Dynamic Patient Scheduling in Hospitals

T. O. Paulussen, A. Zöller, A. Heinzl
Dept. of Inf. Systems I
University of Mannheim
wifo1@bw1.uni-mannheim.de

A. Pokahr, L. Braubach, W. Lamersdorf
Distrib. Systems and Inf. Systems
University of Hamburg
info@visis.informatik.uni-hamburg.de

Abstract: In this paper, we focus on treatment scheduling for patients in hospitals. Scheduling and coordinating patients in hospitals is faced with a high amount of complexity due to the inherent dynamics of the processes and the distributed organisational structure of hospitals. To this end, our multi-agent system MedPAge (Medical Path Agents) is presented, in which patients and hospital resources are represented as autonomous agents. For coordination, our market mechanism MedPaCo (Medical Path Coordination) is described, in which the patient agents negotiate with each other based upon individual health state dependent cost functions - over the scarce hospital resources. To incorporate stochastic processing times, a decision theoretic approach is introduced in which delays are viewed as risk by the patient agents. Further, a first approach to handle variable pathways is given. Finally, we describe a hospital simulation system which allows the benchmark of different coordination mechanisms including the current practice in hospitals.

1 Introduction

Scheduling and coordinating patients in hospitals is faced with a high amount of complexity [DL00][HJF95]. This complexity stems from the inherent dynamics of the processes and the distributed organisation structure of hospitals, as they are divided into several autonomous wards and ancillary units. For treatment, patients visit different units according to their illness [Sch90]. However, the pathway of the patients through the hospital is confronted with uncertainties. Because it is in the nature of diagnostics to gain additional information about the patients' diseases, the necessary medical treatments are often not completely determined at the beginning of the treatment process [PJDH03]. Further, the duration of the examinations and treatments are stochastic, due to the individuality of the patients. Additional problems for the patient-scheduling in hospitals arise from complications and emergencies. The immediate need of treatment for emergency patients causes disturbances in the schedule. Complications, which may occur during a treatment, result in waiting times and changed pathways for other patients. This results in variable pathways and stochastic processing times.

Therefore, the main contribution of this paper is the introduction of a novel multi-agent based distributed approach to patient scheduling under variable pathways and stochastic process durations. In the second section we present the conceptual framework of our approach. In the third section the current realisation of our inter-unit coordination mecha-
nism is described. This article closes in the forth section with conclusions and an outlook to further work.

2 Conceptual Framework

Due to the distributed structure of hospitals and the described complexity of hospital processes, we have chosen to adopt a multi-agent based approach. The autonomy of agents allows to maintain the integrity of the existing organisational structure of hospitals [JFN00][DL00]. Further, agents are able to react flexible to changes and disturbances (e.g. emergencies and complications) through pro-activeness and reactivity [PJDH03][Jen01]. Therefore, patients and hospital resources are implemented as autonomous agents with individual goals.

In contrast to the resource agents that only see the patients as entities to be treated, the patient agents only see the medical actions as tasks that need to be performed. Due to these opposing forces, the patient agents ensure that the resource agents also consider the treatments of the patients outside their unit (without any explicit knowledge of them) and vice versa [PJDH03][AG99].

For the coordination of the patients, i.e. to allocate the patients to the scarce hospital resources, we decided to use a market mechanism, in which the patient agents negotiate with each other in order to reach their individual goals [PRH01]. Within a market mechanism only prices for specific goods are communicated, keeping all other information private to the market participants [WWWWM01]. Additionally, a market facilitates a dynamic environment, where the market participants take actions according to their current (dynamically changing) situation. The price mechanism leads to an efficient resource allocation because the resources are assigned to the agents that are willing to pay the highest price (assuming that the agents bid rationally, these are the agents that gain the highest utility from this resource) [PJDH03][Wei94].

2.1 Health state dependent utility functions

To be able to evaluate their current schedule and to calculate demand and supply prices for time slots, the patient agents need worth functions [RZ94]. Because the priority of the patients is determined by their health condition, we introduce health state dependent cost functions, where the disease of a patient is viewed as disutility (decrease in quality of life) [PJDH03]. For the necessary cardinal measurement of health, we rely on the concept of years of well being [Tor87][PZ90], because it handles the health state progress over time.

Because the loss of utility adds up as long as the disease is not cured, we interpret this disutility over time as opportunity costs for not curing the disease right away [PJDH03]. These opportunity costs $C(t)$ equal the difference between the achievable health state through treatment $z$ and the patient’s health state development over time without treatment $H(t)$. The health state of a patient can either remain constant or can decrease over time.
In case of a decreasing health state we assume a linear reduction for practical reasons, i.e.
\[ H(t) = s - bt, \] where \( s \) denotes the initial health state and \( b \) the decrease rate [PJDH03].
From this we get
\[
C(t) = \int_0^t (z - H(t)) \, dt = at + \frac{b}{2}t^2; \quad a = z - s.
\]
Figure 2.1 shows an exemplary course of an illness with linear reduction of the health state, resulting in a quadratic, respectively convex opportunity cost curve.

Due to the individuality of the patients, the durations of the treatments and examinations in hospitals are stochastic. Therefore it is necessary for the agents to consider this uncertainty in the bargaining process. For this reason, the cost function \( C(t) \) has to be extended to a cost function \( \tilde{C}(\mu, \sigma) \) based upon the expected mean \( \mu \) and variance \( \sigma^2 \) of the starting time distribution \( \varphi(t, \mu, \sigma) \). To calculate the value of \( \tilde{C}(\mu, \sigma) \), the starting time distribution \( \varphi(t, \mu, \sigma) \) has to be weighted with the cost function \( C(t) \) of the patient agent. From this we get
\[
\tilde{C}(\mu, \sigma) = \int_{-\infty}^{\infty} \varphi(t, \mu, \sigma)C(t) \, dt.
\]
Based upon decision theory the variance of the envisaged starting time for a task is viewed as risk (of delay), where a linear opportunity cost curve indicates risk neutrality, because the benefit from the chance to start earlier compensates (in case of a symmetric distribution function) the disutility through the chance of a delayed start. A convex opportunity cost function on the other hand indicates risk adversity, because the possible gains from an early
start are outweighed by the possible losses due to a delayed start [Sch91]. This should be illustrated by the following example equation, using a normal distribution and our health state dependent cost function. The expected costs \( \tilde{C}(\mu, \sigma) \) for a patient agent for a timeslot with a mean starting time \( \mu \) and variance \( \sigma^2 \) can now be calculated by

\[
\tilde{C}(\mu, \sigma) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \left( at + \frac{b}{2} t^2 \right) dt = a\mu + \frac{b}{2}(\mu^2 + \sigma^2)
\]

where we see that the variance \( \sigma^2 \) depends on \( b \). With regards to decision theory, we can now interpret the health decrease rate \( b \) as a determent of the agent’s attitude to risk, i.e. for \( b = 0 \) the agent is risk neutral and for \( b > 0 \) the agent is risk adverse [Sch91]. This is illustrated in figure 2.2, where curve A shows a risk adverse and curve B a risk neutral preference or cost function of the patients.

![Figure 2.2: Stochastic treatment duration.](image)

However, if the starting time distribution is not symmetric, even risk neutral patients are sensitive to different variances. This is reasonable, because a change of variance of non-symmetric distribution functions will result in a changed expectation value.

Because the starting time distribution of a task depends on the finishing time distribution of the predecessor in that resource and the previous task of the patient, the distributions are not statistically independent. Therefore an analytic approach using series-parallel reduction [KLS75] would require central knowledge of the complete network in advance. For this reason, we use Monte Carlo simulation to determine the starting time distributions, in which the agents exchange a set of random variables – drawn from the distributions of the specific tasks – during negotiation.
2.2 Coordination Mechanism

The main goal of the MedPaCo coordination mechanism is to minimise the health state adjusted stay time of the patients, which is equivalent to an overall minimisation of suffering for the patients. The basic idea of our coordination mechanism is that the patient agents try to buy into resource time slots for the needed treatments and examinations [PJDH03].

To ensure feasible (i.e. conflict free) initial task appointments for the patients, all new treatments and examinations are scheduled on a first-come first-served (fcfs) basis [PJDH03][PRH01]. These initial appointments determine the budget (or better to say the initial opportunity costs) of the agents.

Based upon this initial schedule, the agents try to improve their schedule in order to reduce their opportunity costs. The prices they are willing to pay for a specific time slot result as the difference between the cost-value of the current allocation and the cost-value for the wanted appointment, according to their individual cost function described in the previous section.

Because in this approach the (opportunity) costs for a medical action increase over time, the patient agents must try to schedule their treatments and examinations as early as possible. In case of a resource conflict, i.e. if a demanded time slot is already occupied by another patient agent, the initial demander must try to buy the time slot from the current owner. With respect to the properties of a market mechanism, the agents act in a rational, self-interested manner. Therefore, the owners of the time slots will only release their time slots, if the price offered equals the losses invoked through rescheduling [PJDH03].

The detailed negotiation process goes as follows [PJDH03]:

1. A patient agent initiates a negotiation for rescheduling, if the pathway (additional or obsolete medical actions) or the health state of his patient has changed.

2. The initiating agent selects the task with the highest possible improvement (difference between the costs of the current owned and the best reachable time slot) and contacts the resource agent which is responsible for the execution of this task.

3. The resource agent reserves that time slot and contacts all affected patient agents, i.e. the agents who currently own this interval, and informs them about the proposal of the initiator.

4. The affected patient agents (sellers) try to reschedule to the first nonreserved time slots (see step 3) and notify the initiator about their costs due to rescheduling. To prevent cycles, reserved intervals cannot be demanded by other agents.

5. If the alternative time slots for the sellers are already occupied, they again become demanders for those time slots and accumulate the invoked costs. Here, order constraints can invoke additional rescheduling in other resources.

6. After all prices are computed and submitted to the initiator, the initiator compares his expected gains from rescheduling to the total price asked for this interval. If the
gains exceed the costs he accepts or rejects otherwise, and the negotiation for this
time slot terminates.

The former initiator continues his rescheduling activities by opening new negotiations for
the next task with the (now) highest possible improvement until he cannot improve any
task any further. Previously rejected time slots will not be considered unless these time
slots are released by their owners. For concurrency issues only one (randomly chosen)
agent can initiate a negotiation at a time.

As described earlier, the whole pathway of the patients through the hospital cannot be de-
termined a priori due to uncertainties about the diseases of the patients. With respect to
current hospital practice, the treatment process of the patients (i.e. their pathway), can be
divided into successive task-assignment intervals in which the physicians determine the
next set of treatments and examinations for their patients, based upon the current indica-
tions derived from previous diagnostics. Therefore, these assignment intervals are tightly
coupled with the ward rounds of the physicians, resulting in different lengths of the inter-
vals between the wards. For example, the interval for patients in intensive care units are
shorter (several times a day) than for patients in regular wards (once a day).

While the set of assigned tasks can be scheduled directly as described above, assumptions
about the possible future treatments have to be made. To this end, we compute - based upon
empirical data – the probability \( P_{v,i+1}(n_{v,i}+1) \) of a treatment \( v \) to be assigned in the next
interval \( i+1 \) for the \( n_{v,i}+1 \) time \( (n_{v,i} \geq 0) \), where \( n_{v,i} \) denotes the number of assignments
of this task in previous intervals. To schedule these unassigned tasks, the duration \( d_v \) of
these tasks \( v \) are weighted by their probability, i.e. \( \hat{d}_v = P_{v,i+1}(n_{v,i}+1) \times d_v \). Through
this, buffers according to the probability of the treatments and examinations are created in
the resources. If a prearranged task gets assigned in the next interval, it will be scheduled
at full length. Otherwise the task will be removed. However, new tasks can always be
added to the schedule using the fcfs-rule.

3 Current Realisation

The coordination mechanism presented in the previous section tries to overcome the tradi-
tional difficulties in hospital scheduling and needs to be tested against the existing mech-
anism and other approaches to validate its usefulness. To test the new mechanism, it is
necessary to provide - besides the coordination core itself - an environment for simulating
the approach in a hospital scenario. This allows for watching the coordination in action
and for collecting statistical data, which can be used as objective comparison criteria. In
the following an overview about the general architecture of our system ”MedPAge”, as
well as the technical infrastructure, the simulation layer and the realisation of coordination
strategies is given.
3.1 MedPAge System Architecture

In figure 3.1 the components of the MedPAge system are outlined. The system is divided into five conceptually different functionalities. The GUI consists of the user interfaces according to the different actors in the system. Most important is the simulation GUI, which exhibits the complete control infrastructure that is necessary for defining the simulation setup properties and starting the simulation run. When the simulation takes place this control center is additionally the monitoring infrastructure for observing the actual system state. Furthermore, all other actors offer interaction possibilities and therefore allow the modification of important properties at runtime.

The core of the MedPAge system is composed of various interacting agents. Backbone is the simulation control agent which is responsible for starting all system agents. This means in detail that it has the responsibility to set up the time service agent, the event generator agent as well as the resource agents and the initial patient agents. Task of the time service agent is the timely synchronization of all agents that participate in the simulation run while the event generator creates time points for important occurrences. These agents
are essential parts of the simulation environment and therefore are described accurately in the following sections. Main actors of the system are the patient and resource agents, which negotiate appointment slots following the given coordination mechanism. The initial patient agents represent patients that are already admitted in the hospital. During the simulation run, new patient agents are created according to the arrival of new patients in the hospital and they are removed from the system when patients leave the hospital. The coordination mechanism itself and further needed functionality are encapsulated in separate agent modules, which are called capabilities [BHR +00]. So called knowledge agents embody legacy hospital information system components and are designated for a future MedPAge hospital integration. With the help of the Agent.Hospital gateway agent the MedPAge system can communicate with external multi-agent systems. The mechanism allows to address certain functionalities of the MedPAge application from the outside and trigger external actions from within the system.

The data tier serves for the persistent storage of hospital information. On the one hand static information about the available hospital resources (including personnel data) is contained, on the other hand dynamic information about the agreed appointments and the executed treatments is saved. The latter information is read by the patient agents and updated by the resource agents during a simulation run. Additionally, existing legacy hospital information system components are located in this tier.

The evaluation components provide services for the analysis of the collected data from simulation runs. A statistics module offers functionality for calculating key data like mean waiting time for patients and efficiency of hospital resources. On top of the statistics module it is planned to realize a benchmarking module, which summarizes important data and allows to compare the different coordination strategies.

3.2 Technical Infrastructure

The technical infrastructure is based on two fundamental building blocks that are concerned with agent and persistency services (see figure 3.2). Basic agent services as the agent lifecycle management, agent communication and search facilities are provided by the FIPA-compliant agent platform [PC01]. These basic services are enhanced with a rational agent layer following the BDI-metaphor [RG95], which enables the usage of goal-oriented concepts at the implementation level. It therefore simplifies the development with the introduction of high-level agent-oriented programming concepts [PBL03]. The

---

1 The MedPAge project is part of the German priority research programme DFG-SPP-1083 “Intelligent Agents in Real-World Business Applications” which is made up of research projects dealing with agent based solutions to problems in the manufacturing and hospital logistics domain. The Agent.Hospital initiative has developed a scenario in which the different independent research projects cooperate based on superordinated process flows. For more information on Agent.Hospital and the DFG-SPP-1083 see http://www.realagents.org.

2 The concretely used components are (available as open source):
   - The JADE agent platform: http://sharon.cselt.it/projects/jade
   - The Jadex BDI extension for JADE: http://sourceforge.net/projects/jadex
   - The MySQL relational database: http://www.mysql.com
   - The Cayenne object-relational mapping framework: http://www.objectstyle.org
The persistency infrastructure consists of a relational database management system, which is connected with an object-relational mapping layer. This mapping layer is responsible for allowing object-oriented access to the data by making the underlying relational database model transparent.

### 3.3 Hospital Simulation Layer

Foundation of the MedPAge simulation layer is a domain-independent agent-based time synchronization component \([\text{BPL} + \text{04}]\). This FIPA-compliant service enables the agents to act synchronously with respect to a global advancing clock. The global clock describes the simulation time in the system and is advanced according to the collected time-requests of the service participants. The participants therefore control the simulation run by their communication with the time service.

When a simulation run is initiated, the information from the hospital model is used to create the hospital infrastructure consisting of initial patient and resource agents. During the run, the event generator agent uses different random distributions to approximate real arrival rates of patients and other occurrences like emergencies. It therefore decides at what time the next arrival or emergency will take place. As simulation time advances it becomes necessary to call the patients with due appointments to the resources. This is generically done with the call-patient module which is the driving force of the simulation. The resource agents are notified by the time service when a treatment start time is reached, and try to call the patient. When the patient is unavailable due to another ongoing treatment, the treatment has to be delayed until the patient is available again. If possible, the resource
will perform another treatment first. The treatments are simulated in the resource agent by using the time service to wait for the duration of the treatment. When a treatment is done, the actual treatment start and end times are stored in the database for later evaluation. The call-patient module is independent of the used coordination mechanism. This approach ensures that the agents have the same generic perception of the (simulated) environment, regardless of the coordination mechanism currently executed. Therefore different strategies can be flexibly be plugged into the agents, and compared to each other under the same conditions.

3.4 Coordination Layer

The coordination mechanisms have been designed using techniques such as AUML [AP01]. From the AUML diagrams the conversations an agent has to support were derived. For all of the conversations plans have been implemented for each participant. The strategy realisation is explained exemplarily using the MedPaCo mechanism, which relies on the FCFS strategy to create initial schedules for new reservations. In the FCFS strategy a newly created patient agent instantly makes reservations for all treatments contained in it’s clinical pathway template. For each treatment the patient agent negotiates with the corresponding resource agent using the FIPA contract net interaction protocol [Fou02] allowing both agents to come to an agreement concerning suitable time slots.

For the MedPaCo strategy the patient agent has a plan that initiates the protocol for a single optimisation round, in order to improve the current schedule (see figure 3.3, left hand side). At the resource agent, reacting on the request of a patient agent, a plan is executed that manages the optimisation round by requesting other patients to free their time slots as needed (see centre of figure). A second plan is executed at a patient agent
reacting to every subsequent request of a resource that a time slot has to be given away (right hand side). The resource determines that a new valid schedule has been created when no more reservations have to be moved. Then it checks if the new schedule is an improvement over the current schedule using the cost information supplied by the patients in the change-reservation requests. In this case inform messages are sent to all participating patient agents. All agents update their local beliefs and the resource agent also updates the MedP Age database with the new schedule. When the new schedule does not represent an improvement, failure messages are sent and the temporary schedule is discarded. In either case a new optimisation round may now be initiated by some other patient agent.

One can see that in MedPaCo (and also in the other coordination mechanisms) some plans react to the receipt of a message from another agent, while other plans pro-actively initiate
new conversations without being triggered by a message (e.g. making an appointment for a new treatment or initiating an optimisation round). These pro-active plans are triggered by the internal goals of the agent, which are created by the underlying call-patient mechanism or in response to events produced by the event generator agent.

### 3.5 Implementation

A prototype of the system has been implemented and is used to verify the applicability of the architecture and the proposed coordination mechanism. For easy access to the system functionality, a control center was created that allows for configuring, starting and
observing simulation runs. During a simulation run patients are admitted in the hospital according to the arrival rate and distribution properties of the event generator. The control center offers three different views (master data, simulation, evaluation).

Figure 3.4 shows the simulation tab, which allows to monitor the progress of a simulation run. In the lower area, the timetable with the current reservations is shown from the patients perspective. It is also possible to change the view for observing the resources perspective. In the timetable the currently planned reservations are shown as hatched blocks, while already performed treatments are painted in solid color. Therefore, one can see in the screenshot that the simulation time has advanced to 4 o’clock.

The evaluation tab (figure 3.5) visualises key data generated by the evaluation module. In the picture, the average and total waiting times of the patients are shown. In addition, other key data such as the utilisation of resources and average waiting times for the different treatments can be observed. The evaluation module is also active while the simulation is running, therefore the changes in the key data can be monitored on line.

4 Conclusions and Outlook

In this paper, we described a distributed, multi-agent based approach to patient scheduling in hospitals. This approach implements the patients as well as the hospital resources as software agents. We explained the conception and implementation of our MedPaCo coordination mechanism, in which the patient agents negotiate with each other over the scarce hospital resources. In this context, health state dependent cost functions which the patient agents use to evaluate their current schedule and to compute bid and ask prices for time slots during negotiation were developed in this paper. Stochastic processing times were handled through the introduction of a decision theoretic approach in which the patient agents view possible delays as risk. Through the notion of task-assignment intervals, we provided a first concept to handle the problem of variable pathways in hospitals.

Further, we presented the architecture and technical infrastructure of our implemented multi-agent system. We have shown how the layered architecture and the use of capabilities and the BDI agent model facilitated a flexible design, where the simulation environment was developed independently of the coordination mechanisms. The simulation environment therefore allows to test and compare the coordination mechanisms under the same conditions, which are based on empirical data that was collected from hospitals.

Future work will focus on further improvement and validation of our current coordination mechanisms using extensive trials and benchmarks. Final goal of our project is the deployment of the system into hospitals.
5 Acknowledgements

We would like to thank DFG, Deutsche Forschungsgemeinschaft (German Research Council) for funding this work under the project title Agentenbasierte Planung und Koordination funktionsübergreifender Aktivitäten in medizinischen Behandlungspfaden (MedPAge) as part of SPP 1083.

Literaturverzeichnis


