Patient Journey Optimization Using A Multi-Agent Approach

Chung Ho CHOI

Abstract

Apart from the physical pains they suffer, most patients nowadays have to endure long waiting times during their patient journeys. While the alleviation of physical pains can mostly be done by the use of drugs, psychological stresses resulting from long waiting times are by no means trivial and always pose a threat to patients' life. Hence, an increasing attention has now focused on shortening the length of patient journey by scheduling patients in a more efficient way. In doing so, because of the decentralized structure of hospital settings, conventional centralized approaches such as operations research are hard to be applied and motivates the use of a multi-agent approach. In this paper, we propose a multi-agent framework for scheduling patients in a decentralized and efficient manner. Particularly, in order to demonstrate the effectiveness of the proposed multi-agent scheduling framework, simulations were performed based on a dataset containing about five thousand cancer patient journeys.

1 Introduction

Not to mention the physical pains they encounter, most patients nowadays also have to confront with psychological stresses resulting from long waiting times during their patient journeys. While physical pains can be alleviated by the use of drugs, psychological stresses resulting from long waiting times are by no means trivial and always pose a threat to patients' life. Since how a patient feels is well-linked to his or her medical condition [3, 8], we should not underestimate the impacts brought by such psychological stresses and thus long waiting times should not be tolerated.

In order to best utilize the existing resources to minimize undesired long waiting times, a well-designed scheduling algorithm is crucial [9]. Meanwhile, though conventional operations research methods have been found effective for centralized scheduling problems [10, 11, 5], most of them are not suitable for hospital settings where decentralized structures are found [6, 9, 2]. As a result, during the design of a patient scheduling algorithm, a multi-agent approach is proposed.

A multi-agent approach is characterized by emphasizing on local interaction and self-organization of different entities being modeled. These properties make it especially suitable for tackling complex tasks with a lot of stakeholders [13, 1]. Multi-agent methods have found applications in a variety of problem domains, such as airport resource scheduling [4], load allocation in transportation logistics [7], supply chain management [12], etc. Recently, it has also been applied to patient scheduling in [6, 9] with some initial success demonstrated. And yet, there are limitations. Paulussen *et al.*, in [6], assume that a quantified health state can be accurately derived as a utility measure for guiding the scheduling process. In [9], Vermeulen *et al.* did not consider the temporal constraints between the treatment operations during the scheduling process.

The objective of this study is to explore the extent to which a patient journey can be improved by better coordinating and mobilizing resources distributed at different medical units. In particular, we formulate the scheduling problem according to the cancer treatment practice in Hong Kong. We propose the use of a multi-agent framework in which autonomous agents interact with each other to arrive an effective overall schedule with reduced waiting times. To evaluate the effectiveness of the proposed framework, we made use of a patient identity anonymized dataset collected by Hospital Authority in Hong Kong which contains 4720 cancer patients with a diagnosis period spanning over 6 months, and have carried out simulations with the proposed approach given different settings of the environment.

The rest of the paper is organized as follows. A patient scheduling problem is formulated in Section 2. Section 3 and Section 4 present the details of the proposed agent-based scheduling algorithm. Section 5 presents some experimental results and Section 6 concludes the paper.

2 PROBLEM FORMULATION

In this section, we first briefly describe the establishment of the cancer clusters in Hong Kong. Then, we formulate the patient scheduling problem for cancer treatment as an optimization problem and explain how our proposed multi-agent framework can be adopted to address the distributed nature of the problem.

2.1 Cancer Patient Treatment - A Hong Kong Scenario

In Hong Kong, there are seven cancer clusters. Figure 1 shows the geographical distribution of the seven cancer clusters in Hong Kong. With the objective not to reveal the performance of individual clusters, we denote the set of the seven clusters as $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_7\}$. Currently, on-demand information exchange among the clusters for scheduling patients is not yet extensively used. That is why it is common for cancer patients to be scheduled to receive treatments at only one cancer cluster, even though some of the treatments could be provided earlier by other clusters.

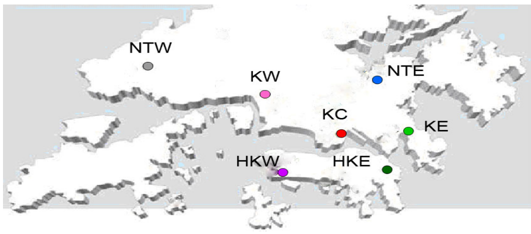

Figure 1. Seven geographically distributed cancer clusters in Hong Kong.

Generally speaking, once the case is suspected to be cancer for a patient, the doctor will specify the patient a treatment plan which contains a sequence of treatment operations. We denote the set of treatment operations as $\Gamma = \{\text{radiotherapy, surgery, chemotherapy}\}$.

To carry out the treatment operations, medical resources are needed. We denote the set of medical resources (or units) as $A = \{\text{radiotherapy unit, operation unit, chemotherapy unit}\}$. We assume that one treatment operation can only be performed at one medical unit of the corresponding type. A patient journey is defined as the duration from the date of histopathological diagnosis to the date of the last treatment operation completed.

2.2 Formulation

Let $K := A \times \mathcal{C}$ be the cartesian product of A and $\mathcal{C}$ giving the complete set of medical units, $M := K \rightarrow \Gamma$ be an one-to-one mapping between K and Γ specifying the treatment type of the medical units, and P be the set of cancer patients being scheduled.

Also, given a patient i, let N^i_Γ denotes the number of treatment operations needed, D^i_0 denotes the diagnosis date, D^i_j denotes the date of the j^{th} treatment operation where $1 \leq j \leq N^i_\Gamma$, $V^i_j \in K$ be the unit at which the j^{th} treatment operation is performed, $Tr^i_j \in \Gamma$ be the type of treatment for the j^{th} operation, C_k be the daily capacity (i.e. number of patients that could be treated) of medical unit $k \in K$, T_t be the duration (in days) of treatment type $t \in \Gamma$, and Z be the set of dates on which patient scheduling is being considered.

With the assumption that all the patients are being treated equally in terms of urgency, the scheduling problem can be formulated as:

$$\min_D \sum_{i=1}^{|P|} \sum_{j=1}^{N^i_\Gamma - 1} (|D^i_j - D^i_{j+1}|) \tag{1}$$

with the following constraints to be satisfied:

$$D^i_{j+1} > D^i_j + T_{Tr^i_j} \tag{2}$$

$$\forall d \in Z \quad \left|\{i : D^i_j = d \wedge V^i_j = k \wedge Tr^i_j = M(k)\}\right| \leq C_k \tag{3}$$

$$D^i_j > D^i_0 > 0 \tag{4}$$

The objective function in (1) is to minimize the time lags between treatment operations for cancer patients. Constraint (2) ensures the temporal constraints between treatment operations are not violated, constraint (3) is used to ensure all medical units are operating within their capacities. Constraint (4) ensures that patients would only be scheduled to receive treatment operations after their diagnoses.

3 SCHEDULING FRAMEWORK

Theoretically, patient waiting times could be minimized by optimizing (1). However, it is impractical to do so as it is hard to assume that a cancer cluster is willing to share its real-time resource allocation related data (e.g., C_k) with other clusters due to both technical and managerial reasons.

Hence, in this section, we propose the use of the multi-agent approach which tries to model each stakeholder as an autonomous agent and emphasizes on local interactions among the agents. It aims to minimize the information sharing requirement among the clusters and yet to obtain a good enough suboptimal result for the (global) patient journey optimization. In our proposed framework, there are two types of agents, namely *patient agents* and *resource agents*. They interact via some designed protocol for achieving the aforementioned optimization.

3.1 Patient Agent

A patient agent is used to represent one cancer patient and is denoted as P_i with $i = 1, 2, ..., |P|$. It stores the patient's treatment plan. As it is common that some treatment operations have to be performed in prior to another, the set of treatment operations to be received by a patient has to be ordered to satisfy certain temporal constraints. Hence, each patient agent P_i maintains an ordered set $Tr^i = \{Tr_1^i, Tr_2^i, ...Tr_{N_\Gamma^i}^i\}$ as its treatment plan.

3.2 Resource Agent

A resource agent is used to manage a specific medical unit. Here, we denote R_{ab} as a resource agent representing medical unit $a \in A$ at cancer cluster $b \in \mathcal{C}$. Each resource agent has full access to the schedule of the medical unit it represents, but not the others.

3.3 Scheduling Algorithm

We adopt a two-phase scheduling algorithm similar to what being proposed in [6, 9]. For each newly diagnosed cancer patient, a treatment plan is first designed and then the corresponding treatment operations are initially scheduled (initial assignment phase). Then, a timeslot-swapping process is enforced for shortening the patient journey (rescheduling phase). Here we assume that any two patient agents are willing to exchange their timeslots as far as none of their schedules is worsen (as suggested in [9]) and none of the temporal constraints as specified in *Eq.* (2) is violated.[1] Algorithm 1 gives a high-level description of this two-phases scheduling algorithm.

Algorithm 1 Scheduling Algorithm

```
for every patient agent P_i do
    Initial assignment based on P_i's treatment plan
    for each P_i's treatment operation do
        Rescheduled to be performed earlier by exchanging timeslot with another patient agent with the help of the resource agent (rescheduling phase)
        if No involving parties are worsened in terms of their resulting overall schedules then
            The exchanging process is proceeded
        end if
    end for
end for
```

4 AGENT COORDINATION

In this section, more details about the scheduling algorithm are given, including 1) how the patient agents interact with the resource agents, and 2) how some "unnecessary" swappings can be rejected so as to further improve the scheduling optimality.

[1] This assumption may imply that some policy-wise incentive has to be in place so that different medical units are willing to share their resources in this manner, which however is not the main focus of our study.

4.1 A bidding process for agent matchmaking

Figure 2 shows our proposed framework. As what have been introduced in Sections 3.1 and 3.2, there are two types of agents, namely patient agents and resource agents. In order to show clearly the coordination between agents, we further categorize patient agents into *initiating patient agents* and *target patient agents*. Initiating patient agents P_I are those patient agents who initiate a request for timeslot exchange. Target patient agents P_G are the others who are willing to participate in the exchanging process.

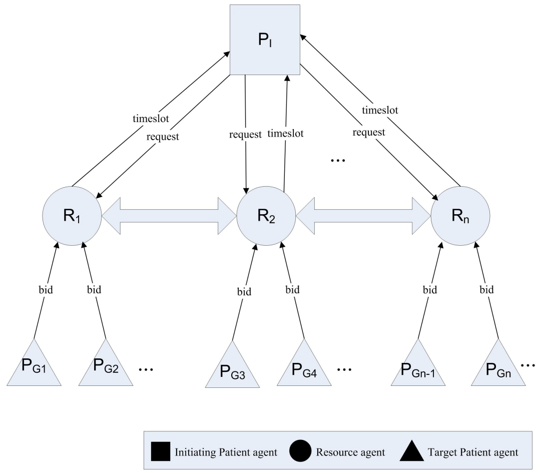

Figure 2. The proposed agent coordination framework for patient scheduling.

With the objective of shortening its patient journey, an initiating patient agent P_I first sends out a request for rescheduling to the corresponding resource agents R_{ab}. The request includes the earliest possible start date (*EPS*) and the latest possible start date (*LPS*) of its associated treatment operation. In order not to violate the temporal constraints between treatment operations, the *EPS* can be defined as:

$$EPS_j^I = D_{j-1}^I + T_{Tr_{j-1}^I} + \delta_1. \tag{5}$$

Note that δ_1 denotes how many days a patient should be admitted (if needed) before a treatment operation to be carried out. In our experiment, we set to be one. In practice, this value could be designated by healthcare providers in order to better suit their needs. With a similar argument, LPS is defined as:

$$LPS_j^I = D_j^I - 1. \tag{6}$$

Once a resource agent receives a request with EPS and LPS, it will first check whether there are available timeslots released by deceased patients which can fulfill the request. If yes, the released timeslot will be assigned to the initiating patient agent. If not, the resource agent will then pass the request to those patient agents (target patient agents, P_G) which reserved resources of the same type in the period from EPS to LPS. Those target patient agents who have received the request will submit a bid to the resource agent in response.

There are several factors needed to be considered in computing the bid value.

- First, the target patient agent should not have its last treatment operation in its treatment plan to be exchanged, or its last treatment operation has then to be performed later and thus it would end up with a lengthened patient journey.

- Second, as it is impractical to reschedule a patient's treatment operation without prior notification, we assume that the exchange of timeslots would not be considered if the initiating patient will have less than a week's time of notification.[2]

- Third, the target patient agent also has to ensure that the temporal constraints between its treatment operations would not be violated after the exchanging process.

Taking into account the above considerations, the bid value submitted by a target patient agent P_G is formulated as:

$$Bid^G = (D^G_{j_t} - EPS^I_{j_i}) + Last + Noti + Temp, \quad (7)$$

where $Last$, $Noti$ and $Temp$ are three binary variables. $Last = 0$ if the j_tth operation is not the last one for P_G, or ∞ otherwise. $Noti = 0$ if there is a week's time of notification for the target patient agent to be notified, or ∞ otherwise. $Temp = 0$ if there are no temporal constraints violated, or ∞ otherwise.

Among all the target patient agents, the one with the lowest bid value will be accepted and the timeslot swapping between the initiating agent and target agent will be proceeded. If two bids are found to be numerically identical, the resource agent will select one at random.

4.2 A coordination process for rejecting unnecessary swappings

A timeslot swapping as described in the previous section sometimes does not necessarily lead to ultimate improvement in patient journey. To illustrate that, suppose there is a patient agent with 3 treatment operations to be rescheduled. In case the last treatment operation could not be rescheduled to be performed earlier, any rescheduling of the first 2 treatment operations are essentially useless as the duration of the whole journey remains unchanged (see Figure 3(a)). As another example, even a shortened patient journey can be achieved, rescheduling of the first 2 treatment operations could also be useless if the rescheduling of the last treatment operation cannot be benefited from the rescheduling of the first two (see Figure 3(b)).

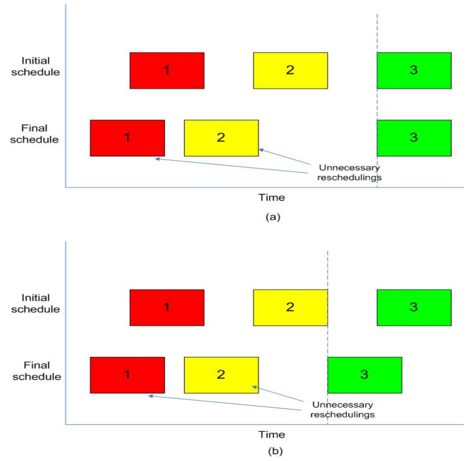

Figure 3. Unnecessary reschedulings.

In order that these useless swappings can be rejected so as to be reserved for other potentially more useful swappings, the scheduling algorithm could be modified in such a way that a resource agent after identifying the most optimal bid among the target patient agents will not notify the initiating patient agent immediately. Instead, it will pass the bid to the resource agent which is responsible for the succeeding treatment operation of the initiating patient agent. Having received such a bid, the resource agent could derive a new EPS, denoted as ${}^{(new)}EPS^I_{j_i+1}$. Clearly, unnecessary swappings occur if that resource agent could not find a bid among those received from the target patient agent P_G such that ${}^{(new)}EPS^I_{j_i+1} \leq D^G_{j'_t} \leq EPS^I_{j_i+1}$, where $Tr^I_{j_i+1} = Tr^G_{j'_t}$. In that case, the resource agent will notify its antecedent to discard the bid such that the corresponding timeslots would not be exchanged. In general, such a succeeding resource agent consultation process can be carried out in a recursive manner.

5 EXPERIMENTS

In order to evaluate the effectiveness of the proposed multi-agent framework, we first obtained a dataset containing the scheduled treatment plans of 4720 cancer patients being treated at the seven cancer clusters in Hong Kong with a diagnosis period spanning 6 months (from 1/7/2007 to 31/12/2007). The average length of the patient journey among all cancer clusters is 82.4 days. Based on the dataset,

[2]In general, the time of notification can be adjusted according to the real situation.

we have carried out simulations with the following 4 experimental settings:

Setting 1: Patient agents are willing to exchange timeslots with others whenever there is a Pareto improvement.

Setting 2: It is assumed that only 20% of the patients of each cancer cluster are allowed to undergo timeslot swapping.

Setting 3: It is assumed that patients are reluctant to travel for a long distance even though some of their operations can be scheduled earlier, and thus only swappings between two nearby cancer clusters are allowed. In particular, the neighborhood relationships are assumed to be

- $C_1 \to C_2$ or C_3
- $C_2 \to C_1$ or C_3
- $C_3 \to C_1$ or C_2 or C_4 or C_5 or C_6
- $C_4 \to C_3$ or C_6
- $C_5 \to C_3$ or C_7
- $C_6 \to C_3$ or C_4
- $C_7 \to C_5$

where $\alpha \to \beta$ implies that patients admitted in cancer cluster α would only be swapped to its neighboring cancer cluster β.

Setting 4: Timeslots released by deceased patients are allocated to the patient agents who have the longest patient journeys at a time point.

Given the four aforementioned settings, Figure 4 shows the average length of patient journey among the seven cancer clusters in Hong Kong.

The experimental results obtained show that, on average, the average length of journey can be reduced by 6.0 days for those 4720 cancer patients if no restriction is imposed on the exchange of timeslots whenever there is a Pareto improvement (Setting 1). Given only 20% of patients per cancer cluster are allowed for timeslot exchange (Setting 2), we found that the average length of journey could still be reduced by an average of 3.4 days. With the geographical restriction on allowing only swappings between nearby cancer clusters (Setting 3), the average length of journey can also be reduced by 4.4 days.

However, it should also be noted that according to Figure 5, the maximum length of journey remains unchanged. The reason is obvious as no one is willing to swap with those with the longest length of journey. Reductions on the maximum length of patient journey can only be observed for Setting 4 where the released timeslots due to deceased patients are allocated to those with the longest patient journey.

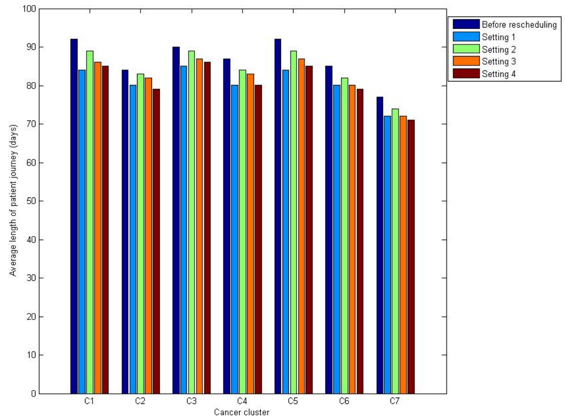

Figure 4. Average length of patient journey among the seven cancer clusters under 4 different settings.

5.1 Simulations revealing the impacts of varying the unit capacities

For all the results presented so far, it is assumed that the capacity of each medical unit is fixed. To study the cost-effectiveness of increasing the capacities of medical units for patient journey optimization, we had conducted several simulations in which timeslots were additionally allocated to each medical unit. Particularly, 3 different timeslot-allocation strategies were performed in our simulations:

1. Timeslots were added on a *daily* basis. In our simulation, 2 timeslots were added daily to each medical unit.

2. Timeslots were added on a *weekly* basis. In our simulation, 14 timeslots were added weekly to each medical unit.

3. Timeslots were added on a *monthly* basis. In our simulation, 60 timeslots were added monthly to each medical unit.

It is worth to note that, let say, if timeslots were added on a weekly basis, cancer patients would then be scheduled on a weekly basis too (i.e. a cancer patient would be scheduled to receive treatment operation in a certain week if the corresponding weekly capacity does not exceed its limit).

By Figure 6, it was found that when the timeslot-allocation strategy was changed from a daily basis to a weekly basis, and then to a monthly basis; the utilization of medical units will increase subsequently. The reason is that when the medical resources were allocated in a more

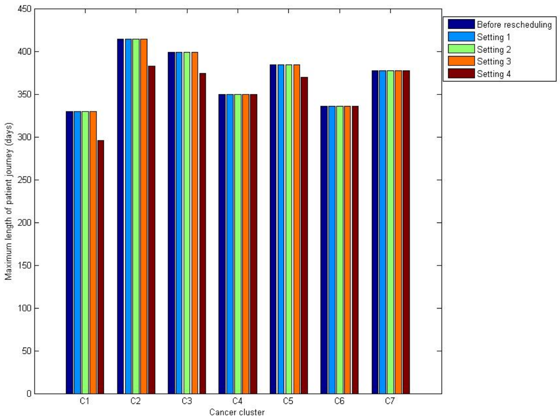

Figure 5. Maximum length of patient journey among the seven cancer clusters under 4 different settings.

flexible way (i.e. with a wider time frame considered), the possibility of assigning a cancer patient to receive treatment operation on a particular date will increase as a result.

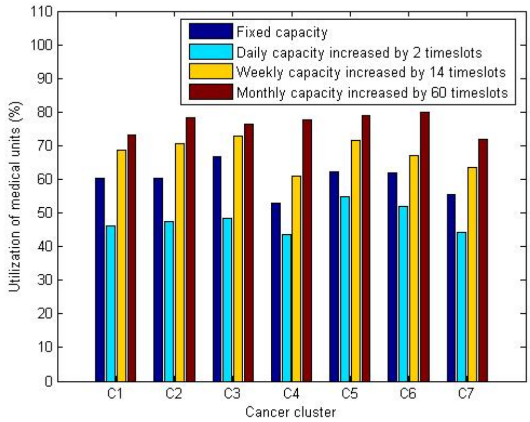

Figure 6. Utilization of medical units by varying the timeslot-allocation strategies.

Meanwhile, it was also observed that, both the average and maximum length of patient journey will drop significantly when additional timeslots were allocated to each medical unit. In particular, such significant drop can be found before (see Figure 7 and Figure 8) and after (see Figure 9 and Figure 10) the rescheduling phase. In fact, among the 3 timeslot-allocation strategies, we found that a monthly-basis strategy is the most effective one in achieving an optimal patient journey.

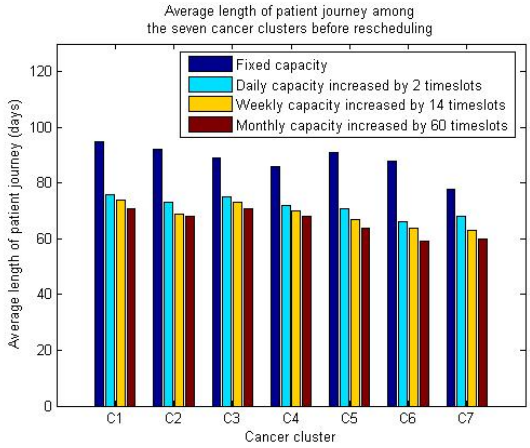

Figure 7. Average length of patient journey among the seven cancer clusters by varying the timeslot-allocation strategies (before rescheduling).

5.2 Simulations revealing the impacts of reducing the durations between treatment operations

During our simulations, we had also tried to investigate the impacts induced by the durations between treatment operations. At this time, while all the unit capacities are fixed, we reduced all the treatment durations by half.

Interestingly, it was observed that when the treatment durations were reduced by half, a significant reduction in both the average and maximum length of patient journey can also be achieved before (see Figure 11 and Figure 12) and after (see Figure 13 and Figure 14) the rescheduling phase. With such insight, we got an important implication: whenever the durations between treatment operations can be minimized in accordance with medical reasons, cancer patients could then enjoy themselves with less undesired waiting times. In other words, during the practical implementation, it is important for the healthcare provider to carefully and unbiasedly quantify such durations between treatment operations.

6 CONCLUSIONS

In this paper, a multi-agent framework was proposed for patient journey optimization. Particularly, by applying the framework, the shortening of a patient journey will not lengthen the journeys of the others. Also, all the temporal constraints among the treatment operations for each patient would not be violated during the scheduling process.

The effectiveness of the proposed framework has been demonstrated by applying it to a dataset containing 4720 scheduled treatment plans of cancer patients admitted to

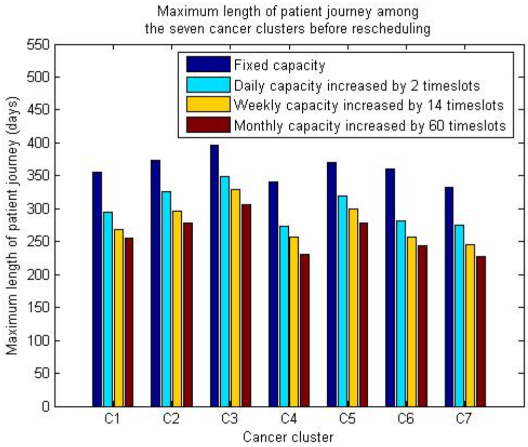

Figure 8. Maximum length of patient journey among the seven cancer clusters by varying the timeslot-allocation strategies (before rescheduling).

hospitals in Hong Kong. The effects of varying the unit capacities and treatment durations on the overall reduction in length of patient journey are also studied.

Because of the limited resources during practical implementation; in the near future, rather than routinely allocate a fixed amount of additional timeslots to each cancer cluster as what had been demonstrated earlier, we are going to assess how resources (or timeslots) should be allocated to cancer clusters in a more efficient and unbiased way such that the overall patient journey could be shortened in a greater extent.

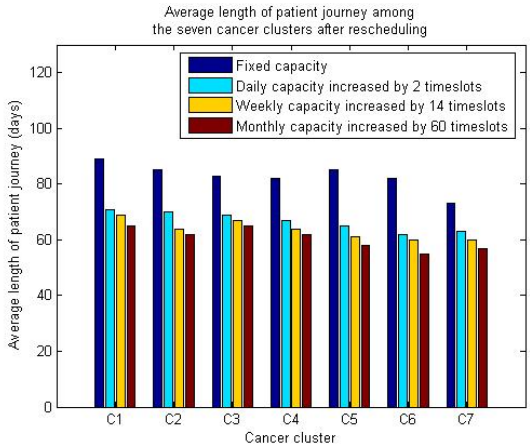

Figure 9. Average length of patient journey among the seven cancer clusters by varying the timeslot-allocation strategies (after rescheduling).

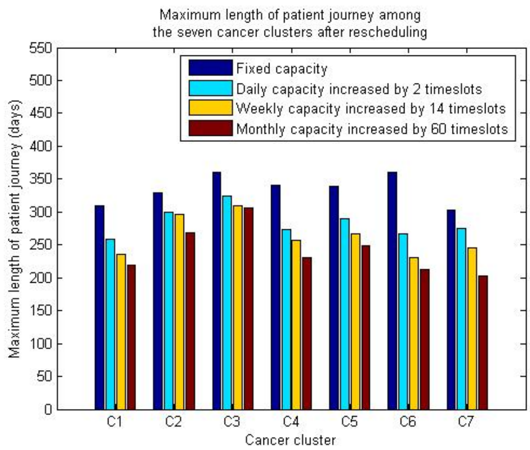

Figure 10. Maximum length of patient journey among the seven cancer clusters by varying the timeslot-allocation strategies (after rescheduling).

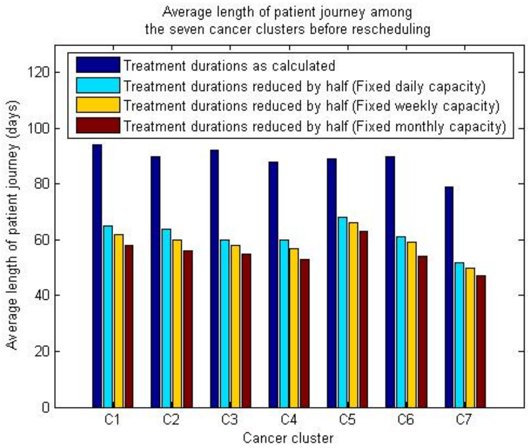

Figure 11. Average length of patient journey among the seven cancer clusters by reducing the treatment durations by half (before rescheduling).

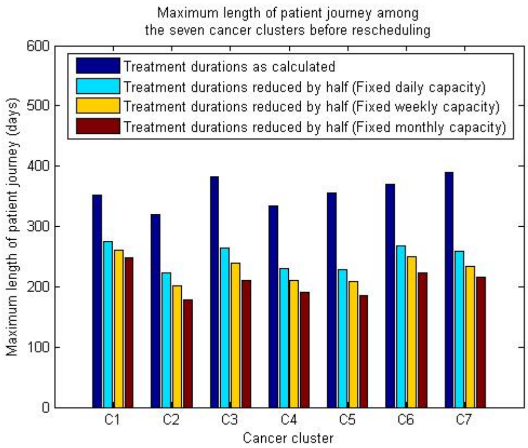

Figure 12. Maximum length of patient journey among the seven cancer clusters by reducing the treatment durations by half (before rescheduling).

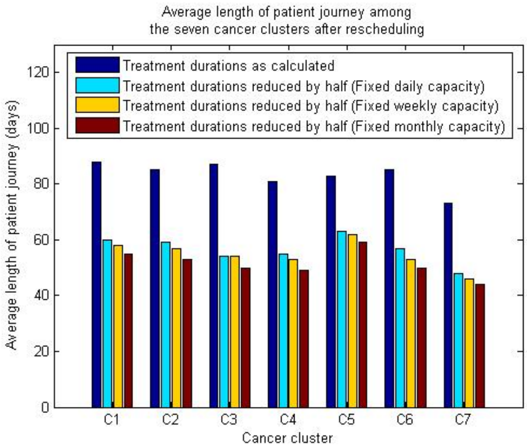

Figure 13. Average length of patient journey among the seven cancer clusters by reducing the treatment durations by half (after rescheduling).

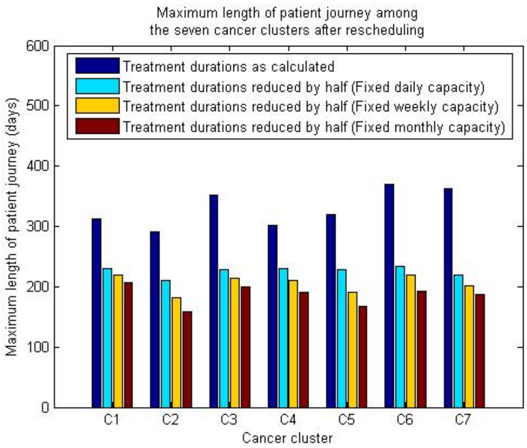

Figure 14. Maximum length of patient journey among the seven cancer clusters by reducing the treatment durations by half (after rescheduling).

ACKNOWLEDGEMENT

This is to acknowledge Hospital Authority in Hong Kong for providing the dataset to support this study.

References

[1] H. Czap and M. Becker. Multi-agent systems and microeconomic theory: A negotiation approach to solve scheduling problems in high dynamic environments. In *HICSS '03: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 3*, page 83.2, Washington, DC, USA, 2003. IEEE Computer Society.

[2] K. Decker, J. Li, and Y. Demazeau. Coordinating mutually exclusive resources using gpgp. *Autonomous Agents and Multi-Agent Systems*, 3:200–0, 2000.

[3] Department of Health, UK. Now I feel tall: What a patient-led NHS feels like. 2005.

[4] X. Mao, A. Mors, N. Roos, and C. Witteveen. Coordinating competitive agents in dynamic airport resource scheduling. In *MATES '07: Proceedings of the 5th German conference on Multiagent System Technologies*, pages 133–144, Berlin, Heidelberg, 2007. Springer-Verlag.

[5] J. Patrick and M. Puterman. Reducing wait times through operations research: optimizing the use of surge capacity. *Healthc Q*, 11(3):77–83, 2008.

[6] T. O. Paulussen, I. S. Dept, K. S. Decker, A. Heinzl, and N. R. Jennings. Distributed patient scheduling in hospitals. In *Coordination and Agent Technology in Value Networks. GITO*, pages 1224–1232. Morgan Kaufmann, 2003.

[7] V. Robu, H. Noot, H. La Poutré, and W.-J. van Schijndel. An interactive platform for auction-based allocation of loads in transportation logistics. In *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 3–10, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[8] M. Simunovic, A. Gagliardi, D. McCready, A. Coates, M. Levine, and D. DePetrillo. A snapshot of waiting times for cancer surgery provided by surgeons affiliated with regional cancer centres in Ontario. *CMAJ*, 165(4):421–425, August 2001.

[9] I. Vermeulen, S. Bohte, K. Somefun, and H. La Poutre. Improving patient activity schedules by multi-agent pareto appointment exchanging. In *CEC-EEE '06: Proceedings of the The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*, page 9, Washington, DC, USA, 2006. IEEE Computer Society.

[10] J. Vissers and R. Beech. *Health operations management : patient flow logistics in health care*. Routledge, 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN, 2005.

[11] J. Vissers, J. Bekkers, and I. Adan. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*, 16, 2005.

[12] M. Wang, J. Liu, H. Wang, W. K. Cheung, and X. Xie. On-demand e-supply chain integration: A multi-agent constraint-based approach. *Expert Systems with Applications*, 34(4):2683 – 2692, 2008.

[13] G. Weiss. *Multiagent systems : a modern approach to distributed artificial intelligence*. Cambridge, Mass. : MIT Press, 1999.