need to be requested (e.g. task “send fax to pathology”). At the first consultation, the doctor decides which diagnostic tests are necessary before the next visit of the patient. A doctor can make a selection from the following medical tests that may be performed: a lab test, an X-ray, an MRI scan, a CT-scan, a preoperative assessment, and an examination under anesthetic. For the first visit (task “make conclusion”), MRI (task “MRI”), CT (task “CT”), and examination under anesthetic (task “examination under anesthetic”) an appointment is required. These are schedule tasks as indicated by the calendar icon in Figure 3. The other medical tests: lab (task “lab”), and X-ray (task “x-ray”) are walk-in facilities for which no appointment is required. So, these are flow tasks as indicated by a single person icon in the figure for them. A special situation applies for the preoperative assessment for which a walk-in facility exists (task “walk-in pre-assessment”), but also an appointment can be made (task “pre-assessment”). So, task “walk in pre-assessment” is a flow task and task “pre-assessment” is a schedule task.

[4] In this figure the YAWL notation is used to describe the ordering of tasks. Note that later YAWL is used as a workflow engine.

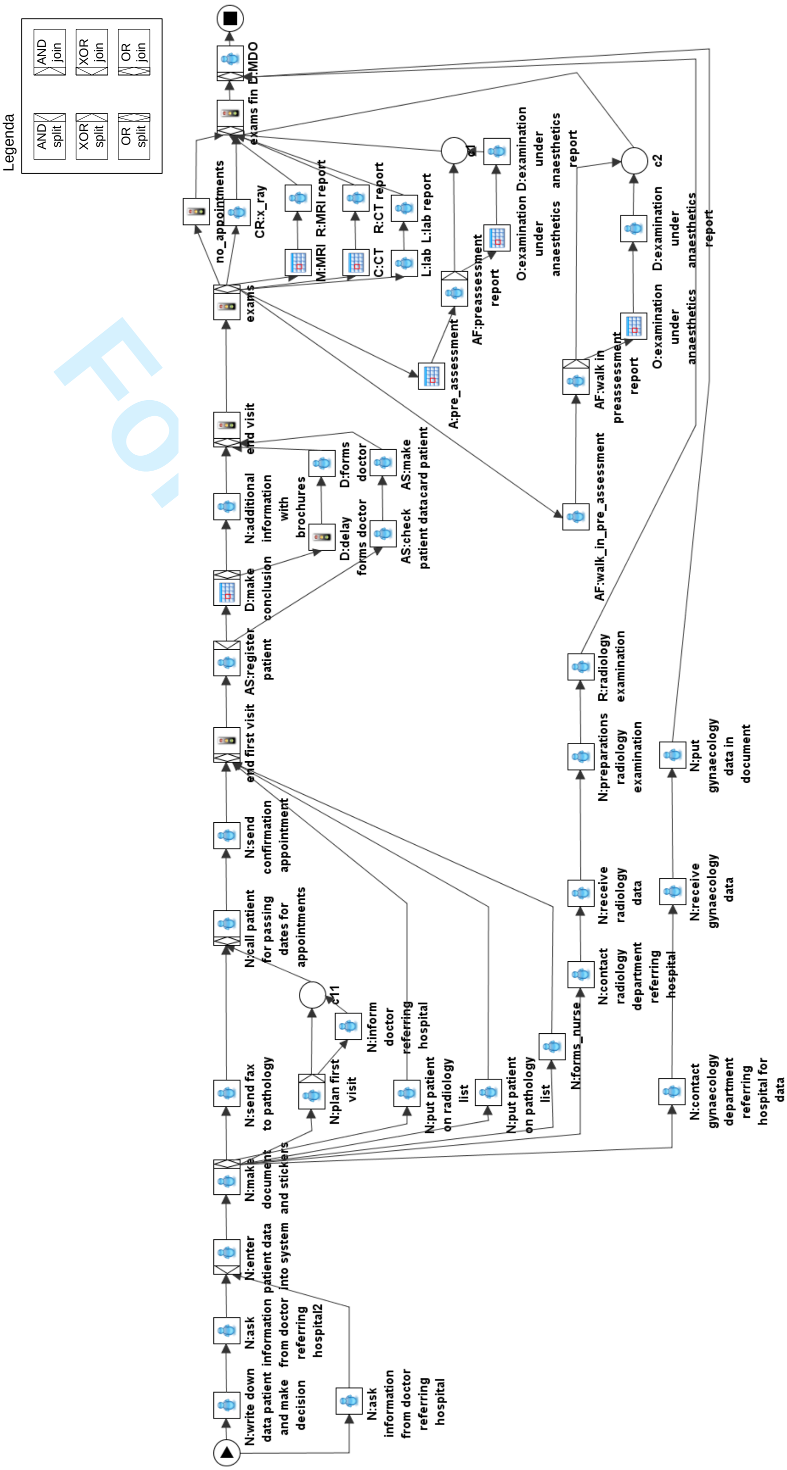


Fig. 3. Screenshot of the YAWL model showing the diagnostic process of the gynecological oncology healthcare process. The flow tasks are indicated by a person icon and the schedule tasks are indicated by a calendar icon. The tasks indicated by a traffic

If during the telephone call at the very beginning of the process, it is already clear that diagnostic test(s) are necessary, then appointments are made for them. Note that for the MRI, CT, pre-assessment, and examination under anesthetic tasks we do not consider the preceding tasks at the respective departments as including them would significantly complicate the simulation model. Our experience is that for these departments, once an appointment is known, the preceding steps are performed in time so that an appointment can always take place. At the bottom of the figure a sequence of tasks are modeled (starting with task "contact radiology department referring hospital") involving contacting the radiology department of the referring hospital to request radiology data. Moreover, a second sequence of tasks are modeled (starting with task "contact gynecology department referring hospital for data") involving contacting the referring gynecology department and requesting they send their data to the AMC.

For this healthcare process we perform several different simulation experiments for the time period from 02-07-2007 to 19-03-2008 (9 months). During this time period, a group of 142 patients follow the process depicted in Figure 3. This means that the execution of the required tasks in the process is simulated for each patient. Note that this also involves the required (re)scheduling of appointments for the schedule tasks. As performance measures we consider several performance indicators specified by the AMC for the healthcare process and compare them with the realized values of these indicators.

3.2 Simulation model

In Figure 2, part of the topmost net of the conceptual model is shown, presenting the main components of the schedule-aware WfMS. The set-up of the simulation model is as follows.

Figure 4 shows schematically the set-up of the simulation model that will be used for the experiments, together with its inputs and outputs. In the figure, we see the main components of the system: workflow engine, scheduling service, workflow client application, and calendars. Additionally, we see rectangles from which an arc leads to a given component. These rectangles indicate the inputs that are used by the simulation model before or during a simulation run. These inputs are either required for initialization purposes or to steer the simulation. Moreover, the rectangle with name "System" shows that the system is split-up in a CPN part (rectangle with name "CPN") and a prototype part (rectangle with name "Prototype") in which parts of the actual implemented system are used. Below, each aspect of the simulation model will be briefly discussed.

Performance measures

In the AMC, for the group of patients we are considering some service levels have been defined for which reliable historical data could be easily obtained. As we are dealing with a group of patients suffering from cancer, a quick diagnosis is of the utmost importance. The following patient related service levels are defined for 90% of the patients:

- *First visit*: The first visit should take place within seven calendar days of the initial telephone call.

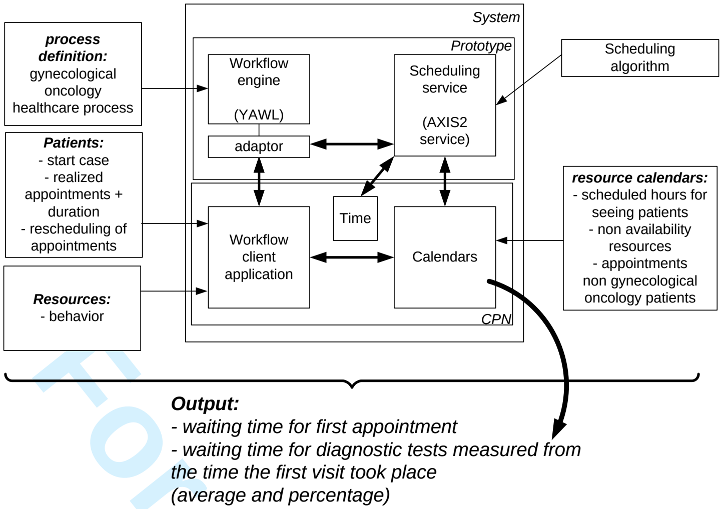


Fig. 4. Overview of the set-up of the simulation model with corresponding inputs and outputs.

– *Diagnostic steps*: All diagnostic tests should be completed within 14 calendar days of the first visit.

Therefore, we define the following performance measures for the simulation model:

– the waiting time for the first visit at the outpatient clinic of gynecological oncology (task "come to a conclusion"), measured from the time the appointment is made till the appointment takes place.
– the waiting time for the diagnostic tests performed (tasks "MRI", "CT", "pre-assessment" and "examination under anesthetic"), measured from the time that the first visit takes place till the time of the respective appointment for the diagnostic test.

For these performance measures we calculate the *average* as this allows for comparison with historical data (discussed in more detail in Section 3.4) and for comparing differences in outcomes of experiments. Moreover, for determining whether the patient related service levels are met, we also calculated the *percentage* of patients that have the first visit within seven calender days and the percentage of patients of which all diagnostic tests are performed within 14 calendar days of the first visit.

Of course many more performance measures can be imagined. However, in that case, data that can be used for comparison could not be easily obtained (so that either lengthy observations are needed or that patient documents need to be checked which are both very time consuming) or is are stored in any of the information systems of the AMC.

Note that with regard to the validation and the experiments to be performed, in which certain input parameters are manipulated, it is vital for the AMC to clearly understand the resulting effect onto the performance measures defined above. In other words, the focus is on delta analysis rather than absolute performance metrics. One of the challenges that needs to be faced in this context is that the healthcare process we are studying is not in a steady-state. For example, due to holidays patients are arriving at a lower rate, doctors are less available and so on. In order to allow for delta analysis, certain environment variables will be controlled instead of approximating them by a stochastic distribution[5]. When discussing the inputs of the simulation model below, these controlled variables will be discussed.

Resources

First, we elaborate on the resources required for performing the tasks in the gynecological oncology process. The tasks in Figure 3 have different requirements with regard to the resources that are allowed to complete the tasks. So, every task has a prefix ending with a colon, indicating the required role for completing a certain task. Table 1 shows the specific role associated with each prefix and specifies the number of resources belonging to a role.

Table 1. For every prefix in Figure 3, the specific role is indicated together with the number of resources belonging to the role.

Prefix	Role name	Number of resources	Kind of task
N	nurse	3	flow
AS	administrative staff	2	flow
D	doctor	7	schedule / flow
O	operating rooms	4	schedule
M	MRI	2	schedule
C	CT	2	schedule
A	anesthesia	2 / 3	schedule
L	lab	1	flow
CR	conventional radiology	1	flow
R	radiology (flow tasks)	1	flow
AF	anesthesia (flow tasks)	1	flow

For example, we see that the "make conclusion" task has prefix "D" indicating that this task may only be performed by a person having the doctor role. For all the schedule tasks the presence of the patient is also required. For the anesthesia department a somewhat special situation applies as during the simulated period a change in the calendar organization took place. Initially, only two calendars were available for scheduling patients, where later on this increased

[5] Of course it is possible to try and capture seasonal effects, human behavior, etc. in the simulation model. However, given the data available and our limited understanding of these phenomena, this is impractical.

to three. For the situation of three calendars, a better distinction is made in appointments for more ill patients which need a longer appointment than less ill patients for which a shorter appointment suffices. Moreover, for the new situation, blocks are reserved for walk-in patients. In this way, the time used by doctors for seeing patients could be used more efficiently.

Resource Calendars

All the resources mentioned in Table 1 have each their own calendar indicating the availability of the resource. As we are performing a transient analysis for a period of 9 months the contents of the calendars and the initial content of them is very important [28]. Moreover, the availability of resources is difficult to be approximated by a stochastic distribution (e.g. holidays, maintenance physical equipment). Therefore, in order to allow for delta analysis, the contents of the calendars of the resources that are performing *schedule tasks* are based on historical data from X/Care, i.e the AMC electronic calendar system. More specifically, based on the information from this system, the following data is included in the simulation model for the calendar of a resource:

- The scheduled hours for seeing patients.
- The non-availability of resources (during scheduled hours).
- During a simulation run, appointments are only made for gynecological oncology patients by our system. So, the appointments that are made for patients which do not belong to this group are considered as time that a resource is not available. Moreover, so called no-shows, appointments that are made where the patient does not show-up, are also included in the calendar as time that a resource is unavailable.

For example, on Thursday 5 July 2007, a doctor sees patients from 9 'o clock until 12 'o clock. However, from 10 'o clock until 15 minutes past 10 the doctor is not available because they take a scheduled break. An example of an appointment is that a patient is seen from 9 'o clock till 10 'o clock.

For the MRI, CT, and operating rooms it is important to note that the calendar shows the availability of the room instead of the availability of a specific resource. For the operating rooms a doctor from the gynecological oncology department needs to be present. At the AMC, every department is allocated a specific amount of time which can be used for performing surgeries and the required resources for performing this surgery are determined at a later time. For the calendars of the operating rooms a somewhat special situation applies as for these calendars we only have data about the surgeries performed (and not their actual availability). Based on these appointments and the average occupation rate (85%) for these calendars, we determined the scheduled hours for performing surgeries. Note that the organization of the calendars of the doctors and the operating rooms are organized in such a way that when a doctor is performing a surgery, that doctor is not available for seeing patients at the outpatient clinic and the other way around.

For the resources performing *flow tasks*, the scheduled office hours are based on interviews with hospital experts (e.g. a nurse is present on working days from

08:00 to 16:30). However, for the diagnostic tests for which a walk-in facility exists (i.e. no appointment is needed), only one calendar is defined for each of them. As for these facilities, a patient is seen almost immediately when he/she arrives, we consider that only one dedicated resource exists who sees gynecological oncology patients. Recall that for the lab, x-ray, and pre-assessment a walk-in facility exists.

Moreover, the “MRI report”, “CT report”, “pre-assessment report”, “walk in pre-assessment report”, “lab report”, and “radiology examination” tasks are all related to the reporting required after a diagnostic test took place or some data has been examined (task “radiology examination”). Therefore, as all these tasks take place at a department other than gynecological oncology, we consider that for each of them, only a single resource performs the task.

Note that doctors may both perform schedule and flow tasks. However, the calendars of these resources are organized such that they show the availability of doctors for performing schedule tasks. In order for a doctor to still perform flow tasks, it has been decided to assign an unlimited amount of time for performing them. So, a doctor can always perform a flow task, except when a schedule tasks needs to be performed. Consequently, this would mean that a doctor is always eagerly waiting for performing a flow task, which does not hold in reality. Therefore, a delay is taken into account such that assignment does not immediately take place. Note that the “forms doctor” in Figure 3 is the only flow task which is performed by a doctor. The “delay forms doctor” task is only added for routing purposes. Note that this is a well-known problem when modeling human actors, especially if they are involved in multiple processes. See [4] for more examples.

Patients

For the group of 142 patients that follow the healthcare process, an appointment at the outpatient clinic itself is created and some diagnostic tests are completed. In total, the number of appointments registered for an MRI, CT, pre-assessment, and examination under anesthetic are 45, 25, 66, and 16 respectively.

For patients, several important events need to be considered for the simulation. These are (1) the arrival process of patients; (2) the selection of tests for a certain patient (e.g. MRI, CT); (3) the duration of appointments; and (4) the rescheduling of appointments. All these events are influenced by human behavior which is difficult to approximate by a stochastic distribution. For example, the diagnostic tests that will be performed and the length of them can not be considered mutually independent. The actual scheduling of an appointment is determined by a multitude of (human) factors, including patient preferences, scheduling heuristics used by a scheduled, etc. At the same time, people tend to be flexible. Moreover, recall that some diagnostic tests are ordered before the first visit of the patient and some of them only after the first visit.

A similar remark applies for the rescheduling of appointments which also can not be considered independently of each other. For example, if for a patient a first visit is rescheduled also the MRI appointment needs to be rescheduled as they are scheduled on the same day and the MRI needs to take place after the first visit.

Remember that the contents of the calendars of the resources that are performing *schedule tasks* are based on historical data from the AMC electronic calendar system. Moreover, as already indicated earlier, our focus is on delta analysis and not the absolute performance of the process as circumstances change all the time. Therefore, the following approach is taken.

The electronic calendar system of the AMC contains data about appointment creation and rescheduling. Additionally, it contains data about the scheduled duration of the appointment. Therefore, for the "flow" of the patient through the healthcare process, subsequent events in the process for a patient will be performed in the simulation at the *same time* and in the *same way* as they happened in reality. Consequently, the distribution of the events listed below will be the same as those in reality.

- The start of the healthcare process itself. This is based on the registration date of the first consultation with the doctor.
- The duration of an appointment (e.g. the consultation with a doctor). The actual duration of the appointment that will be scheduled by the system has the same duration as the appointment that happened in reality. For example, for the first visit there are in total 142 appointments for which 4 of them took 30 minutes, 112 took 60 minutes, and 20 of them took 90 minutes.
- The rescheduling of an appointment.

The difficulty of approximating above mentioned events by a stochastic distribution can be illustrated by considering the arrival process of patients.

In the information systems of the AMC, the arrival time of patients is only registered in days. This means that the exact timestamp is not available although in reality patients arrive throughout the day. As a consequence, we investigated whether the stochastic process can be modeled with a *compound Poisson process*. Arrival processes in which customers arrive in *batches* can be captured by a compound Poisson distribution. Amongst others, as requirement to use this kind of distribution, the interarrival times of batches need to follow an exponential distribution. A histogram showing the frequency of the interarrival times of the batches, expressed in days, is shown in Figure 5. A goodness-of-fit test (Kolmogorov-Smirnov with result $P = 0.0$) shows that the interarrival times of batches can not be modeled by an exponential distribution. As a consequence, random arrivals of the batches and independence of them can not be assumed.

Note that the replaying of events as mentioned above and the fact that the contents of the calendars of resources that are performing schedule tasks is based on the content of the AMC calendar system, does not mean that the behavior in a simulation run is deterministic. The actual scheduling of appointments is dependent on the number of cases in the system, the actual completion of the tasks in the healthcare process, patient preferences, and the heuristics used by the scheduler. Additionally, the way that resources perform tasks in the process, i.e. the precise behavior of them, occurs on a non-deterministic basis. The deterministic aspects of the simulation model are discussed below.

Resource behavior

The behavior of the resources undertaking the tasks in the healthcare process

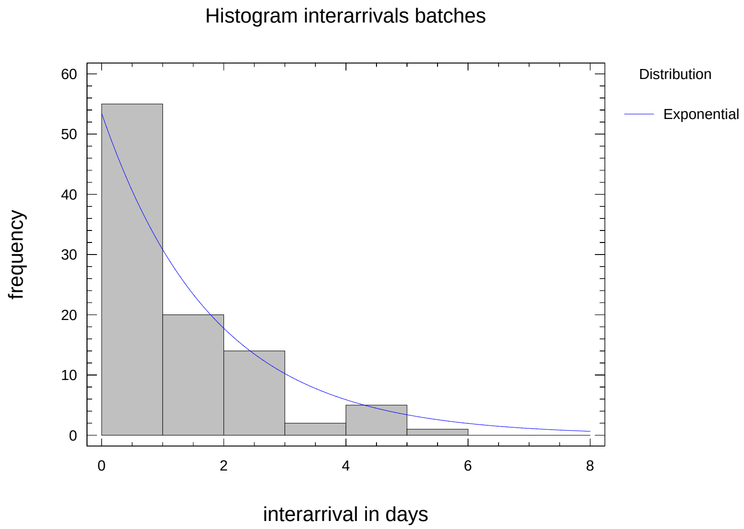


Fig. 5. Fit of the interarrival times of batches with an exponential distribution ($\lambda = 0.551$).

also deserves some attention. The way in which tasks are performed and the time spent on performing these tasks is configured in the following way:

- In order to estimate the duration of flow tasks by a stochastic distribution, interviews with healthcare specialists have taken place and initial observations have been made when observing the process. As a consequence, we decided to approximate the time spent by a resource on a flow task by a normal distribution.
  Note that the duration of schedule tasks is already determined beforehand as appointments for them are made. For simplicity reasons, we assume that schedule tasks are completed within their assigned timeslot.
- In practice, nurses working at the gynecological oncology department perform their duties on a case-by-case base. Therefore, these resources perform flow tasks belonging to the same case for as long as remaining flow tasks for that case are available. Once there are no further flow tasks for the case, flow tasks for other cases may be performed. Once further flow tasks become available for a case, they do not necessarily need to be performed by the same resource.
- For the flow tasks it is assumed that a resource is working on one task only. So, a resource may only allocate one flow task to his or herself and must complete the task. If there is not enough time left for completion of the task on the day itself, the next time that the resource is available, he or she will continue working on that task. Note that as a consequence of our assumption that schedule tasks are performed within their assigned timeslot, none of these tasks need to be partially performed at another day.

- The resources possessing the "nurse" and "administrative staff" roles do not spend all of their time on performing work for gynecological oncology patients. One possible solution, described in [27], is to assume limited availability of people. Based on initial observations with the process and interviews with healthcare specialists, we decided to model that they only spent 50% of their time working on these tasks and 50% of their time is available for working on other tasks (e.g. helping a doctor which is seeing a patient). Note that a more refined solution can be found in [4]. However, this is outside the scope of this paper.
- The assumption that a resource is eagerly waiting for work and immediately reacts to any task that is available does not hold for every task (see also [4]). Therefore, for the tasks "forms doctor", "examination under anesthetic report", "MRI report", "CT report", "lab report", and "radiology examination" a delay is taken into account so that assignment does not immediately take place. Furthermore, the execution of the "receive radiology data" and "receive gynecology data" tasks are dependent on the actual sending of data to the AMC hospital. Clearly, these two tasks are performed at departments outside the AMC. For these tasks a similar delay as mentioned before is taken into account. Note that the "delay forms doctor" task in this context is only added for routing purposes. The delays follow a normal distribution and are based on interviews with medical experts.

Scheduling algorithm

Remember that with regard to the gynecological oncology healthcare process that we are simulating we are only focusing on non-acute patients which all have equal priority. In this way, for the AMC it is vital that the patients need to be diagnosed as quickly as possible. Consequently, for the scheduling of appointments, the scheduling requirements mentioned in the "scheduling service" subsection of Section 2.2 are taken into account. However, no additional scheduling requirements (e.g. priorities between patients, maximum time between tasks in the process) are taken into account.

Therefore, for the scheduling of appointments by our system the following scheduling algorithm is used. The scheduling of appointments is done automatically, which means that there is no user involvement. Starting with the tasks in the graph for which a work-item exists, it is determined which schedule tasks need to be (re)scheduled. Once we know that tasks are able to be scheduled, they are scheduled. Moreover, these tasks are scheduled on a sequential basis in order to avoid conflicts involving shared resources.

For the actual scheduling of an appointment, a search is started for the first opportunity where one of the resources of a role can be booked for the respective work-item. If found, an appointment is booked in the calendar of them. As the patient for which the case is performed also needs to be present at the appointment, this is also taken into account. For example, for Figure 1, if a case is started, an appointment is created for task "give radiotherapy" in the calendars of "Jane", "Jo", and the patient, or "Jane", "Anne" and the patient.

3.3 Replacement of components

An important requirement for the deployment of the system in the AMC is that the operational performance of the *implemented* system needs to be validated. Therefore, the simulation experiments are performed using part of the CPN model as shown in the "System" rectangle of Figure 4. As can be seen in Figure 4, the workflow client application and calendars component are kept in the CPN model, whereas the workflow engine and scheduling service component are replaced by the concrete implementation of these components by making connections between the CPN model and the corresponding components in the actual software. In this way, *we can control the behavior of the resources in the system, and the initial contents of the calendars of the resources, while still being connected to the actual system components.* For the correct working of the scheduling service it is important that the notion of time in this component is the same as that in the simulation model. This is realized by the connection between the scheduling service and the time subnet in the CPN model which ensures that the time in the scheduling service and the CPN model are synchronized.

The workflow component is realized using the open-source WfMS YAWL [1] and a service which acts as an adaptor in-between YAWL and the workflow client application and in-between YAWL and the scheduling service. The scheduling service component is implemented in Java as a service which communicates with the WfMS via SOAP messages. The communication between the components of the CPN model and the concrete implementation of the components is realized using Comms/CPN. Comms/CPN is a library that offers the necessary infrastructure to establish communication between CPN models and external processes. In this way, interactions between CPN models and the physical environment are possible [14].

As a consequence of using the YAWL WfMS in our simulation model, the process that has to be followed by the patients, needs to be entered into the YAWL system. Therefore, as process definition, the YAWL model, modeling the gynecological oncology healthcare process as shown in Figure 3, is used.

Note that including parts of the implemented system (the engine and the scheduling service) in the simulation model has its expenses on the computational complexity of the simulation model itself. *For one replication* performed by a modern PC (Intel Core2 Duo 2.33 GHz, 4 GB of RAM), *it takes around 15 hours to complete.* However, as we want to validate the operational performance of the implemented system we consider it as vital to include parts of the developed system in the simulation model.

3.4 Validation

The validation of a simulation model is a non-trivial but important step in the simulation process. By performing a validation it is determined whether the right model has been built [7]. In [29], 15 different validation techniques and tests used in simulation model validation and verification are mentioned. Most of them are found in literature and can be used either subjectively or objectively. The most well-known are animation, face validity, and predictive validation. In our case the simulation model is validated by animation, historical data, face validity, and by means of traces. These techniques have been selected as our simulation is terminating, historical data is available, and information from medical experts could be obtained. Moreover, due to the specific set-up of the simulation model in which components in the conceptual model are replaced by their implemented counterpart, obtaining information about the behavior of the system is vital.

The gynecological healthcare process, shown in Figure 3, has been validated by *animation* as described in [23]. When animating the process it was asked to users whether something was missing or should be added.

By assessing *face validity* individuals knowledgeable about the system are asked whether the model and/or its behavior are reasonable [29]. During the process of validating the model, medical experts have been asked about the inputs and outputs of the model which led to improvements in the model. For example, initially it appeared that waiting times for the first visit where increasing over time. It appeared that the historical data involving the contents of the calendars of the gynecological oncology doctors did not include all scheduled blocks for seeing patients.

Moreover, remember that the organization of the calendars of the resources, that are performing schedule tasks, are based on historical data from the AMC electronic calendar system and that subsequent events in the process for a patient are performed in the simulation at the *same time* and in the *same way* as they happened in reality. In this way, we believe that this also leads to an increased face validity.

By using *traces* as a means to validate the model, the behavior of different types of specific entities in the model are traced (followed) through the model to determine if the models logic is correct and if the necessary accuracy is obtained [29]. Due to the fact that our simulation is terminating, for one replication several process related steps for a patient (e.g. completing a task, start of the process), have been recorded in a so-called "event log". Moreover, we also included the actual appointments that are made for a case in the log. Based on such an event log, which contains all the process steps taken for each case, a wide range of process mining techniques comes into reach. The basic idea of process mining is to learn from observed executions of a process. The ProM framework features an extensive set of analysis techniques which can be applied to real-life logs [32]. In our case, it allows for viewing and checking the correctness of several steps performed by the system.

As historical data is existing, validating the model using *historical data* is possible. The following approach has been taken. We use the performance measures as described in the "Performance measures" subsection. For the 142 patients that followed the gynecological oncology process, the second column in Table 3 shows the average waiting times that are realized in reality for the defined performance measures. Note that for the realized waiting times some outliers needed to be removed. We used Box and Whisker plots to visualize these outliers after which they could be removed. In total for the first visit, MRI, CT, anesthesia, and operating rooms, respectively 11, 1, 0, 0, and 0 values have been removed. It is remarkable that for the first visit 11 observations have been removed. All the removed appointments had a waiting time longer than 4 weeks and they where quite separated from the other observations. Based on discussions with medical specialists we considered them as not relevant as a typical treatment process takes 4 weeks, whereas for the removed values the waiting time for the first visit was already more than 4 weeks.

Remember that for the actual scheduling of an appointment, a search is started for the first opportunity that precisely one of the resources of a role can be booked for the respective appointment. However, the actual scheduling of an appointment is determined by a multitude of (human) factors, including patient preferences, scheduling heuristics used by a scheduler, etc. Our experience is that human behavior can not easily be captured and can involve many different factors. Moreover, the making of appointments can not be considered independently from each other. For example, often it is tried to schedule multiple appointments for a patient on one day.

Therefore, a simple solution to this problem, taking these factors into account, is to add a delay to the earliest time that an appointment may be booked such that the average waiting time for each appointment matches the figure realized in reality. The delay that has been added for each specific kind of appointment can be seen in Table 2 and follows a normal distribution with an average and a variance.

Table 2. For the delays, for which a normal distribution is used, the average is presented in the average column and the variance is presented in the variance column.

Appointment type	**Delay** (normal distribution)	
	average	variation
first visit	1859	2500
MRI	1742	2500
CT	7313	2500
pre-assessment	1636	2500
examination under anesthetic	8939	2500

Using these delays and the simulation model as described in Section 3.2, we ran 10 replications in which 142 patients followed the gynecological oncology process. The corresponding results are shown in Table 3. For each performance measure the second column shows the average value realized in reality, whereas the last four columns show the simulation results obtained for 10 replications. For the simulation results, respectively, the average, standard deviation (sd), and the lower bound (LB) and upper bound (UB) for the corresponding 95% confidence interval are shown. Note that for pragmatic reasons only 10 replications have been performed. Each replication requires more than 15 hours on a powerful computer.

We see that only for the MRI, the average waiting time realized in reality is within the upper and lower bound of the confidence interval. For the other appointments we see that the average waiting time realized by our simulation model is still very close. As can be seen in the table, the confidence intervals are rather small. However, in reality an appointment only takes place during scheduled hours. We expect that this fact seriously complicates the validation of our simulation model as an appointment can not be scheduled at any point in time. Therefore, given that the realized average waiting times for each of the appointments are very close to either the lower bound or upper bound of the confidence interval, except for the MRI, we consider the simulation model to be valid.

Recall that for the AMC it is of the utmost importance that the operational performance of a healthcare process, when supported by our system, needs to be at least as good as the operational performance of the process executed in reality. For the gynecological oncology healthcare process, we see that for every type of appointment a delay needs to be added to meet the average waiting time realized for each type of appointment in reality, which means that the requirement for this healthcare process is met. Therefore, we expect in general that our system does not negatively impact the operational performance of the processes it supports and consider the requirement as satisfied for all healthcare processes.

Table 3. Validation. For a 95% confidence interval, the average, standard deviation, lowerbound (LB), and upperbound (UB) values for each performance measure are presented. Each row in the table represents the average waiting time (AWT) for a specific appointment. FV, MRI, CT, PRE, and SU represent respectively the first visit, MRI, CT, pre-assessment, and examination under anesthetic. In the AWT column, an arc represents the time in between the two appointments. Note that all figures, except the standard deviation, are presented in minutes. The figures in brackets are presented in days.

Average Waiting Time (AWT)	Realization	Simulation (10 replications)			
	Average	Average	sd	LB	UB
FV	11333 (7,9)	11070 (7,7)	182,5	10941 (7,6)	11201 (7,8)
FV → MRI	7489 (5,2)	7534 (5,2)	451,5	7211 (5,0)	7857 (5,5)
FV → CT	8853 (6,1)	9064 (6,3)	173,2	8941 (6,2)	9188 (6,4)
FV → PRE	4030 (2,8)	3761 (2,6)	90,7	3696 (2,6)	3825 (2,7
FV → SU	13733 (9,5)	13069 (9,1)	169,3	12948 (9,0)	13190 (9,2)

3.5 Experiments

As indicated above, the average waiting time for diagnostic tests is an important measure in the context of the healthcare process. By performing a number of simulation experiments we want to obtain some quantitative insights with respect to these measures. In the "Experiment 1" and "Experiment 2" subsections different experiments will be considered.

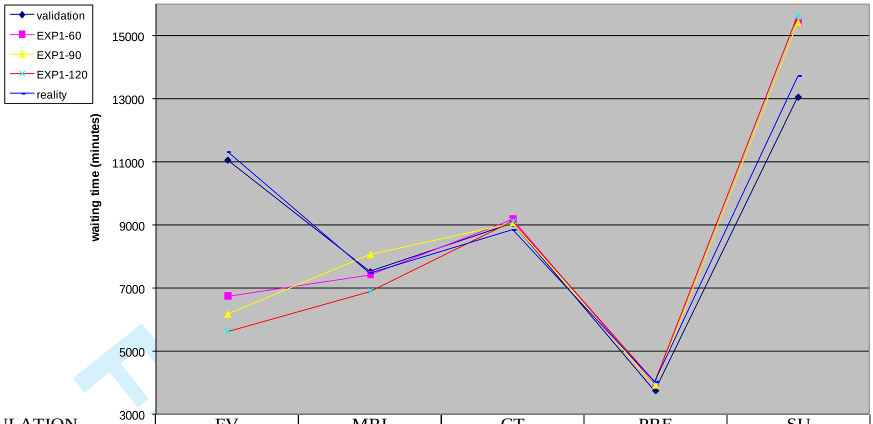
Experiment 1 Recall that as service level for the group of patients we are studying, it has been defined that for 90% of them, (1) the first visit should take place within seven calendar days after registration of the patient, and (2) all diagnostic tests should be completed within 14 calendar days after the first visit. However, for the first visit, the service level is not met as in reality only 47% of the patients have an appointment within 7 calendar days. For the first visit the average waiting time is 11333 minutes (7,9 days). Note that for the other appointments, the required service level is met. For the simulated system we have that only 51% of the patients have an appointment within 7 calendar days and that the average waiting time is 11070 minutes (7,7 days).

To examine how this situation might be remedied, it has been decided to add capacity for seeing new patients to already existing calendars, i.e., we assume that the AMC is not hiring new doctors. Instead, we investigate for the current medical staff available how much capacity needs to be added so that the service level is met.

In total, 142 appointments for a first visit are registered of which 4 of them took 30 minutes, 113 took 60 minutes, 5 took 75 minutes, and 20 took 90 minutes. So, the majority of appointments either take 60 or 90 minutes. Note that the figures are the same for the system in reality and the simulated system. Therefore, the following three variations were examined: for a selected resource, every week, at the same day, an additional 60, 90 and 120 minutes have been added for seeing *new* patients respectively (and not for patients which need a regular check-up or a follow-up meeting).

For every experiment, 10 replications of the simulation model have been carried out to be able to provide standard deviations. Note that each replication requires more than 15 hours on a fast powerful computer. The obtained results for these three experiments can be seen in Figure 6 which focusses on the average waiting time for each kind of appointment. Figure 6 is split-up in three parts. First, for each experiment, the average for each performance measure is visualized in the graph. Second, the table with name "SIMULATION RESULTS" shows first for every performance measure the average waiting time (avg) that has been realized in reality. The next rows show for each performance measure the average (avg) and standard deviation (sd) that are obtained for the validated simulation model ("validation"), and the figures obtained for adding a capacity of 60 ("EXP1-60"), 90 ("EXP1-90"), and 120 ("EXP1-120") minutes for seeing new patients respectively.

The table with name "T-TEST" shows the results of t-tests in order to determine whether the observed average for a certain performance measure is statistically significant from zero between two experiments. Respectively, the outcome of t-tests for the validated simulation model and the experiment in which 60 minutes are added ("validation $\leftrightarrow$ EXP1-60"); the experiment in which 60 min-



SIMULATION RESULTS	FV		MRI		CT		PRE		SU	
	avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
reality	11333	-	7489	-	8854	-	4030	-	13734	-
validation	11071	183	7534	452	9064	173	3761	91	13069	169
EXP1-60	6772	81	7447	284	9204	124	3968	109	15420	190
EXP1-90	6188	28	8073	468	9092	108	3938	82	15424	221
EXP1-120	5640	38	6907	214	9131	77	4049	121	15649	258
T-TEST	t-value	P-value	t-value	P-value	t-value	P-value	t-value	P-value	t-value	P-value
validation <-> EXP1-60	68,09	0,00	0,51	0,61	-2,08	0,05	-4,62	0,00	-29,23	0,00
EXP1-60 <-> EXP1-90	21,57	0,00	-3,61	0,00	2,16	0,04	0,68	0,50	-0,04	0,97
EXP1-90 <-> EXP1-120	36,68	0,00	7,09	0,00	-0,92	0,37	-2,31	0,04	-2,09	0,05

Fig. 6. Results for the experiments in which for a selected resource, every week, at the same day, an additional 60, 90 and 120 minutes have been added for seeing new patients. For every average waiting time (AWT) performance measure the average (avg) in minutes (avg) and the standard deviation (sd) are shown in the simulation results table part. Additionally, in the "t-test" table part, the result of t-tests are shown to determine whether the observed average waiting time of two experiments is statistically significant from zero. For each experiment, the average for each performance measure is visualized in the graph.

utes are added and the experiment in which 90 minutes are added ("EXP1-60 $\leftrightarrow$ EXP1-90"); and the experiment in which 90 minutes are added and the experiment in which 120 minutes are added ("EXP1-90 $\leftrightarrow$ EXP1-120") are shown.

More specifically, the t-test has been constructed to determine whether the difference between the average of two experiments equals 0,0 (null hypothesis) versus the alternative hypothesis that the difference does not equal 0,0 at the 95% confidence level ($\alpha = 0,05$). It is assumed that the observed values of each experiment come from a normal distribution and that the variances of them are not equal. In case $P < 0,05$ (column "P-value"), this implies that the null hypothesis is rejected. For example, the difference of the average waiting time of the first visit obtained for the experiment in which 60 minutes are added ("EXP1-60") and the experiment in which 90 minutes are added ("EXP1-90") is statistically significant as $P = 0,00$.

For the first visit, we see that adding an additional capacity of 60 minutes significantly lowers the average waiting time. Adding an additional 90 and 120 minutes significantly lowers the average waiting time again, but does not have such a dramatic impact as when adding the first 60 minutes. When adding an additional 60, 90, and 120 minutes for seeing new patients, the percentage of patients that have an appointment within 7 calendar days is 82, 88, and 93 respectively. For the MRI, adding an additional capacity of 60 minutes does not significantly change the average waiting time. However, when adding an additional capacity of 90 minutes, the average waiting time significantly increases, whereas adding an additional capacity of 120 significantly lowers the average waiting time. For the CT, adding an additional capacity of 60 minutes significantly increases the average waiting time. However, when adding an additional capacity of 90 minutes, the waiting time significantly decreases, whereas adding an additional 120 minutes does not significantly change the average waiting time. Both for the pre-assessment and the surgery, the average waiting time significantly increases when adding an additional capacity of 60 minutes whereas the average waiting time not significantly changes when adding an additional capacity of 90 minutes. The average waiting time for them significantly increases again when adding an additional capacity of 120 minutes. Note that for the surgery, when adding an additional capacity of 60 minutes, this has a more dramatic impact on the average waiting time compared to the pre-assessment.

In general, for the appointments other than the first visit, we see on average that there is far less impact on the average waiting time for them compared to the waiting time for the first visit. This can be easily explained as for these appointments, no additional capacity has been added for seeing patients. However, for the examination under anesthetic appointments, there is a remarkable increase in average waiting time for them. This shows that the making of examination under anesthetic appointments is the next bottleneck in the process.

As answer for the question of how much capacity needs to be added for seeing patients such that for 90% of the patients the first visit takes place within 7 calendar days, we have that for a selected resource, every week, at the same day, an additional 120 minutes needs to be added.

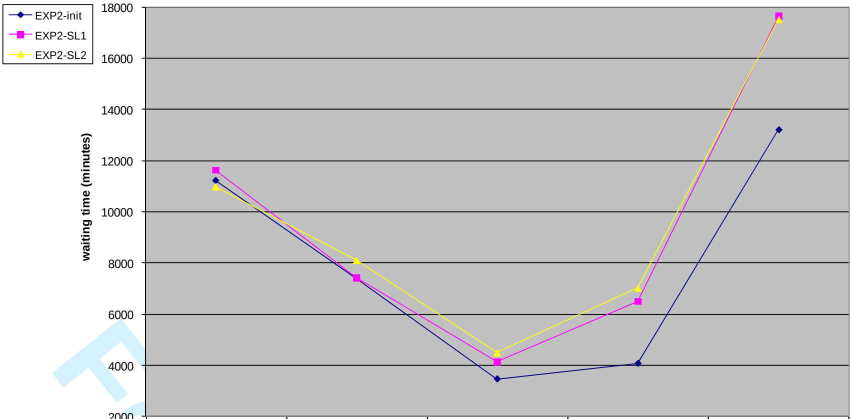
Experiment 2 Another service level that is important for the AMC is that appointments for diagnostic tests are scheduled on the same day (not when rescheduling) with a 1-4 hour gap between them. As candidates for this service level we consider the appointments for the MRI, CT, and pre-assessment. All these appointments occur after the first visit and they may be scheduled in any order. The examination under anesthetic is not a candidate as it requires a pre-assessment as a prerequisite step for which either an appointment needs to be made or a walk-in facility exists.

In order to be able to fully examine the impact of this rule, we consider the following three variations: (1) an appointment for the CT, MRI, and pre-assessment is scheduled for the very first opportunity that all required resources are available ("EXP2-init") (not necessarily on the same day); (2) the appointments for the CT, MRI, and pre-assessment are scheduled on the same day when they need to be scheduled ("EXP2-SL1"); and (3) similar as (2) but with at least one hour in between and at most four hours between the appointments for the CT, MRI, and pre-assessment ("EXP2-SL2"). Recall that the appointments for the MRI, CT, and pre-assessment may either be scheduled at the very beginning of the process or during the first visit. Appointments for the CT, MRI, or pre-assessment are only scheduled on the same day if this is either requested at the beginning of the process or during the first visit. For example, if at the beginning of the process an MRI and pre-assessment are requested and during the first visit also a CT is requested, then only the MRI and pre-assessment are scheduled on the same day. Note that as a consequence of scheduling appointments together, for each replication of the simulation model in total 21 times an MRI and a pre-assessment are scheduled together, 8 times a CT and a pre-assessment are scheduled together, and 2 times a CT, MRI, and pre-assessment are scheduled together.

In order to understand the true impact of the two service levels, we decided to configure the simulation model such that the rescheduling of an appointment is avoided as much as possible. Therefore, we removed from the model that an appointment is rescheduled on request of the hospital or the patient. In addition, a small delay of 1636 minutes is added (the smallest of the delays that are added for the MRI, CT, and pre-assessment appointments as can be seen in Table 2) to the earliest time that an MRI, CT or pre-assessment appointment may be scheduled. Note that for the first visit and the examination under anesthetic the same delay is added as defined in Table 3.

For every experiment, 10 replications of the simulation model have been carried out to be able to provide standard deviations. The obtained results for these three experiments can be seen in Figure 7 which focusses on the average waiting time for each kind of appointment. Figure 7 is split-up in three parts. First, for each experiment, the average for each performance measure is visualized in the graph. Second the table with name "SIMULATION RESULTS" shows for every performance measure the average (avg) and standard deviation (sd) that are obtained for the three experiments. In a similar fashion as for the first experiment, the table with name "T-TEST" shows the results of t-tests in order to determine whether the observed average of two experiments is statistically significant from zero for a certain performance measure.

For the first visit, we see that for the average waiting time there is no statistically significant difference between the first and second experiment. However, between the second and third experiment, there is a statistically significant difference. For the MRI, there is no significant increase in waiting time for the second experiment compared to the first experiment. However, the average waiting time in the third experiment is significantly higher compared to the second experiment. For the CT and pre-assessment, the average waiting time in the second experiment compared to the first experiment and the average waiting time in the third experiment compared to the second experiment are both significantly higher. Note that for the pre-assessment the increase in average waiting time from the first to the second experiment is higher compared to the increase for the CT. Finally, for the surgery, the average waiting time in the second experi-



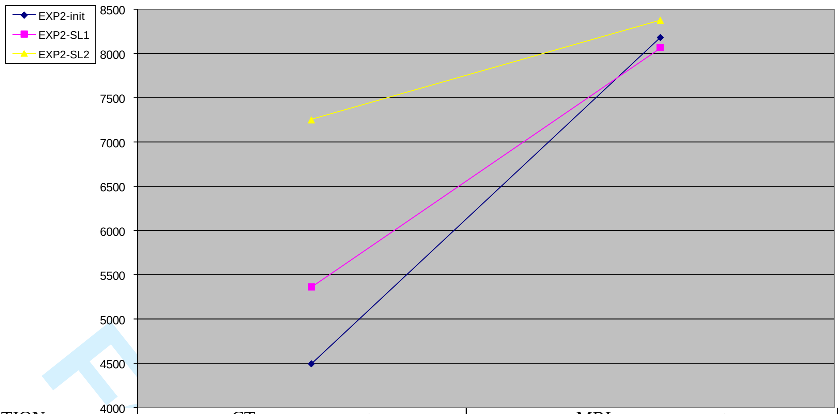
SIMULATION RESULTS	FV		MRI		CT		PRE		SU	
	avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
EXP2-init	11218	506	7404	542	3459	319	4068	114	13198	278
EXP2-SL1	11620	743	7432	429	4132	398	6497	335	17635	991
EXP2-SL2	10980	24	8104	596	4490	363	7016	466	17510	1447
T-TEST	t-value	P-value	t-value	P-value	t-value	P-value	t-value	P-value	t-value	P-value
EXP2-init <-> EXP2-SL1	-1,42	0,18	-0,13	0,90	-4,17	0,00	-21,72	0,00	-13,62	0,00
EXP2-SL1 <-> EXP2-SL2	2,72	0,02	-2,89	0,01	-2,11	0,04	-2,86	0,01	0,22	0,82

Fig. 7. Results for the experiments in which the appointments for MRI, CT, and pre-assessment are scheduled for the first opportunity that all resources are available ("EXP2-init"), the appointments for CT, MRI, and pre-assessment are scheduled on the same day ("EXP2-SL1"), the appointments for CT, MRI, and pre-assessment are scheduled on the same day but with 1 hour in-between and at most 4 hours between them ("EXP2-SL2"). For every average waiting time (AWT) performance measure the average waiting time (avg) and the standard deviation (sd) are shown in the "simulation results" table part. Additionally, in the "t-test" table part, the result of t-tests are shown to determine whether the observed average waiting time of two experiments is statistically significant from zero. For each experiment, the average for each performance measure is visualized in the graph.

ment compared to the first experiment is significantly higher. However, for the average waiting time in the second and third experiment there is no significant difference.

In general, we can see that as a consequence of scheduling the appointments for MRI, CT, and pre-assessment together, the average waiting time for the pre-assessment and examination under anesthetic considerably increases whereas the average waiting time for the MRI and CT slightly increases. In order to investigate these differences, we had a look at the combinations where the appointments for an MRI, CT, and pre-assessment are scheduled together. Remember that in total 21 times a MRI and a pre-assessment are scheduled together, 8 times a CT and a pre-assessment are scheduled together, and 2 times a CT, MRI, and pre-assessment are scheduled together.

Therefore, in the "SIMULATION RESULTS" table part of Figure 8, for the three different experiments, we show for the MRI-pre-assessment and the CT-



SIMULATION RESULTS	CT-pre assessment		MRI-pre assessment	
	avg	sd	avg	sd
EXP2-init	4492	586	8182	1233
EXP2-SL1	5362	790	8067	1085
EXP2-SL2	7252	862	8375	1086
T-TEST	t-value	P-value	t-value	P-value
EXP2-init <-> EXP2-SL1	-2,80	0,01	0,22	0,83
EXP2-SL1 <-> EXP2-SL2	-5,11	0,00	-0,64	0,53

Fig. 8. The average waiting time for the appointment of the MRI-pre-assessment and the CT-pre-assessment combination which is scheduled at the latest point in time. For every combination the average (avg) and the standard deviation (sd) is shown in the simulation results table part. Additionally, in the "t-test" table part, the result of t-tests are shown to determine whether the observed average of two experiments is statistically significant from zero.

pre-assessment combination, the average waiting time for the appointment of the combination that is scheduled at the latest point in time. In the "T-TEST" table part, the results of t-tests are presented, in a similar way as for the previous experiments, in order to determine whether the observed average waiting time for the combinations is statistically significant from zero for two experiments. Note that for each experiment, the average waiting time for each combination is visualized in the graph.

For the average waiting time of the MRI-pre-assessment combination we see that there is no statistically significant difference between the first and second experiment and between the second and the third experiment. However, for the average waiting time of the CT-pre-assessment combination this does not hold as for the second experiment the average waiting time of the combination is significantly higher compared to the first experiment. The same remark also holds for the average waiting time for the third experiment compared to the second experiment. Note that for the CT-pre-assessment combination, the average waiting time for the third experiment compared with the second experiment more dramatically increases than for the second experiment compared with the first experiment.

Moreover, if a pre-assessment is scheduled in combination with another appointment we found that this has quite some impact on the average waiting time for a pre-assessment. For the experiments “EXP2-init”, “EXP2-SL1”, and “EXP2-SL2” we considered the patients where a pre-assessment is scheduled at the same time with another appointment (this means that for the “EXP2-SL1” and “EXP2-SL2” experiments the appointments are indeed scheduled on the same day, whereas for the “EXP2-init” experiment this does not necessarily need to be the case). We found that the average waiting time for the pre-assessment for “EXP2-init” and “EXP2-SL1” is respectively 3686 and 7454 minutes which is significantly different (t-test with α=0,05, t=-16,41, P=0,00). For “EXP2-SL2” the average waiting time is 7963 minutes which is not significantly different compared with the average waiting time of the “EXP2-SL1” experiment.

In general, these results show that when applying the first service level, in which the appointments for CT, MRI, and pre-assessment are scheduled together, this has quite some impact on the average waiting for the CT and pre-assessment. In case in addition the second service level is applied, this has impact on the average waiting times for the MRI, CT, and pre-assessment which all significantly increase. Note that the average waiting time for the pre-assessment dramatically increases when applying the two service levels. This is related to the fact that for 21 appointments a pre-assessment is scheduled together with an MRI of which the average waiting time is higher than for a pre-assessment. Finally, the average waiting time for the examination under anesthetic also significantly increases when applying the two service levels. This can be explained by the fact that the average waiting time for the pre-assessment significantly increases when scheduling together with an MRI or CT.

4 Related Work

The use of discrete-event simulation in healthcare has been around since the seventies and there are numerous studies reporting on its successful application in order to improve efficiency and reduce costs. Several review papers have been written on the conduct of simulation studies in healthcare clinics [19] showing its wide spread use in this area including laboratory studies, emergency services, and the national health system. Good overviews of literature can be found in [19, 13, 33]. In this context, [19] reports that there are very few articles that report on using simulation to study complex multi-facility healthcare delivery systems. That is, most studies report on the analysis of individual units within multi-facility clinics or hospitals. In our simulation experiments we take the scheduling of workitems for the whole gynecological oncology workflow into account which means that the scheduling of tasks across multiple facilities (e.g MRI, CT) is taken into account.

Mainstream development of workflow technology started in the seventies, with office automation systems, such as Office talk by Ellis [11], but has become more mature in the late nineties. At the moment, several hundred WfMS exist and they have become “one of the most successful genre of systems supporting cooperative working" [10]. The commonly cited advantages of applying workflow technology are faster and more efficient process execution [27, 21, 15]. However, the uninformed introduction of workflow technology can have unforeseen impacts on the execution of the processes within an organization. Simulation of the workflow allows for the identification of several kinds of problems such as the existence of bottlenecks [8]. Most of the workflow simulation studies focus on the validation of a certain process by optimizing the corresponding process definition via simulation [9]. The standard approach is to convert the process definition into a formal model, and then simulation is applied using this converted model. Examples of these can be found in [6, 8, 16, 22]. As we have discussed in this paper, we have used components of the WfMS that we developed in performing simulation experiments, together with parts of the conceptual model. To the best of our knowledge, we are not aware of another approach using both the conceptual model and parts of the implemented system to perform simulation experiments. An approach that is related to our work is described in [9] in which a simulation module can be embedded in an existing Business Process Management System. Clearly, some components of the system are included in the simulation model.

In [25] it is indicated that there are successful implementations of workflow systems in healthcare but "widespread" adoption and dissemination is the exception rather than the rule. One of the problems that has to be dealt with in order to support healthcare processes by WfMS is that flexibility needs to be provided by the system [24, 30]. Unfortunately, current workflow systems fall short in this area, a fact which is recognized in the literature [3, 5, 12, 20]. To our best knowledge, [26] is the only work reporting on the application of simulation as a preliminary step for the subsequent implementation of a workflow (careflow) management system in the healthcare domain. However, the scheduling of appointments is not considered and their focus is on steady-state analysis instead of on transient analysis.

5 Discussion

In this paper, we have presented the design and implementation of a WfMS augmented with calendar-based scheduling facilities. For the implemented system, our aim was to validate its operational performance and to investigate its impacts on operational performance for different configurations of the system. This has been done by taking the gynecological oncology healthcare process and investigate the impacts of our system on several performance indicators specified for the process. The gynecological oncology healthcare process is a non-trivial and complex healthcare process which involves the scheduling of appointments across multiple facilities.

Simulation model
As a result of our aim of validating the operational performance of the implemented system, parts of the implemented system have been included in the sim-

ulation model (the workflow engine and the scheduling service). Consequently, one run of the simulation model takes already more than 15 hours on a modern PC showing that this approach is computationally expensive. For the validation of the simulation model and for each of the different experiments, a total of 10 replication runs has been performed which is rather limited. We would have preferred to use more replications to narrow down the confidence interval further.

For the gynecological oncology healthcare process, several performance measures are defined. For the AMC it is vital to understand the impact on the defined performance measures when investigating different configurations of the system. In order to simplify delta analysis, certain environment variables are controlled instead of approximating them by a stochastic distribution. To this end, it has been decided to replay for every patient its arrival, the selection of tests, the duration of appointments, and the rescheduling of appointments. Moreover, the organization of the calendars of resources performing schedule tasks is based on historical data of the electronic calendar system of the AMC. By controlling these external factors we can have a more accurate comparison of the different scenario's. Additional reasons for controlling these environment variables is that we are investigating a time period of 9 months for which the healthcare process under study was not in a steady state. In fact, typically hospital processes are rarely in steady state.

Moreover, due to the fact that the process is not in a steady state, it is required to control certain environment variables as approximating them by a stochastic distribution would lead to significant different results. In order to confirm this statement we have performed an additional experiment to demonstrate that by not controlling the arrival process of patients, this leads to completely different results. Please remember that in the information systems of the AMC, the arrival time of patients is only registered in days. So patients arrive in batches. In this additional experiment, we had a dataset of the interarrival times of batches and a dataset for the size of the batches. These datasets were based on the arrivals of patients during the time period of 9 months that have been simulated. Based on these datasets, a simulation experiment consisting of 10 replications has been carried out in which the next arrival of the batch and the size of the batch is sampled from these datasets. Moreover, the simulation model has been configured in the same way as the simulation model that has been used for the validation experiments.

In Table 4, the average and the standard deviation for the validation experiment and the experiment in which the arrival of patients is sampled, are shown. For all the performance measures, except the waiting time for the examination under anesthetic, the average of the two experiments is significantly different (this has been determined by performing t-tests to asses whether the observed average of two experiment is statistically significant from zero). This shows that by not controlling the arrival process of patients, this leads to completely different results.

Table 4. The results for the validation experiment are shown in column "Validation". The results for the experiment in which the arrival of patients is sampled are shown in column "Sampling". For every average waiting time (AWT) performance measure the average (avg) and the standard deviation (sd) are shown. The figures in brackets are presented in days.

Average Waiting Time (AWT)	Validation		Sampling	
	Average	sd	Average	sd
FV	11070 (7,7)	182,5	2091 (1,5)	120,9
FV $\rightarrow$ MRI	7534 (5,2)	451,5	2376 (1,7)	111,65
FV $\rightarrow$ CT	9064 (6,3)	173,2	8408 (5,8)	1585,0
FV $\rightarrow$ PRE	3761 (2,6)	90,7	3982 (2,8)	294,8
FV $\rightarrow$ SU	13069 (9,1)	169,3	11824 (8,2)	447,0

A consequence of our decision that the organization of the calendars of resources performing schedule tasks is based on historical data, is that patients that did not belong to our patient group and which had an appointment for one of these tasks, the appointment already had been included to the calendar of the respective resource. In that way, the timeslot is unavailable for scheduling gynecological oncology patients. As future work, we would like to study the scheduling of appointments over multiple processes.

The scheduling of appointments is done automatically by our system. For the actual scheduling of an appointment, a search is started for the first opportunity that precisely one of the resources of a role can be booked for the respective appointment. However, the actual scheduling of an appointment is determined by a multitude of (human) factors, including patient preferences, scheduling heuristics used by a scheduler, etc. Moreover, the making of appointments can not be considered independently from each other, especially when multiple departments are involved. Therefore, a simple solution to this problem, taking these factors into account, is to add a delay to the earliest time that an appointment may be booked such that the average waiting time for each appointment matches the figure realized in reality. As future work, we would like to improve the way these appointments are scheduled such that human related factors are taken into account, for example, by using the concept of "chunks" explained in [4].

For our experiments, our main focus is on the scheduling of appointments. Therefore, in our simulation model a basic approach has been taken to capture the behavior of resources performing flow and schedule tasks. For example, several kinds of resources were only working part-time on the processes at hand or a delay is taken into account so that assignment of a workitem does not immediately take place. However, in reality capturing the way people actually work is complicated by many factors. For an elaborate discussion on the problems when capturing human behavior in business processes and possible solutions, we again refer to [4].

As performance indicators for the gynecological oncology healthcare process, we have the average waiting time for the first visit, and the average waiting times for the MRI, CT, pre-assessment, and examination under anesthetic measured from the time the first visit took place. Based on historical data, it was not possible to validate the simulation model based on a 95% confidence interval. As the average waiting times for the different types of appointments obtained by multiple replications of the simulation model were very close to the figures realized in reality, we still considered our model as valid. Complicating factors for validating the model are that the making of appointments can not be seen independently from each other and involves human factors. Additionally, appointments are only taking place during office hours. Finally, the performance measures are defined in such a way that it is assumed that resources are always available. Therefore, in the future we plan to investigate approaches to alleviate this problem.

Experiments

By performing various simulation experiments for the gynecological oncology healthcare process, we have obtained some quantitative insights into the average waiting time for the various appointments. We have seen that for a selected resource, every week, at the same day, adding an additional one hour for seeing new patients already seriously decreases the average waiting time for a first visit to the outpatient clinic. By adding an additional two hours for seeing new patients, the desired service level is met, i.e., for 90% of the patients the first visit takes place within 7 calendar days after registration of the patient.

A limitation for this experiment is that only for a selected resource every week, at the same day, additional time has been added for seeing patients. In case the management of the department of gynecological oncology wants to investigate different configurations (e.g. selecting multiple resources, or adding time at different days), the simulation model can be used to investigate the impacts.

A limitation of our experiments of scheduling appointments on one day is that the organization of the calendars of the resources involved has been unmodified. Preferably, when scheduling appointments on one day, the calendars of resources at different medical departments are tuned in to each other such that appointments taking place at different departments can easily be scheduled on one day. As future work we like to focus on a so-called one-day diagnostic trajectory for the gynecological oncology healthcare process. In other words, how do all calendars need to be organized such that all diagnostic tests that are required (MRI, CT, pre-assessment, lab, x-ray) are scheduled on one day, including a visit to the outpatient clinic.

6 Conclusion

In this paper, we described the design and implementation of a WfMS augmented with calendar-based scheduling facilities. Instead of just offering workitems via a work-list, as is the case in most existing WfMSs, they can also be offered as a concrete appointment in a calendar taking into account which preceding tasks are necessary and whether they have been performed.

In the AMC many patient centered, critical processes are running. Although the schedule-aware WfMS has been developed to support AMC's healthcare processes, it is vital for them that the new system seamlessly integrates with the running processes in the hospital and that it does not degrade the operational performance of the healthcare processes that it should support.

Therefore, we performed computer simulations in order to investigate the differences between the current situation and a new situation using our schedule-aware WfMS. In order to perform computer simulations, the conceptual model which has been used for specifying and developing the schedule-aware WfMS is also used for simulating the operational performance of the system. One of the important characteristics of this approach is the tight coupling between the conceptual model and the implemented system. In this way, parts of the system are simulated while connected to the actual system components. Moreover, the obtained simulation results hold for the implemented system and show that the system is going to work in practice. Together with the fact that for the gynecological oncology healthcare process we have shown that the correct operation of the implemented schedule-aware WfMS is ensured, we can conclude that the system can be applied safely in the AMC.

For the simulation experiments, a healthcare process involving multiple departments has been considered. The different experiments clearly show that it is difficult to capture a process in which many human factors are involved. Future work is needed in order to better capture human behavior in simulation models related to both the execution of tasks and the scheduling of appointments.

Additionally, our experiments shows how complex typical healthcare workflows are. Therefore, it is important that these kind of processes are supported by a schedule-aware WfMS such that the whole workflow is taken into account. In this way, support is offered for the timely execution of tasks and that there is enough time in between two scheduled appointments.

References

1. W.M.P. van der Aalst, L. Aldred, M. Dumas, and A.H.M. ter Hofstede. Design and Implementation of the YAWL System. In *Proceedings of CAiSE'04*, 2004.
2. W.M.P. van der Aalst and K.M. van Hee. *Workflow Management: Models, Methods, and Systems*. MIT Press, Cambridge, MA, 2002.
3. W.M.P. van der Aalst and S. Jablonski. Dealing with Workflow Change: Identification of Issues and Solutions. *International Journal of Computer Systems, Science, and Engineering*, 15(5):267–276, 2000.
4. W.M.P. van der Aalst, J. Nakatumba, A. Rozinat, and N.C. Russell. Business process simulation : how to get it right? Computing Science Reports 08/21, Eindhoven University of Technology, Eindhoven, 2008.
5. W.M.P. van der Aalst, M. Weske, and D. Grünbauer. Case Handling: A New Paradigm for Business Process Support. *Data and Knowledge Engineering*, 53(2):129–162, 2005.
6. J.S. Bae, S.C. Jeong, Y.H. Seo, Y.H. Kim, and S.H. Kang. Integration of Workflow Management and Simulation. *Computers & Industrial Engineering*, 37:203–206, 1999.

7. J. Banks, J.S. Carson, and B.L. Nelson. *Discrete-Event System Simulation*. Prentice-Hall, Upper Saddle River, NJ, 1996.
8. D.Y.K. Chan. Design an Effective Workflow in Simulation. In *The 8th International Conference on Computer Supported Cooperative Work in Design Proceedings*, pages 324–328, 2004.
9. B.K. Choi, D. Lee, and D.H. Kang. DEVS Modeling of Run-time Workflow Simulation and Its Application. In *22nd European Conference on Modeling and Simulation*, 2008.
10. P. Dourish. Process Descriptions as Organizational Accounting Devices: the dual use of workflow technologies. In C.A. Ellis and I. Zigurs, editors, *Proceedings of the ACM 2001 international conference on supporting group work*, pages 52–60, 2001.
11. C.A. Ellis. Information Control Nets: A Mathematical Model of Office Information Flow. In *Proceedings of the Conference on Simulation, Measurement and Modeling of Computer Systems*, pages 225–240, Boulder, Colorado, 1979. ACM Press.
12. C.A. Ellis, K. Keddara, and G. Rozenberg. Dynamic change within workflow systems. In N. Comstock, C. Ellis, R. Kling, J. Mylopoulos, and S. Kaplan, editors, *Proceedings of the Conference on Organizational Computing Systems*, pages 10 – 21, Milpitas, California, August 1995. ACM SIGOIS, ACM Press, New York.
13. W. England and S. Roberts. Applications of Computer Simulation in Health Care. In H.J. Highland, L.G. Hull, and N.R. Neilsen, editors, *Proceedings of the 1978 Winter Simulation Conference*, pages 665–676. Institute of Electrical and Electronics Engineers, 1978.
14. G. Gallasch and L.M. Kristensen. Comms/CPN: A Communication Infrastructure for External Communication with Design/CPN. In K. Jensen, editor, *Proceedings of the Third Workshop on the Practical Use of Coloured Petri Nets and CPN Tools (CPN 2001)*, volume 554 of *DAIMI*, pages 75–90, Aarhus, Denmark, August 2001. University of Aarhus.
15. S. Graeber. The Impact of Workflow Management Systems on the Design of Hospital Information Systems. In *AMIA 2001 symposium proceedings*, 2001.
16. A. Greasley. A Simulation of a Workflow Management System. *Work study*, 52(5):256–261, 2003.
17. M. Helfert, P. Henry, S. Leist, and G. Zellner. Healthcare Performance Indicators - Preview of frameworks and an approach for healthcare process-development. In K.S. Soliman, editor, *Information Management in Modern Enterprise: Issues and Solutions - Proceedings of the The 2005 International Business Information Management Conference*, pages 371–378, 2005.
18. K. Jensen, L.M. Kristensen, and L. Wells. Coloured Petri Nets and CPN Tools for Modelling and Validation of Concurrent Systems. *International Journal on Software Tools for Technology Transfer*, 9(3-4):213–254, 2007.
19. J.B. Jun, S.H. Jacobson, and J.R. Swisher. Application of Discrete-Event Simulation in Healthcare Clinics: A survey. *Journal of the Operational Research Society*, 50:109–123, 1999.
20. M. Klein, C. Dellarocas, and A. Bernstein, editors. *Adaptive Workflow Systems*, volume 9 of *Special issue of the journal of Computer Supported Cooperative Work*, 2000.
21. J.G. Kobielus. *Workflow Strategies*. IDG Books, Foster City, CA, USA, 1997.
22. J.J. Li, F. Casati, and M.C. Shan. Business Process Simulation with HP Process Manager. *Simulation Series*, 34(1):291–295, 2002.
23. R.S. Mans, W.M.P. van der Aalst, N.C. Russell, P.J. Bakker, A.J. Moleman, K.B. Lassen, and J.B. Jørgensen. From Requirements via Colored Workflow Nets to

an Implementation in Several Workflow Systems. *Transactions on Petri Nets and Other Models of Concurrency (ToPNoC) III*, 5800:25–49, 2009.

24. L. Maruster, W.M.P. van der Aalst, A.J.M.M. Weijters, A. van den Bosch, and W. Daelemans. Automated Discovery of Workflow Models from Hospital Data. In C. Dousson, F. Höppner, and R. Quiniou, editors, *Proceedings of the ECAI Workshop on Knowledge Discovery and Spatial Data*, pages 32–36, 2002.
25. M. Murray. Strategies for the Successful Implementation of Workflow Systems within Healthcare: A Cross Case Comparison. In R.H. Sprague, editor, *Proceedings of the 36th Hawaii International Conference on System Sciences*. IEEE Computer Society Press, 2003.
26. S. Quaglini, E. Caffi, A. Cavallini, G. Micieli, and M. Stefanelli. Simulation of a Stroke Unit Careflow. In V.L. Patel, R. Rogers, and R. Haux, editors, *Proceedings of the 10th World Congress on Medical Informatics*, volume 84/2001 of *Studies in Health Technology and Informatics*, pages 1190–1194. IOS Press, 2001.
27. H.A. Reijers and W.M.P. van der Aalst. The Effectiveness of Workflow Management Systems: Predictions and Lessons Learned. *International Journal of Information Management*, 56(5):457–471, 2005.
28. A. Rozinat, M.T. Wynn, W.M.P. van der Aalst, A.H.M. ter Hofstede, and C.J. Fidge. Workflow Simulation for Operational Decision Support using design, historic and state information. In M. Dumas, M. Reichert, and M.C. Shan, editors, *Business Process Management (6th International Conference, BPM 2008)*, volume 5240 of *Lecture Notes in Computer Science*, pages 196–211. Springer-Verlag, Berlin, 2008.
29. R.G. Sargent. Verification and Validation of Simulation Models. In M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference*, pages 130–143, 2005.
30. M. Stefanelli. Knowledge and Process Management in Health Care Organizations. *Methods Inf Med*, 43:525–535, 2004.
31. L.J. Stewart and A. Lockamy. Improving Competitiveness Through Performance Measurement Systems. *Healthcare Financial Management*, 55(12):46–50, 2001.
32. W.M.P. van der Aalst, B.F. van Dongen, C.W. Günther, R.S. Mans, A.K. Alves de Medeiros, A. Rozinat, V. Rubin, H.M.W. Verbeek, and A.J.M.M. Weijters. ProM 4.0: Comprehensive Support for Real Process Analysis. In J. Kleijn and A. Yakovlev, editors, *Application and Theory of Petri Nets and Other Models of Concurrency (ICATPN 2007)*, volume 4546 of *LNCS*, pages 484–494. Springer, Berlin, 2007.
33. Y. Yang, K. Sullivan, P. Wang, and K.D. Naidu. Applications of Computer Simulation in Medical Scheduling. In *Proceedings of the Joint Conference on Information Sciences*, 2000.