An intelligent real-time scheduler for out-patient clinics: A multi-agent system model

Jyoti R Munavalli
CAPHRI School for Public Health and Primary Care, Maastricht University, The Netherlands; BNM Institute of Technology, India

Shyam Vasudeva Rao
Forus Health, India; Maastricht University Medical Centre, The Netherlands

Aravind Srinivasan
Aravind Eye Care System, India

GG van Merode
Maastricht University Medical Centre, The Netherlands

Abstract
Scheduling of resources and patients are crucial in outpatient clinics, particularly when the patient demand is high and patient arrivals are random. Generally, outpatient clinic systems are push systems where scheduling is based on average demand prediction and is considered for long term (monthly or bimonthly). Often, planning and actual scenario vary due to uncertainty and variability in demand and this mismatch results in prolonged waiting times and under-utilization of resources. In this article, we model an outpatient clinic as a multi-agent system and propose an intelligent real-time scheduler that schedules patients and resources based on the actual status of departments. Two algorithms are implemented: one for resource scheduling that is based on predictive demand and the other is patient scheduling which performs path optimization depending on the actual status of departments. In order to match resources with stochastic demand, a coordination mechanism is developed that reschedules the resources in the outpatient clinics in real time through auction-bidding procedures. First, a simulation study of intelligent real-time scheduler is carried out followed by implementation of the same in an outpatient clinic of Aravind Eye Hospital, Madurai, India. This hospital has huge patient demand and the patient arrivals are random. The results show that the intelligent scheduler was able to reduce the waiting time and under-utilization of resources.

Corresponding author:
Jyoti R Munavalli, Maastricht University Medical Centre, Maastricht 6200 MD, The Netherlands.
Email: jyothimunavalli@gmail.com; j.munavalli@maastrichtuniversity.nl

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
real-time scheduler improved the performance measures like waiting time, cycle time, and utilization significantly compared to scheduling of resources and patients in isolation. By scheduling resources and patients, based on system status and demand, the outpatient clinic system becomes a pull system. This scheduler transforms outpatient clinics from open loop system to closed-loop system.

Keywords
multi-agent systems, push-pull systems, real-time coordination, real-time scheduling, walk-in patients

Introduction
Outpatient clinics (OPCs) are push systems where the capacity for care is delivered based on projected average demand. Care is provided whenever resources are ready, not at the moment the patient asks for it. A system where the care is delivered when the patient asks for it, based on system status, is a pull system. OPCs are open loop systems with an unstable demand that get affected by its external and internal environment. In addition, the department-centric control of operations (decentralized) and lack of resource (staff and doctors) coordination between departments result in long waiting times, cycle times, and under-utilization of resources. Therefore, we believe that OPCs should transform to a pull system from a traditional push system by converting its open-loop system characteristics to closed-loop system characteristics.

OPCs have a complex environment with lots of participants with different roles, such as departments, resources, patients, logistics, information systems, and most importantly human interactions between different groups of people like the staff, patients, and management. Each group has its own goals and limitations, which sometimes conflict. A patient’s goal is to get quality service in time and a manager’s goal is to optimize quality, cost, flow, and efficiency. OPCs perform resource scheduling and patient scheduling in isolation.

Multi-agent systems (MASs) have been extensively used in solutions for real-world problems as they are adaptive and reactive to the dynamic changes in the environment. Earlier, MAS has been applied for patient-centric, staff-centric, and organization-centric applications. MAS acts like a self-analyzing and self-correcting network because of which we model the OPC of Aravind Eye Hospital (AEH), Madurai, India as a MAS. Its working conditions are explained later. In this study, we schedule both patients and resources considering each other’s status in the OPC, not in isolation as both, resource and patient scheduling are interconnected or interdependent. When patient arrival is highly variable and uncertain, the OPC system requires handling the situation in both the perspective. Certain amount of variability is managed through patient scheduling that identifies the optimal pathway for each patient. But the patient scheduling alone cannot manage if the variability (patient demand) increases beyond certain limit/threshold. In such a situation, resources have to be rescheduled. For this, a coordination mechanism is required that reschedules resources in all the departments depending on the demand and flow balancing.

As our main contribution, we present an intelligent real-time scheduler that schedules and coordinates resources and patients in synchronization with the actual system status of the OPC. That is, the resources adapt, as per changes in the system that are caused due to variability and uncertainty in demand. In this MAS approach, the OPC is viewed as collections of more than one type of agents; patient, doctor, and manager; all with independent goals. These agents interact and negotiate towards achieving OPC-wide (global) goals. We apply Takt time management for resource (re)scheduling and a hybrid ant agent algorithm for patient scheduling. In coordination mechanism, agents play a game (a formal description of a strategic situation) of bidding to improve waiting times and utilization. We extensively evaluate our intelligent real-time scheduler
in a precisely simulated environment under various scenarios and report about its implementation in the OPC of AEH.

The remainder of this article is organized as follows: section ‘Literature review’ presents literature review. Section ‘Materials and methods’ describes materials and methods: case study on AEH, data collection, model development, experimental design, simulation, and implementation study. Section ‘Intelligent real-time scheduler’ presents an intelligent real-time scheduler wherein we define agents and develop real-time coordination mechanism to implement interactions, coordination, and negotiations. Section ‘Results’ reports the results from different scenarios created from real case studies. We discuss our findings in section ‘Discussion’ and conclusion in section ‘Conclusion’.

Literature review

There is vast literature present on resource scheduling and patient scheduling. First, resource scheduling and then patient scheduling is presented. Later, we discuss about the literature on MASs that are extensively used for scheduling.

Resource scheduling in OPCs

OPCs often schedule resources ahead of time (weekly, monthly, or bimonthly), before the arrival of patients and resources are determined at an aggregate level because of which planning and scheduling fail to incorporate variability and uncertainty in demand. The hierarchy in management limits the decision-making by frontline employees who actually deal with variability and uncertainty. The current methods of resource scheduling that are based on average demand and average service times do not fully reflect reality. The mismatch in demand and supply results in waiting time and under-utilization. OPCs often view these frustrating delays as a capacity shortage problem rather than as a poor capacity management. Munavalli et al. propose demand-driven scheduling that adapts resources to match with the demand. The resources are scheduled based on short-term (hour-wise) demand prediction to incorporate short-term variability. Yet, there exists a difference between planned and actual demand. In addition, resources are shared between different departments in order to optimize cost and this further increases workflow complexity. As a result, patients wait for resources and resources wait for patients at different locations in the OPC. The resource scheduling model in Munavalli et al. is implemented in OPC of AEH and it has reduced the waiting times (from 66.3 to 39.0 min).

Patient scheduling in OPCs

On the other hand, patient scheduling is based on the availability of resources (already planned and scheduled), mostly by appointment scheduling systems. Patients have to visit multiple departments for various tests, diagnosis, or treatment. Patients are scheduled at registration or at other departments. Patients are unaware of waiting times at following departments in their pathways; as a result, they wait at different departments. A study schedules all the patients on their arrival, to the pathway with minimum waiting time. The scheduling model was implemented in OPC of AEH and reduced the waiting time (66.3 ± 18.7 min and 44.2 ± 11.6 min). As seen in the literature, scheduling of patients and resources in OPCs have been performed in isolation.

Despite the potential importance of coordination and integration of these issues, we find surprisingly few studies on this (van der Ham, et al., 2019). A study conducted by White et al. shows that integration of patient scheduling, capacity scheduling, and patient flow improves the patient’s
experience and the clinic’s operational performance. This study describes capacity allocation, various appointment scheduling policies, and different patient flow configurations but does not optimize its performance in real time.

**MASs**

MASs have been extensively applied to the complex healthcare environment. MAS has been used in modelling emergency departments, operation theatres, and inpatient hospitals, also used for medical planning and diagnosis where multiple strategies are analyzed. Paulussen et al. describe patients and resources as agents and show that scheduling and coordinating patients increase efficiency. Patient agents (PAs) compete for timeslots of scarce resources. Keith Decker developed a coordination mechanism that modelled the hospital as a MAS, with different rule-based agents bidding for resources or timeslots. Generalized partial global planning (GPGP) is a scheduling coordination approach that provides a planner or plans retriever to create task structures that aim to achieve agent goals and a scheduler that attempts to maximize utility. A GPGP approach along with MAS modelling reduced patient’s stay (cycle time) and increases throughput in a spatially distributed hospital. Deshpande et al. extended the GPGP approach by providing coordination mechanism for resource sharing across hospitals with multiple objectives like quality, cost, and duration to optimize and this reduced the complexity. Štiglic and Kokol modelled hospitals as a MAS to monitor and forecast patient demand for a week, based on which the resource scheduling was adapted for that week. A scheduling agent evaluated the available timeslots and waiting times before fixing the appointments for patients to reduce the average waiting times.

Vermeulen et al. proposed multi-agent Pareto appointment exchanging for patient scheduling. The patients’ schedule is improved by virtual agents, assigned to individual patients, which negotiate and exchange appointments to reduce the waiting time. Zöller et al. describes patient scheduling that allows patients to bid for the earliest treatment based on the resources that are auctioned. Agents representing patients compete for treatment appointments in the fictitious market place. The resource agents (RAs) auction off timeslots, and if a resource is free, its next timeslot is assigned to the PA with the highest bid. Each PA determines the benefit of a treatment as the price; it is willing to pay for it. The utility of a resource is defined according to how much it improves a patient’s health. This approach reduced waiting times and improved resource occupancy in the hospital, but it lacks to incorporate scheduling of walk-in patients.

MASs have been used in patient scheduling where patients as agents compete for treatment appointments. In order to achieve this purpose, the RAs auction off timeslots (for appointments) corresponding to their capacity. If the resource is free, its next time slot is assigned to the PA with the highest bid. Utilities are defined to improve the patient’s health by providing earlier appointments. Agents are also used in negotiation over scarce resources. Also, variable pathways in an OPC can be assessed and handled efficiently. Efficiency is increased by rescheduling the pathways depending on information of the system. Štiglic and Kokol have scheduled patients and nurses by adaptive scheduling but in long term, not in short-term. Nurse scheduling is performed on the basis of workload and patient predictions on a weekly basis by using an MAS. Resource sharing and interaction between them provide better and timely care to patients. Interactions between the people in OPCs are represented by a MAS where agents interact and cooperate to collectively solve problems. Not only in healthcare, but MAS approach has been widely used in holonic manufacturing and management. Ant-agent algorithms have been used for job shop scheduling. The resources are assumed to be fixed in the layout. When scheduling orders, resources are also scheduled.
Hospitals are open-loop systems that are decentralized in control with unstable demand. Schumann et al. have explained centralized and decentralized systems. And centralizing it would reduce the horizontal freedom of decision-making. It would reduce staff involvement in decision-making and its optimization would become objective rather than subjective. It is observed from the literature and also practice that OPCs lack coordination in scheduling resources, that too in real time because patient and resource scheduling are performed in isolation. Resources are scheduled for over a period of time (months–days–hours) in advance, but they are not rescheduled in real time. This study proposes a coordination mechanism with system’s approach for scheduling resources and patients, in real time. This way of coordination synchronizes patient and resource scheduling and allows active participation of agents (resources). Usually, in simulation world, the humans/staffs are depicted as agents and the recommendations from the study are implemented. But, in this study, humans are active agents which are goal oriented and they try to achieve their goals. To move the OPC from the push system to pull system, we use decentralized system of control in operations. Now, the question is, with this approach are we better able to control the processes?

Materials and methods

Outpatient clinic in AEH

This study has been conducted in AEH, Madurai, Tamil Nadu, India, which is world’s one of the largest eye care provider. It has performed 401,529 surgeries and treated 2,396,864 outpatients during 2014–2015. The hospital functions with assembly-line efficiency, adhering to strict quality norms, process standardization, and cost control, and receives high patient volume. The OPC under study has two units that are identical (have same departments and the same number of resources) and treat patients of age greater than 35 years. All patient arrivals are random (no appointment systems are used) and independent, making patient demand highly variable and uncertain. In addition, resources are scheduled much ahead of time based on average demand. The OPC has no control over input, as the OPC provides same-day care for all patients. The OPC is decentralized (department-centric control) with different types of people having different goals and decision-making approach is hierarchical (control flows vertically and upwards). It manages two types of patients: new patients and review patients. The OPC is open from 7:00 a.m. to 6:00 p.m. and follows the rule: zero at 10, which means that the patients who arrive before 9:00 a.m. to the OPC are provided care by 10:00 a.m. For this purpose, the OPC transfers resources with the same skills from various specialty clinics to the OPC, based on their availability. However, this is performed manually and only as a reaction to situations where the patient load is high and the queues are building up. Departments such as New Registration (NR) and Review Registration (RR) are common to both units, and each unit has five departments, namely, Vision (V), Refraction (RF), Tension (TN), Dilatation (DL), and Preliminary and Final Examination (PE and FE). The average processing times for the departments NR, RR, V, RF, PE, TN, DL, and FE are 2, 5, 2, 10, 10, 1, 31 (1 min for processing and 30 min for waiting to get the eyes dilated), 5 (all in minutes), respectively. The OPC has in-house two software that are Integrated Hospital Management System (IHMS) and Clinical Management System (CMS) that record process and patient information. For clarity, only Unit 1 is shown in Figure 1. All the queues (1–7) are on the basis of First Come First Served (FCFS) method. The patient flow arrows show the possible pathways for new and review patients. Now, we describe the current methods of resource scheduling, patient scheduling, and coordination in the OPC, and then present its decision-making control.

Managers are the local controllers (C) who plan, schedule, and control the activities manually in the departments. Different managers are responsible for scheduling the resources (r) like
ophthalmologists and paramedical staff. The OPC schedules the ophthalmologists on the basis of their availability after academic (teaching and research) activities and surgery schedules, and this scheduling is performed once in a month. During peak hours, if (Queue $n >$ threshold workload) $\rightarrow$ (resources ‘$r$’ is added to $n$th department $D_n$, such that $r \leq R_T$, the total number of resources) OR shorten the lunch break time of already working resources to control waiting times, $w(t)$. When upstream departments work faster, the patients flood the downstream departments that are unready to handle the increased workload. Similarly, when upstream departments work slower, the downstream departments wait for patients. The lack of coordination among departments increases the unregulated waiting time in some departments and under-utilization in the same or other departments. The time taken by the manager to react to the change and take corrective measures is called the reaction time $r(t)$ (an exogenous variable) that affects the waiting time. The reaction time varies as the resources to be transferred may be busy elsewhere in AEH. Therefore, the reaction time depends on the availability of resources at the time of need and the kind of managerial measures taken.

The patient workflow in the OPC starts with registration and completes with a final examination. The scheduler in the IHMS schedules patients to the units during registration. It schedules patients alternately to Units 1 and 2 so that both units have equal load distribution. The patient moves through various departments in either of the two pathways: NR–V–PE–RF–TN–DL–FE for new patients and RR–PE–RF–TN–DL–FE for review patients. The order of RF and PE can be interchanged. Around 5 per cent of the patients exit after PE.

There are different types of resources in the OPC, as shown in Figure 2. It is managers who take the decisions regarding planning and scheduling, whereas the frontline employees who actually perform tasks face the challenges at the operational level. Mid-level ophthalmic personnel (MLOP), medical records department (MRD) staff, and ophthalmologists are the frontline employees. Each member of the MLOP group has similar roles and this group is shared among three departments (V,
A MLOP-manager controls this group through a MLOP head. Another MLOP group meant exclusively for refraction is controlled by MLOP-manager through the refraction MLOP head. The senior and junior ophthalmologists are shared between PE and FE and are controlled by the Ophthal-manager. The MRD staff members (NR and RR) are scheduled by the MRD-manager. All managers in the OPC report to the patient care manager. The managers have horizontal control in the department (can take decisions). The patient care manager takes measures when problems arise (reactive). The frontline employees follow the instructions and do not take scheduling decisions. The limitations on decision control and lack of coordination resulted in inefficiency in operations. OPC is decentralized and push system, and this research attempts to make it a pull system so that the planning, scheduling, and coordination happen in real time.

Data collection and analysis

The initial data collection began with interviews of hospital staff which also helped to understand the AEH workflow. More details were collected through patient and process data from IHMS and CMS from January 2012 to June 2012. The collected data included patient demand, arrival times, in-time and out-time of patients and resources, resource schedule, and load distribution in both units. The waiting times, cycle times, patient mix, reaction times, service times, and utilization were extracted from the collected data. Patients’ personal data like their name, age, address, and medical diagnostics were not collected. Therefore, an ethical approval for this study was not required. But, permission from the organization was taken for this study. A data fitting tool: EasyFit (EasyFit is the best commercially available software available to help in fitting data to probability distributions. It is fast and accurate, easy to use) was used to determine the probability distributions of service time and patient arrival time. The data analysis showed that the patient arrival pattern had two peaks, at around 8:00 a.m. and 10:00 a.m. Therefore, a Bimodal Poisson distribution was selected to generate model arrival times (equation (1))

\[ P = \{v_1, v_2\} \quad \text{and} \quad \lambda = \{\lambda_1, \lambda_2, P\} \]

where \( P \) is the sum of two Poisson distributions with mean arrivals \( \lambda_1 \) and \( \lambda_2 \), and mixed with proportions \( v_1 = 0.35 \) and \( v_2 = 0.65 \). The goodness-of-fit test for input and output distribution was conducted using the Kolmogorov–Smirnov test. Emergency cases are rare and the accident cases were excluded from the data analysed. The managers and the information technology (IT) department in the OPC verified the workflow of the model using flowcharts and structured walks.
Model development

MASs are often simulated using agent technology like JADE but we used .NET platform as the hospital was already using it. Moreover, it was easier to extend the simulation model during implementation. The .NET framework was chosen to create communicating agents. It provides unified sets of class libraries and built-in support for the multi-protocol request–response communication between agents. The message transport mechanism delivered messages to agents based on delegated method and publication mechanism. SSMS (SQL Server Management Studio) was chosen to store gathered and analyzed data over the life-cycle of the system. Patients and resources were the entities whose progress was tracked through the OPC. The model was developed on the predefined operation logics such as patient type, pathways, departments, resources, service times, arrival times (in-time and out-time), and reaction time. Randomly generated service times were uniformly distributed between the minimum and maximum processing times from empirical data of each department. The managers and the IT department of the OPC verified the model. Furthermore, the model was calibrated by assigning the reaction time randomly between 20 and 30 min to improve the accuracy. The simulation model was run with the empirical data and the performance measures, namely, waiting times, cycle times, utilization, and load distribution, were collected. The results of the simulation model were compared with the empirical data of the OPC for validation as shown in Table 1 and there was no statistical difference between the two.

Experimental design

Patient demand and their arrival time are important and have impact on performance measures. In this study, patient and resources were scheduled depending on their arrival time and demand, respectively. AEH has stochastic patient demand with an average of 1800 patients/day with 30.8 per cent of the monthly patient demand being 1000–1600 patients/day (low demand), 49.9 per cent of it being 1600–2000 patients/day (medium demand), and 19.3 per cent of it being >2000 patients/day (high demand). We tested two sets of scheduling rules, namely the existing model and the intelligent real-time scheduler. The reaction time becomes extremely important while scheduling

| Patient demand | Performance measures | Existing AEH | Simulation model | p value |
|----------------|-----------------------|--------------|-----------------|--------|
|                |                       | Mean | SD   | Mean | SD   |       |
| Low            | Waiting times (min)   | 48.6 | 12.5 | 47.2 | 11.1 | 0.5   |
| Medium         |                       | 68.2 | 18.6 | 69.9 | 17.2 | 0.4   |
| High           |                       | 82.1 | 25   | 82.2 | 22.5 | 0.4   |
| Low            | Cycle times (min)     | 98.9 | 14.3 | 96.9 | 13.2 | 0.5   |
| Medium         |                       | 122.3| 17.8 | 120.7| 18.1 | 0.5   |
| High           |                       | 138.9| 27.1 | 135.9| 24.9 | 0.4   |
| Low            | Load distribution     | 3.3  | 1.2  | 1.7  | 0.9  | 0.4   |
| Medium         | in number of patients | 4.1  | 2.5  | 2.4  | 1.2  | 0.5   |
| High           |                       | 4.3  | 2.8  | 3.9  | 1.6  | 0.5   |
| Low            | Resource utilization (%) | 76   | 3.2  | 73   | 2.1  | 0.4   |
| Medium         |                       | 78   | 2.4  | 76   | 2    | 0.6   |
| High           |                       | 82   | 2.2  | 80.9 | 3.1  | 0.5   |

OPC: outpatient clinics; AEH: Aravind Eye Hospital; SD: standard deviation.
resources in real time. Therefore, we selected the reaction time (in min) as three levels: \(r(t_1)\), that is, \(r(t_1) \leq 10\), \(11 \leq r(t_2) \leq 20\), and \(21 \leq r(t_3) \leq 30\) min. Reaction times were assigned randomly to the departments in the selected range. In addition, to analyse the effect of reaction times on departmental performance, we selected six combinations of reaction times based on service times (high and low). There were in total \(2^1 \times 3^1 \times 9^1 = 54\) experiments, and the performance measures were recorded for all the experiments in the design. A full factorial experiment was conducted to estimate the effect of selected factors on the performance parameters.

**Simulation runs**

The experimental design has been replicated 10 times with 540 runs (for confidence level of 95%) to estimate the variability associated with the phenomenon. The simulation of a day took around 10 min. Different arrival times for the same mean patient arrivals were generated. The same randomizer input was used for simulation with two different scheduling scenarios (existing and proposed) to assure that difference in the results obtained was not due to different inputs. The mean and standard deviation of the waiting times, cycle times, load distribution, and resource utilization were collected. These output results were analyzed and compared with those from the existing model. Analysis of variance (ANOVA) tests were conducted for statistical comparisons at a significance level of 0.05. In addition, ANOVA tests were performed using Minitab to determine the significance of main effects and interaction effects of proposed (real-time) scheduling and reaction time on waiting times. In addition, effects of resource scheduling and patient scheduling were analyzed independently as well as together.

**Implementation of the intelligent real-time scheduler in AEH**

The proposed intelligent real-time scheduler was implemented in the OPC during April 2014 and May 2014. The scheduling algorithms were integrated into the IHMS. Figure 3 shows the timing and interaction of data during real-time scheduling in the IHMS. After a patient arrives at the OPC, a staff member enters the patient details at the registration department and invokes the scheduler. Furthermore, the patient scheduler requests real-time data from the database. The database extracts real-time data from all departments and sends it to the scheduler. The patient scheduler finds the optimal path (it is the pathway that has minimum waiting time by arranging the sequence of departments to be visited) and prints it on the patient information card.

The resource scheduler was implemented in the OPC using the SQL PHP server, Android mobile, and database server as shown. The auction and bidding were managed by an Android application developed to pull data every 30 s from a database (MySQL) via a PHP web-service (communication mode) and send SMS. An android device was connected to a Wi-Fi connection, using a free SMS pack plan for cost-effectiveness. The polling collected the information of the waiting patients in the department. This enabled the auction procedure. The call for the auction was sent to the RAs via the android device. The RAs were provided with mobile phones to bid. This auction-bid communication was performed via web services. The database was updated after the selection of RA. The performance measures were collected from IHMS and CMS. ANOVA tests for simulation and implementation results were conducted for statistical comparisons at a significance level of \(p = 0.05\).

**Intelligent real-time scheduler**

In this section, we describe scheduling of resources and patients along with coordination mechanism, all in real time. The scheduling model uses four control levels, as shown in Figure 4.
Figure 3. Timing and interaction of data during real-time coordination and scheduling in IHMS.

Figure 4. Control structure of natural and artificial agents for coordination in intelligent real-time scheduler in the OPC system (as MAS).
1. A resource pool keeps a record of all resources.
2. Real-time data are collected from all departments of the two units.
3. Patients are scheduled to pathways through an integrated patient scheduler based on the departmental system status.
4. Finally, resources are rescheduled using a coordination mechanism.

All the participants in the OPC were mapped to agent roles in the MAS. Patients, resources in each department, and managers (C1, C2, and C3) are mapped as natural (human) agents (Figure 4). Departments, patient scheduler, route agents (RoAs), and resource scheduler were defined to facilitate the necessary information and control operations and were mapped as artificial agents to implement the coordination mechanism. Patients were passive agents, meaning the patients did not optimize their waiting time/cycle time nor got involved in the optimization process. Rest of the other agents were active agents. First, we explain types of agents and their roles.

**Types of agents**

- **Patient agent (PA).** These agents need care in the OPC. The database is maintained to track their movement in the departments according to clinical pathways.
- **Resource agent (RA).** Ophthalmologists, MRD staff, and paramedical staff such as MLOP staff are the RAs who perform the activities/tasks. The goal of these agents is to maximize their utilization.
- **Manager agent (MA).** Monitors the RAs in their departments and their goal is to keep tab on RA’s idleness.
- **MLOP head agent (MHA).** Follows instructions from MLOP-MA.
- **Department agent (DA).** Keeps track of all patients and resources in the department.
- **Route agent (RoA).** This agent assesses the waiting time in each department.
- **Patient scheduler agent (PSA).** This agent creates RoAs who traverse the departments to find the optimal path for PAs.
- **Resource scheduler agent (RSA).** The role of this agent is of the patient care manager with respect to scheduling and coordination in the OPC. It monitors/tracks waiting time in all departments. It also monitors the resource pool and DAs and identifies the resource requirements in each department. This agent calls for bids from resources and schedules resources to reduce waiting time in the departments. This agent instructs MAs. With this agent, OPC can utilize its system status for scheduling both patients and resources, hence transforming the OPC from push to pull system.

The agents have attributes like name, address, and identification number as shown in the class diagram (Figure 5). An agent’s class consists of agent’s name, services, and goals. The PA has database entries regarding the departments visited, in-time, and out-time. Whenever the patient enters the OPC, the class application uses an interface agent and connects all other agents with the class application. The department, patient, resource, and resource pool information are shared with the RSA. Now, we present the algorithms for patient scheduling, resource scheduling, and real-time coordination.

**Algorithm 1: Real-time patient scheduling algorithm**

This algorithm schedules patients in real time depending on the system (departments) status. A record for each PA is created when a patient enters the OPC. The database stores patient ID, age,
in-time, departments visited, and out-time. At registration, the PA is scheduled to one of the pathways after the following actions are performed. We use hybrid ant-agent algorithm that runs throughout the day:27

- The scheduler agent creates RoAs.
- RoAs gather waiting time information for all departments.
- The scheduler agent uses the collected information from RoAs and then chooses a pathway with the minimum waiting time and schedules the PA to that pathway. The optimal path is identified by using an ant algorithm.65,66 The optimal pathway is identified at the registration and then it is updated at each department the patient is routed through. At each department, after the patient is serviced, the next department to be visited is identified. If there is change in the initial pathway based on the actual system status, then initial pathway is modified.

**Algorithm 2: Resource scheduling algorithm**

This algorithm schedules resources like doctors and paramedical staff/nurses based on the actual demand variation. First, we define Takt time as the time between units of output to be synchronized with the customer demand. In the OPC context, it translates to the average time at which a patient moves out of the OPC.
Takt time synchronizes demand and supply, commonly used in production industries. For example, consider a department with service time of 5 min that has to provide care for 30 patients in an hour. In this case, Takt time = 60 min/30 patients = 2 min/patient. The Takt time of 2 min does not mean that the patients are treated for only 2 min (contradicting the service times), but that every 2 min a patient should move out of the department. If the Takt time is less than 2 min, the service in the department is faster than patient demand and the resources either wait or stay idle. If it exceeds 2 min, then the patient waits. In order to achieve a Takt of 2 min, the department needs \( r \) resources

\[
r = \frac{\text{service time}}{\text{Takt time}} = \frac{5}{2} \approx 3
\]

The details of Takt time management are provided in the literature. To maintain flow balancing between departments, in Munavalli et al., the resources are scheduled based on the short-term demand prediction using Takt time management. We utilize the Takt time management for scheduling resources. Here, the DAs collect the real-time data from their respective departments for calculation of the required number of resources. DA alerts RSA based on the threshold value for waiting times. The threshold value for each department varied as departments varied in processing times. This threshold value is determined to avoid unregulated waiting times in departments. The RSA identifies the number of resources required throughout (all the departments) the OPC based on actual/current patient demand, constraints like consultation rooms and equipment in those departments. The existing resources and the required number of resources are compared:

Let \( r_n \) be the number of resources in the \( n \)th department where \( n = 1, \ldots, N \).

\( x_n \) be the available equipment or consultation rooms in the department.

\( R_n \) is the total available resources for each department.

RSA calculates \( r_n \) by using equations (2) and (3), such that

\[
 r_n \leq x_n
\]

\[
 r_n \leq R_n
\]

\[
 r_n \text{ integer}
\]

Inequality (4) requires that allocated resources be within the available equipment or consultation rooms. Inequality (5) is satisfied when the allocation for each department is within the available resources. In addition, all allocations must be integers. Then it activates the coordination mechanism for transferring of the resources.

**Algorithm 3: Real-time coordination mechanism**

This algorithm provides a mechanism for coordinating resources and patients in real time. Once algorithm 2 identifies the new set of resources (additional resources) required to match supply with
demand, it needs to be communicated to all RAs. RSA uses auction-bidding, as it improves both 
OPC-wide (global) interest and self-interest of bidders (local). It removes the requirement of 
extensive one-to-one negotiations (time-consuming) between managers and resources. It enables 
comparison-based selection and fairly allocates resources to departments. The auction-bidding for 
resource scheduling was implemented through the \( n \)-player Bayesian game with incomplete informa-
tion. The game consists of the following:

Players \( i \in \{1, 2, \ldots, I\} \)

- A set of actions (pure strategies) for each player \( a_i \in A_i \)
- A set of types for each player \( \theta_i \)
- A payoff function (reward) for each player \( R_i, i = 1, \ldots, I \)
- A probability distribution \( p(\theta_1, \ldots, \theta_I) \) over types of players

Here, in this case, the players (also bidders) are the RAs. The auction-bidding is a game between 
RAs where they compete with each other to achieve rewards by improving their utilization. As the 
OPC has a rule of load balancing for each RA, the players have certain strategies and bid for the 
departments with slack resources.

The RSA initializes the auction by broadcasting the call for bids to all RAs in the units and 
resource pool. RA (bidder) \( i = 1, \ldots, I \) observes the call and prepares his or her bid value \( v_i \). All 
the bidders are interested in maximizing their rewards (utilization). There is no real cost associ-
ated. The bidding action shows the responsiveness of the RAs. RAs do not know about other 
RAs bidding status. A set of auction rules or mechanism design will give rise to a game between 
the RAs. Bidders’ information and value are independent (private) from each other. In this case, 
bidders submit their current utilization \( u_i \) (normalized value), time of last transfer \( t_i \), the number 
of transfers till the time \( TR_i \) (at the start of the day it is 1), and the distance between current and 
required department \( d_i \) (stored in the database). The utility function for each bidder is computed 
as follows

\[
UF_i = u_i * TR_i * d_i \tag{7}
\]

With these auction rules, RAs play the game within the strategy space (all possible strategies/ 
options). Bidders submit sealed bids \( b_1, b_2, \ldots, b_I \). The bidder with the lowest value wins the bid 
(vacancy). In addition, the RA with lower reaction time (time taken by an RA to move from one 
department to another department) is selected. The appropriate selection of the winner earns 
rewards for RSA. The goal of RSA is to maximize its rewards. Initially, the reward is set to zero, 
\( R_{old} = 0 \).

While (bidders proposals arrive at RSA) do

\{ 

- Compare all bids and select the bid with Min \( UF_i \)
\}

RSA gets rewards that are calculated on the basis of reduction in waiting time for the action of 
resource transfer \( (w(t)_{new}) \)

\[
R = \frac{w(t)_{old} - w(t)_{new}}{w(t)_{old}} \tag{8}
\]
A Foundation for Intelligent Physical Agents (FIPA) compliant ACL message is used for agent communication. A confirmation message is sent to the winning bidder (RA) and to the related MA. The selected RA then

- Receives the message
- Updates the agent data

\[ R \leftarrow R_{\text{old}} + R \]  \hspace{1cm} (9)

The winner is transferred from the current department or from the resource-pool to the required department and the database is updated. MAs monitor the transfer of resources. If the rescheduled RA does not reach the allotted department by a predefined time (in min), then the RSA again calls for bids. The time taken by the intelligent real-time scheduler to coordinate and reschedule resources is ‘response time’ of the MAS. With this mechanism, both individual goals (utilization) and global goal (waiting time) of the OPC are achieved.

## Results

We report the results of simulation followed by implementation of the proposed scheduling model. In the simulation, the mean cycle time for the existing scheduling and real-time scheduling scheme was 120 ± 19.7 min and 83.3 ± 6.7 min, respectively. The mean waiting time was 66.3 ± 18.7 min and 29.3 ± 9.3 min, respectively. The waiting time reduced by 56 per cent by using the real-time resource and patient scheduling. Statistical measures like mean and standard deviation for the waiting time, cycle time, and resource utilization for the existing versus the real-time scheme are compared in Tables 2 to 4. ANOVA tests (\( p = 0.05 \)) show that performance measures of real-time scheduling are significantly better than those of the existing scheduling scheme.

The average waiting times in all departments were reduced (Figure 6). The figure compares the waiting times in departments for various combinations of different reaction times. Reaction times

### Table 2. Comparison of waiting time (min) of the existing and the real-time scheduling.

| Patient demand | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | Existing scheduling scheme | Real-time scheduling (MAS) | \( p \) value | Existing scheduling scheme | Real-time scheduling (MAS) | \( p \) value | Existing scheduling scheme | Real-time scheduling (MAS) | \( p \) value |
|----------------|-------------------|-------------------|-------------------|---------------------------|---------------------------|-------------|---------------------------|---------------------------|-------------|---------------------------|---------------------------|-------------|
| Mean SD        | Mean SD           | Mean SD           | Mean SD           | Mean SD                   | Mean SD                   | SD          | Mean SD                   | Mean SD                   | SD          |
| Low            | 44.9 9.6          | 12.4 6.8          | 0.001 45.3        | 10.2 7.9                  | 0.001 50.9 11.2           | 30.6 9.1    | 0.003                     |
| Medium         | 57.8 14.3         | 16.9 6.3          | 0.002 66.1        | 15.9 8.9                  | 0.003 70.1 18.9           | 40.1 11.7   | 0.004                     |
| High           | 70.1 19.4         | 22.4 7.1          | 0.004 78.9        | 20.1 10.1                 | 0.004 87.1 23.6           | 53.4 15.2   | 0.006                     |

MAS: multi-agent systems; SD: standard deviation.
Table 3. Comparison of cycle time (min) of the existing and the real-time scheduling.

| Patient demand | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | \( p \) value | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | \( p \) value | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | \( p \) value |
|----------------|---------------------|---------------------|---------------------|-------------|---------------------|---------------------|---------------------|-------------|---------------------|---------------------|---------------------|-------------|
| Existing       | Real-time           | \( p \) value      | Existing            | Real-time   | \( p \) value      | Existing            | Real-time        | \( p \) value | Existing            | Real-time        | \( p \) value      |             |
| existing       | scheduling         | scheme (MAS)       | scheduling         | scheme (MAS)|                      | scheduling         | scheme (MAS)    |              | scheduling         | scheme (MAS)|                      |             |
| Mean           | Mean SD             | Mean SD             | Mean SD             | Mean SD     | Mean SD             | Mean SD             | Mean SD          | Mean SD       | Mean SD             | Mean SD          | Mean SD             |             |
| Low            | 92                  | 10.2                | 63.8                | 7.1         | 0.003               | 99.9                | 9.9                | 75.2          | 8.3         | 0.002               | 120.9            | 11.7            | 86.1 0.003 |
| Medium         | 101.9               | 15.7                | 71.4                | 7.4         | 0.003               | 118.5               | 16.2               | 83.4          | 8.9         | 0.003               | 128.4            | 19.4            | 98.3 0.005 |
| High           | 107.9               | 22.1                | 76.1                | 8.1         | 0.004               | 130                 | 22.9               | 90.1          | 10.9        | 0.005               | 140.1            | 24.1            | 105.3 0.006 |

MAS: multi-agent systems; SD: standard deviation.

Table 4. Comparison of resource utilization (%) of the existing and the real-time scheduling.

| Patient demand | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | \( p \) value | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | \( p \) value | \( r(t_1) \) (min) | \( r(t_2) \) (min) | \( r(t_3) \) (min) | \( p \) value |
|----------------|---------------------|---------------------|---------------------|-------------|---------------------|---------------------|---------------------|-------------|---------------------|---------------------|---------------------|-------------|
| Existing       | Real-time           | \( p \) value      | Existing            | Real-time   | \( p \) value      | Existing            | Real-time        | \( p \) value | Existing            | Real-time        | \( p \) value      |             |
| existing       | scheduling         | scheme (MAS)       | scheduling         | scheme (MAS)|                      | scheduling         | scheme (MAS)    |              | scheduling         | scheme (MAS)|                      |             |
| Mean           | Mean SD             | Mean SD             | Mean SD             | Mean SD     | Mean SD             | Mean SD             | Mean SD          | Mean SD       | Mean SD             | Mean SD          | Mean SD             |             |
| Low            | 75.1                | 1.6                 | 89.7                | 0.9         | 0.007               | 74.2                | 1.8                | 85.8          | 1.7         | 0.003               | 68.4            | 1.9                | 80.1 0.004 |
| Medium         | 77.1                | 1.9                 | 93.8                | 1.1         | 0.004               | 75.4                | 2.1                | 89.4          | 1.9         | 0.003               | 71.1            | 2.1                | 81.1 0.004 |
| High           | 78.2                | 2.1                 | 95.6                | 1.0         | 0.004               | 77.3                | 3.1                | 93.7          | 2.1         | 0.005               | 69.2            | 3.3                | 86.4 0.007 |

MAS: multi-agent systems; SD: standard deviation.

Figure 6. Effect of reaction times on average waiting times in all departments.
were selected on the basis of low and high service times. For example, in mix1, departments that have low \( (L) \) service times \(<5\) min) have reaction times \( r(t_1) \) and departments with high \( (H) \) service times have reaction times \( r(t_2) \). Similarly, in mix 2, departments with low \( (L) \) service times have reaction times \( r(t_1) \) and departments with high \( (H) \) service times have reaction times \( r(t_3) \) and so on.

The two-way ANOVA tests \( (p=0.05) \) showed the significance of main effects and interaction effects of real-time scheduling and reaction time: real-time scheduling \( (0.003) \), reaction time \( (0.072) \), and real-time scheduling \( \times \) reaction time \( (0.039) \). This shows that the main effect of scheduling on waiting times is significant, whereas the main effect of reaction time is not significant. However, the interaction effects of scheduling and reaction time are significant.

As reaction times between departments varied, tests of effects between-subjects were conducted. Main and interaction effects of departments, scheduling, and reaction times on performance measures were analysed (with \( p \) values for department \( \times \) scheduling \( (0.001) \), department \( \times \) reaction time \( (0.50) \), and scheduling \( \times \) reaction time \( (0.003) \)) and are compared in Figure 7. We analysed scheduling (only patient scheduling, only resource scheduling, and both together with real-time

\[\text{Figure 7. Interaction effects of resource scheduling, patient scheduling, and resource coordination.}\]
coordination mechanism) on waiting times. It was found that the main and interaction effects of scheduling are significant in all departments. Similarly, we analyzed the interaction effects of reaction time and scheduling on performance measures. It was observed that interaction effects of reaction are not significant in departments with shorter service times.

We also compare this intelligent real-time scheduler that schedules resources and patients through a coordination mechanism with the resource scheduling\textsuperscript{18} and integral patient scheduling\textsuperscript{27} that were implemented in the OPC of AEH, in isolation as shown in Figure 8.

Table 5 shows the results of before and after implementation of the real-time scheduling model in the OPC (AEH).

| Performance measures | Patient demand | Before | After | p value |
|----------------------|----------------|--------|-------|---------|
|                      | Mean           | SD     | Mean  | SD      |         |
| Waiting times (min)  | Low            | 48.6   | 12.45 | 21.2    | 7.5     | 0.03    |
|                      | Medium         | 68.2   | 18.56 | 33.4    | 9.3     | 0.02    |
|                      | High           | 82.1   | 25.02 | 41.9    | 10.3    | 0.01    |
| Cycle times (min)    | Low            | 98.9   | 14.25 | 70.9    | 9.4     | 0.02    |
|                      | Medium         | 122.3  | 17.83 | 985.9   | 14.5    | 0.01    |
|                      | High           | 138.9  | 27.12 | 98      | 10.9    | 0.01    |
| Resource utilization (%) | Low          | 76     | 3.2   | 83.5    | 2.1     | 0.03    |
|                      | Medium         | 78     | 2.4   | 85.2    | 2.4     | 0.04    |
|                      | High           | 82     | 2.2   | 93.3    | 2.6     | 0.03    |

OPC: outpatient clinics; AEH: Aravind Eye Hospital; SD: standard deviation.

Figure 8. Resource scheduling\textsuperscript{18} patient scheduling\textsuperscript{27} and intelligent real-time scheduler.

Table 5. Comparison of waiting time, cycle time, and resource utilization before and after implementation of the real-time scheduling model in the OPC (AEH).
Table 6 shows the number of resource transfers required throughout the day and the number of times the RSA could actually transfer (in percentage) the resources. Also, it was found that departments like RR, RF, and PE and FE had more number of transfer requests.

**Table 6. Successive rate of resource transfer in the OPC.**

| Patient demand | No. of resources transferred successfully (%) = No. of times resources transferred actually/No. of times the resources needed |
|----------------|-------------------------------------------------------------------------------------------------------------------|
| Low            | 97.8                                                                                                              |
| Medium         | 89.7                                                                                                              |
| High           | 78.25                                                                |

OPC: outpatient clinics.

Table 6 shows the number of resource transfers required throughout the day and the number of times the RSA could actually transfer (in percentage) the resources. Also, it was found that departments like RR, RF, and PE and FE had more number of transfer requests.

**Discussion**

The OPC system consists of disparate departments that are locally controlled and resources that are scheduled on the basis of average demand, ahead of time, and without real-time coordination. The OPC, like other hospitals (as seen in the literature), is an open loop system and is prone to variability and uncertainty. In addition, the way the operations are managed cause congestion resulting in prolonged waiting times and cycle times. In the proposed intelligent real-time scheduler, both resources and patients are scheduled simultaneously and coordination for this is carried out in real time. The scheduling has been performed in two facets, that is in real-time (actual) demand and in global perspective. Our findings confirm that intelligent real-time scheduler reduced the average waiting time by 56 per cent (simulation) and by 51.6 per cent (implementation). It also improved resource utilization by 8.3 per cent with imposed constraints during implementation. The study also demonstrates the effects of patient demand, scheduling rules, and reaction time on waiting times, cycle times, and resource utilization. It was found that global control of operations, scheduling, and coordination in real time improved performance measures, such as waiting times and cycle times, compared to the current method of scheduling in AEH.

Besides scheduling rules and control of operations, reaction time influences waiting times in the OPC. ANOVA was used to determine the main and interaction effects of real-time scheduling and reaction time. Effects of real-time scheduling are significant on performance measures like waiting times, cycle times, and utilization. It shows that improvement in reaction time does improve waiting times but it is not the sufficient criteria for reducing waiting times. Real-time scheduling along with improved reaction time has significant effect on performance measures. Furthermore, the greater reaction times in departments with longer service times affect the performance measures when compared to shorter reaction times in departments with smaller service times.

Real-time scheduling, as studied here, used resource scheduling and patient scheduling simultaneously. The two scheduling methods perform well individually, but together as MAS along with coordination, they perform even better (Figure 7). Patients were scheduled to the pathways with minimum waiting times at the registration by considering the complete pathway. The waiting times at registration departments were not optimized by patient scheduling algorithm as it considered the system data after registration. But this issue is resolved through real-time resource scheduling by rescheduling the resources in such departments on the basis of the demand. The MAS model has real-time coordination with resource pooling and proactive bidding of resources. This improved resource utilization. AEH has dedicated staff who proactively participated in resource bidding. The resources were scheduled to the departments throughout the OPC
system and the managers monitored them. The resources were matched to the varying patient demand in real time. The MAS modelling of the OPC allowed harmonizing the performance of the OPC with conflicting goals of the agents (resources, patients, and management). Coordination and information sharing between agents improved OPC-wide performance. Delays in decision-making due to management hierarchy were reduced, as decisions were taken on the basis of real-time status and information of the departments. When the RSA could not reschedule resources due to non-availability, the patient scheduling took care of reducing waiting times as it chose the pathway that had the minimum waiting time. Therefore, the situation was handled by both resource rescheduling and patient scheduling simultaneously using real-time system status and actual demand. With this, the OPC system transitions from a push system to a pull system.

As discussed in the literature of section ‘Literature review’, Munavalli et al. implemented a predictive resource scheduling based on patient demand and real-time patient scheduling in the OPC of AEH. Both were implemented separately. In this proposed scheduler, we implement resource scheduling, patient scheduling, and coordination among them in real time and simultaneously. The comparison of waiting times of all scheduler models (Figure 8) shows that intelligent real-time scheduler performs better than the other two. The significance of the proposed scheduler is patient and resource scheduling are done simultaneously, not in isolation and in real time with coordination mechanism to synchronize the scheduling. Planning often fails to incorporate short-term variabilities and uncertainties due to arrivals and service times. The intelligent real-time scheduler captures the stochasticity due to walk-in patient demand, coordinates between departments, and (re)schedules the resources and patients in real time. It acts like a self-analyzing and self-correcting network, which incorporates the slightest change in the OPC and effectively manages its patients and resources.

In MAS, the number of interactions increases with the number of active agents. Patients were passive agents; hence, an increase in the patient demand did not affect the load on the network. The resources were active agents, who communicated and negotiated to achieve their goals. The resource rescheduling model increases exponentially when there is an increase in the number of pathways, hospital departments (50 or 100), or resources.

Limitations of this study are as follows: the non-availability of resources at times might increase the waiting times and also patients are not rescheduled rather resources are rescheduled. The auction-bidding game is objective so active participation of the resource agents is a must. The implementation is the cost involved for the database servers and the main (hospital) server, and the limited accessibility (narrow area) of the network. In order to overcome the latter, data can be stored in a common place (the cloud) where all Application Programme Interfaces are clients. These communicate with each other via Service-Oriented Architecture Protocol and use Web Service Definition Language as the communication language between them. The patient and process data were updated by hospital staff in the departments.

This model can be applied to a general hospital if response time, reaction time, and patient pathways can be studied in detail. The outcome of this model in other situations is dependent on the frequency of real-time data collection based on the hospital information system. Therefore, to extend this model to other types of clinics or hospitals, the reaction time as well as the response time should be further studied. The pathways in specialized OPCs are more deterministic and less variable than one would expect in general hospitals. General hospitals have a more complex management hierarchy and more resource types than specialized OPCs.

Health-care settings like OPCs or hospitals differ in their complexity, patient groups, and processes. Therefore, requirements for scheduling and control might vary. The variability and uncertainty in OPCs are not deterministic and need optimization in a short term. However, how we use real-time scheduling system might depend not only on the algorithm but also on other factors, such
as organization of the OPC, the knowledge management system, cross-skills of the staff to facilitate transfer of staff from one department to the other, and the layout of the building (distances between departments). As future work, we intend to explore the real-time scheduling model in complex hospitals with stochastic pathways. Also, to collect real-time data, we intend to utilize Radio Frequency Identification (RFID) tags for patients to update patient and process data, and to identify the resource location.

Conclusion

This study shows how an Intelligent Real-time Scheduler using MAS improves waiting times and cycle times in OPCs. The study incorporates two facets of scheduling: real time and global perspective. Real-time scheduling is based on actual demand and coordinates in real time to reschedule resources. It schedules patients to the optimal pathway depending on the actual status of the departments in the OPC. Global perspective overcomes the issues of unregulated waiting times. Reaction time, an exogenous variable, has an impact on performance measures in open loop systems like OPCs. The study shows that MAS fits in OPC systems where decision-making is complex and also involve many people. It provides a platform for analysing coordination mechanisms in operations management and control. Real-time information about departments, patients, and resources provide insights in decision-making. This study shows that both departmental (local) and OPC-wide (global) performances improve by Intelligent Real-time Scheduler that is based on the system status of all departments in an OPC. The proposed scheduler has been implemented to an eye care OPC and can be extended to general hospitals by considering their demand, pathways, resources, organizational structures, and workflow complexities. The study shows that scheduling resources and patients in coordination and simultaneously improves the performances rather than done in isolation. The OPC system when incorporates system status in decision-making (scheduling of patients and resources) transforms from the push system to the pull system.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jyoti R Munavalli https://orcid.org/0000-0002-0811-1499

References

1. Bergeson SC and Dean JD. A systems approach to patient-centered care. *JAMA* 2006; 296(23): 2848–2851.
2. Hopp WJ and Spearman ML. *Factory physics: foundations of manufacturing management*. Irwin, CA: The McGraw-Hill, 1996, p. 668.
3. Khanna S, Boyle J, Good N, et al. Time based clustering for analyzing acute hospital patient flow. In: *Annual international conference of the IEEE engineering in medicine and biology society*, San Diego, CA, 28 August–1 September, 2012. New York: IEEE.
4. Plasters CL, Seagull FJ and Xiao Y. Coordination challenges in operating-room management: an in-depth field study. In: *AMIA Annual Symposium proceedings*, 8–12 November, Washington, DC, 2003.
5. Khanna S, Cleaver T, Sattar V, et al. Multiagent based scheduling of elective surgery, in principles and practice of multi-agent systems. Berlin: Springer, 2012, pp. 74–89.

6. Khanna S, Sattar A, Maeder A, et al. Intelligent scheduling in complex dynamic distributed environments. In: Medinfo 2007: proceedings of the 12th world congress on health (medical) informatics; building sustainable health systems, Brisbane, QLD, Australia, 20–24 August 2007, pp. 1665–1666. Amsterdam: IOS Press.

7. Vissers JMH, Bertrand JWM and de Vries G. A framework for production control in health care organizations. Prod Plan Contro 2001; 12(6): 591–604.

8. Mansdorf BD. Allocation of resources for ambulatory care: a staffing model for outpatient clinics. Public Health Rep 1975; 90(5): 393–401.

9. Yurko LC, Coffee TL, Fusilero J, et al. Management of an inpatient-outpatient clinic an eight-year review. J Burn Care Rehabil 2001; 22(3): 250–254.

10. Ren Y, Kiesler S and Fussell SR. Multiple group coordination in complex and dynamic task environments: interruptions, coping mechanisms, and technology recommendations. J Manage Inform Syst 2008; 25(1): 105–130.

11. Güney Güler M. A hierarchical goal programming model for scheduling the outpatient clinics. Expert Syst Appl 2013; 40(12): 4906–4914.

12. van Merode F. A prelude of the 2004 Antwerp Quality Conference: targets and target values – integrating quality management and costing. Accred Qual Assur 2004; 9(3): 168–171.

13. Hulshof PJH, Mes MRK, Boucherie RJ, et al. Tactical planning in Healthcare using approximate dynamic programming (Memorandum), Department of Applied Mathematics, University of Twente, Enschede, The Netherlands, 2014.

14. Harper PR. A framework for operational modelling of hospital resources. Health Care Manag Sci 2002; 5(3): 165–173.

15. Munavalli JR, Vasudeva Rao S, Srinivasan A, et al. The optimization in workflow management: ophthalmology. J Health Manage 2016; 18(1): 21–30.

16. Roupppe van der Voort MM, van Merode FG and Berden BH. Making sense of delays in outpatient specialty care: a system perspective. Health Policy 2010; 97(1): 44–52.

17. Vermeulen IB, Bohte SM, Elkhuizen SG, et al. Adaptive resource allocation for efficient patient scheduling. Artif Intell Med 2009; 46(1): 67–80.

18. Munavalli JR, Rao SV, Srinivasan A, et al. A robust predictive resource planning under demand uncertainty to improve waiting times in outpatient clinics. J Health Manag 2017; 19(4): 563–583.

19. van Merode GG, Groothuis S and Hasman A. Enterprise resource planning for hospitals. Int J Med Inform 2004; 73(6): 493–501.

20. Holleman DR Jr., Bowling RL and Gathy C. Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. J Gen Intern Med 1996; 11(4): 237–239.

21. Cayirli T and Gunes ED. Outpatient appointment scheduling in presence of seasonal walk-ins. J Oper Res Soc 2014; 65(4): 512–531.

22. Peng Y, Qu X and Shi J. A hybrid simulation and genetic algorithm approach to determine the optimal scheduling templates for open access clinics admitting walk-in patients. Comput Ind Eng 2014; 72(1): 282–296.

23. Su S and Shih CL. Managing a mixed-registration-type appointment system in outpatient clinics. Int J Med Inform 2003; 70(1): 31–40.

24. Baril C, Gascon V and Cartier S. Design and analysis of an outpatient orthopaedic clinic performance with discrete event simulation and design of experiments. Comput Ind Eng 2014; 78(0): 285–298.

25. Yang X, Han R, Guo Y, et al. Modelling and performance analysis of clinical pathways using the stochastic process algebra PEPA. BMC Bioinformatics 2012; 13(Suppl. 14): S4.

26. Schrijvers G, van Hoorn A and Huiskes N. The care pathway: concepts and theories: an introduction. Int J Integr Care 2012; 12: e192.

27. Munavalli JR, Rao SV, Srinivasan A, et al. Integral patient scheduling in outpatient clinics under demand uncertainty to minimize patient waiting times. Health Informatics J 2020; 26(1): 435–448.

28. Hossain L and Kit Guan DC. Modelling coordination in hospital emergency departments through social network analysis. Disasters 2012; 36(2): 338–364.
29. Gloede TD, Hammer A, Ommen O, et al. Is social capital as perceived by the medical director associated with coordination among hospital staff? A nationwide survey in German hospitals. *J Interprof Care* 2013; 27(2): 171–176.

30. ter Mors A, Yadati C, Witteveen C, et al. Coordination by design and the price of autonomy. *Auton Agent Multi-Agent Syst* 2010; 20: 308–341.

31. van der Ham A, Boersma H, van Raak A, et al. Identifying logistical parameters in hospitals: does literature reflect integration in hospitals? A scoping study. *Health Serv Manage Res* 2019; 32(3): 158–165.

32. White DL, Frohle CM and Klassen KJ. The effect of integrated scheduling and capacity policies on clinical efficiency. *Prod Oper Manag* 2011; 20(3): 442–455.

33. Bogg P, Beydoun G and Low G. When to use a multi-agent systems? In: Barley MW and Kasabov N (eds) *Intelligent agents and multi-agent systems*. Berlin: Springer, 2008, pp. 98–108.

34. Jemal H, Kechaou Z, Ben Ayed M, et al. A multi agent system for hospital organization. *Int J Mach Learn Comp* 2015; 5(1): 51–56.

35. Isern D and Moreno A. A systematic literature review of agents applied in healthcare. *J Med Syst* 2015; 40(2): 43.

36. Schmidt R, Geisler S and Spreckelsen C. Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. *BMC Med Inform Decis Mak* 2013; 13: 3.

37. Paulussen TO, Zöller A, Heinzl A, et al. Patient scheduling under uncertainty. In: *Proceedings of the 2004 ACM symposium on applied computing*. Nicosia, Cyprus, 14–17 March 2004. New York: ACM.

38. Nieves JC, Lindgren H and Cortés U. Agent-based reasoning in medical planning and diagnosis combining multiple strategies. *Int J Artif Intell Tool* 2014; 23(01): 1440004.

39. Paulussen TO, Zöller A, Heinzl A, et al. Dynamic patient scheduling in hospitals. In: Bichler M (ed.) *Coordination and agent technology in value networks*. Berlin: GiTO, 2004, pp. 149–174.

40. Paulussen TO, Jennings NR, Decker KS, et al. Distributed patient scheduling in hospitals. In: *Proceedings of the eighteenth international joint conference on artificial intelligence (IJCAI-03)*, Acapulco, Mexico, 9–15 August 2003. Morgan Kaufmann Publishers.

41. Decker K. Coordinating intelligent agents. In: d’Inverno M, Luck M, Fisher M, et al. (eds) *Foundations and applications of multi-agent systems*. Berlin: Springer, 2002, pp. 1–18.

42. Deshpande U, Gupta A and Basu A. A distributed hospital resource scheduling system using a multi-agent framework. *Iete Technical Review* 2001; 18: 263–275.

43. Štiglic G and Kokol P. Intelligent patient and nurse scheduling in ambulatory health care centers. In: *Engineering in medicine and biology 27th annual conference*, Shanghai, China, 17–18 January 2005.

44. Vermeulen I, Bohte S, Somefun K, et al. Multi-agent Pareto appointment exchanging in hospital patient scheduling. *Service Orient Comp Appl* 2007; 1(3): 185–196.

45. Zöller A, Braubach L, Pokahr A, et al. Evaluation of a multi-agent system for hospital patient scheduling. *Int Trans Syst Sci Appli* 2006; (4): 375–380.

46. Huang J, Han Z, Chiang M, et al. Auction-based resource allocation for cooperative communications. *IEEE J Sel Area Comm* 2008; 26(7): 1226–1237.

47. Tipsuwan Y, Kamonsantiroj S, Srisabye J, et al. An auction-based dynamic bandwidth allocation with sensitivity in a wireless networked control system. *Comput Ind Eng* 2009; 57: 114–124.

48. Jonker CM, Robu V and Treur J. An agent architecture for multi-attribute negotiation using incomplete preference information. *Auton Agent Multi-Agent Syst* 2007; 15: 221–252.

49. Pendharkar PC. Game theoretical applications for multi-agent systems. *Expert Syst Appl* 2012; 39: 273–279.

50. Williamson M, Decker KS and Sycara K. Executing decision-theoretic plans in multi-agent environments. In: *AAAI fall symposium on plan execution*, Massachusetts Institute of Technology, Cambridge, Massachusetts, 9–11 November 1996. Cambridge, MA: AAAI Press.

51. Decker KS. Distributed problem-solving techniques: a survey. *IEEE T Syst Man Cyb* 1987; 17(5): 729–740.

52. Valkenaers P and Brussel HV. Holonic manufacturing execution systems. *CIRP Annals* 2006; 54: 427–432.
53. Schumann R, Lattner AD and Timm JJ. Management-by-exception – a modern approach to managing self-organizing systems. *Comm SIWN* 2008; 4: 168–172.

54. Ramani KV, Mavalankar DV and Govil D. *Strategic issues and challenges in health management*. Los Angeles, CA: SAGE, 2008, p. 227.

55. Prahalad CK. *The fortune at the bottom of the pyramid: eradicating poverty through profits* (Revised and Updated 5th Anniversary ed.). Upper Saddle River, NJ: Wharton School Publishing, 2009, p. 432.

56. Activity-report. Madurai, India: Aravind Eye Care System, 2014–2015, p. 76.

57. Brilliant L and Brilliant G. Aravind: partner and social science innovator (innovations case discussion: Aravind Eye Care System). *Innov Technol Gov Glob* 2007; 2(4): 50–52.

58. Mehta PK and Shenoy S. *Infinite vision – How Aravind became the greatest business case for compassion*. San Francisco, CA: Berrett-Koehler, 2011, p. 336.

59. Chaudhary Modi AG and Reddy K. Right To sight: a management case study on Aravind eye hospitals. *Zenith Int J Multi Res* 2012; 2(1): 447–457.

60. Rangan VK and Thulasiraj RD. Making sight affordable (innovations case narrative: The Aravind Eye Care System). *Innov Technol Gov Glob* 2007; 2(4): 35–49.

61. Andersen MM and Poulsen F. *Beyond strategy: the impact of next generation companies*. New York: Routledge, 2014.

62. Natchiar G, Thulasiraj RD and Sundaram RM. Cataract surgery at Aravind Eye Hospitals: 1988–2008. *Comm Eye Health* 2008; 21(67): 40–42.

63. Li J and Zha H. Two-way Poisson mixture models for simultaneous document classification and word clustering. *Comp Stat Data Anal* 2006; 50(1): 163–180.

64. Karlis D and Xekalaki E. Mixed Poisson distributions. *Int Stat Rev* 2005; 73: 35–58.

65. Gambardella LM and Dorigo M. An ant colony system hybridized with a new local search for the sequential ordering problem. *INFORMS J Comput* 2000; 12(3): 237–255.

66. Dorigo M and Gambardella LM. Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE T Evolut Comput* 1997; 1(1): 53–66.

67. Liker JK. *The Toyota way: 14 management principles from the world’s greatest manufacturer*. New York: McGraw-Hill, 2004, p. 330.

68. Hopp WJ and Spearman ML. *Factory physics-foundations of manufacturing management*. Irwin, CA: McGraw-Hill, 2001, p. 698.

69. Eswaramoorthy M, Kathiresan GR, Jayasudhan TJ, et al. Flow index based line balancing: a tool to improve the leanness of assembly line design. *Int J Prod Res* 2012; 50(12): 3345–3358.

70. Mac Gregor Smith J and Tan B. *Handbook of stochastic models and analysis of manufacturing system operations*. International series in operations research and management science. Vol. 192. New York: Springer-Verlag, 2013, p. 373.