35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Preference Based Load Balancing As An Outpatient Appointment Scheduling Aid

Uthpala Subodhani Premarathne* Fengling Han, Ibrahim Khalil, Zahir Tari†

*National ICT australia (NICTA), School of Computer Science and IT, RMIT University, Melbourne, VIC 3001, Australia.
Email:s3308412@student.rmit.edu.au
†School of Computer Science and IT, RMIT University, Melbourne, VIC 3001, Australia.
Email:fengling.han@rmit.edu.au, ibrahim.khalil@rmit.edu.au, zahir.tari@rmit.edu.au

Abstract— **Load balancing is a performance improvement aid in various applications of distributed systems. In this paper we propose a preference based load balancing strategy as a scheduling aid in an outpatient clinic of an online medical consultation system. The performance objectives are to maximizing throughout and minimizing waiting time. Patients will provide a standard set of preferences prior to scheduling an appointment. The preferences are rated on to a scale and each service request will have a respective preference score. The available doctors will also be classified into classes based on their clinical expertise and the nature of the past diagnosis and the types of patients consulted. The preference scores will then be mapped on to each class and the appointment will be scheduled. The proposed scheme was modeled as a queuing system in Matlab. Matlab SimEvents library modules were used for constructing the model. Performance was analysed based on the average waiting time and utilization. The results revealed that the preference based load balancing scheme markedly reduce the waiting time and significantly improve the utilization under different load conditions.**

I. INTRODUCTION

Appointment scheduling in a healthcare system balances the demand from patients with the service provider availability. As identified in [1], main factors that affect the performance of appointment scheduling are

- patient and provider preferences
- patient arrival variabilities
- service time variability depending on the severity of the health condition
- availability of information technology resources and
- the level of experience of scheduling staff

Patient arrival varaibilites include punctuality, no-shows, cancellations [2][3]. Cancellation of an appointment is when a patient not-arriving for a scheduled appointment with prior notification. No-show is when a patient does not arrive for a scheduled appointment without any prior notification. As revealed in [4], failed-appointments of 12% to 42% have created serious inefficiencies. Figure 1 shows the scheduling problem with waiting time, idle time and the effect of failed appointments at a single service service provider.

For an outpatient clinic the common preference is the date and time of an appointment. As the service facilitating resources are limited and it is often the case where emergency medical cases need to be given priority. So it is not easy to apply preference driven resource reservation models such as those used in the airlines [1]. Potential benefit of patient centric preference driven appointment scheduling can maximize the quality of service delivery [5]. From the patients point of view, they can get the necessary medical attention from the most comfortable service environment. From the medical practitioners point of view the best possible service can be delivered efficiently using their knowledge and expertise from past experiences in handling similar cases. These potential incentives thus intrigues to find out alternative approaches for incorporating preference driven appointment scheduling.

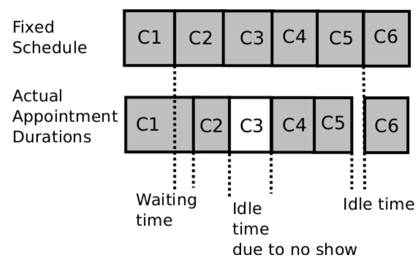


Fig. 1 Appointment Scheduler at a Single Service Provider

To achieve a outpatient consultation system high service satisfaction with high resource utilization we forsee that an online consultation system has the required ability. Yet it needs to be supplemented with a mechanism that can guide the scheduling mechanism to allocate appropriate resources to service requests without compromising the set performance objectives. Load balancing is a well known performance improvement strategy [6]. This concept can empower an online consultation system to efficiently distribute the preference specific appointments to appropriate resources (i.e. medical consultation doctors).

The main contributions of this paper are,

- to propose a preference based load balancing scheme for scheduling outpatient appointments and
- to improve the service performance based on waiting time and throughput by applying the preference based load balancing strategy.

II. Related Work

A. Outpatient Department Model

Patients can be of two classes; those who arrive for the first time and have visited before. These patients may have a prescheduled appointment or walk-ins (i.e. without a prescheduled appointment). Based on the study [7], general patient flow in an outpatient department can be modeled as follows. All the patients have to undergo an initial service selection phase where the authentication of the patient, confirmation of the required service and initial screening is performed. For a first-time or new patient an initial registration is mandate at his stage. Lab tests are performed for the patients with appointments at scheduled appointment times. The patient flow for those who are with appointments for general consultations and walk-ins are shown in Figure 2. The service facilitation is priority based where the walk-ins are given lower priority than the scheduled patients.

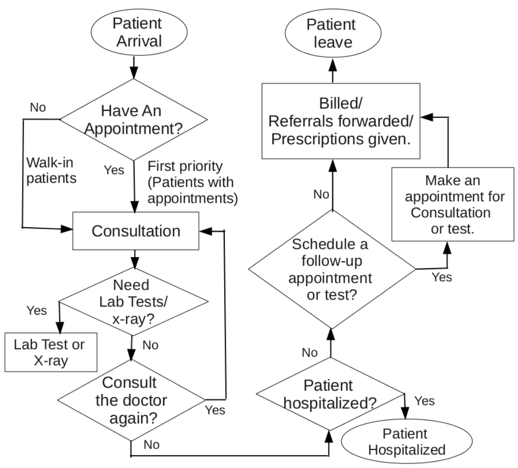


Fig. 2 Patient Flow - Patients with Appointments and Walk-Ins

In an outpatient clinic, future appointments are made to a general practitioner who will be available at a specific date and time. The personal preferences or the medical condition related information (e.g. symptoms) are not necessarily evaluated with the physicians previous experience in handling such medical conditions. Based on this discussion it reveals that there is no explicit preference based selective criteria for scheduling appointments.

B. Applicability of Load Balancing as a Performance Improvement Measure

The concept of load balancing can be effectively applied for appointment scheduling in an online medical consultation system. Consider the scenario where the medical centers are modelled to operate as a peer-to-peer network. The available resources are the doctors and the patients are considered as incoming traffic requesting services from these resources. Load balancing can thus be utilized to assign the incoming service requests to avoid any under-utilized or over-booked resources. When the patients provide a preference based service request the resources that matches the preference criterion can thus be allocated. On the other hand when there are walk-in patients this scenario is further complicated. The walk in patients do not have prior appointments but have to be served as first-come-first-serve basis. Load balancing can be utilized even in this scenario to assist in finding the most likely available doctor for service without overprovisioning appointments. When there are idle slots due to no-shows these can be filled up efficiently in an online consultation system ensuring workload balance between available resources. Therefore it is seemingly a viable effective solution to make use of a load balancing strategy for appointment scheduling in an outpatient clinic.

C. Motivation

Generally an outpatient clinics use block based fixed schedule appointment system and generally the schedule is fixed with minimum reservations for walk-ins or exceptionally fills in to the cancelled or no-show slots [3]. The common preference attributes that are considered are the date and time [8]. Often the patients personal and medical history related information are not accounted in scheduling an appointment. However such information can help to find the most appropriate doctor for consultation by mapping the prefernce criterion on to his/her proven expertise and experience in handling similar cases in the past. Further more patient centric preference driven appointment scheduling can maximize the quality of service delivery [5]. Based on the above discussion, a preference driven appointment scheduling strategy can be a significant service quality improvement factor.

III. Preference Based Load Balancing Model

In [9] it is assumed that patients have pre-notion of a set of preferred appointment slots. The clinic then decides whether to offer the patient an appointment in the set or to reject the patient. However these policies do not smooth out the daily load of the clinic. We propose a preference based load balancing scheme for appointment scheduling (Figure 3).

The proposed scheme has two main operational phases: 1)service based patient classification and priority assignment and 2) load balanced appointment scheduling. The preferences are identified by asking a number of questions from the patient. We assume the patients are punctual so that the effect of tardiness is neglegible. A patient can explicitly state a set of preferences in terms of answers to a standard set of questions with single or multiple possible answers. Patients can have medical condition related and personal preferences. An example of personal preferences would be a female patient preferring to see a female medical practitioner. Preferences based on medical condition would be straight forward. For an example a skin ulcer or a rash would be best to treated by a dermatologist. Based on the given preferences a score will be calculated which will be used to identify the appropriate service class. The clinical appointment would then be scheduled to a medical practitioner, who is assigned to the appropriate service class, with the queue that is least likely to overload. This will improve the quality of service delivery as well as the service satisfaction.

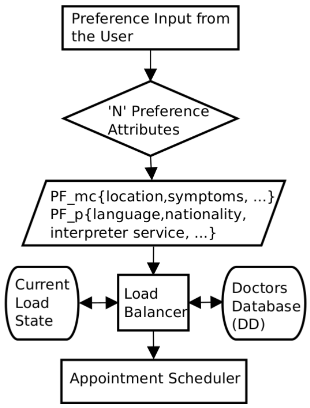


Fig. 3 Preference Based Load Balancing Approach for Appointment Scheduling

We define a patient to enter personal preferences PF_p and medical condition based preferences PF_{mc}. Assume that prospective patients arrive according to a Poisson process with rate λ and have a service time (s). Each service class l has a service time distribution $A_l(s)$. As given in [10][11], the average waiting time for the k^{th} service class is given by,

$$W_{l,k} = \frac{\frac{1}{2}\lambda \int_0^\infty s^2 dA(s)}{(1-\sum_1^{k-1}\rho_l)(1-\sum_1^k \rho_l)} \tag{1}$$

where $\rho_l = \frac{\lambda_l}{\mu_l}$ is the utilization. The service facilitation is assumed to be non-preemptive. In general, most of the out patient departments and out patient medical clinics do not have a large number of doctors. Thus if there are too many preference classes the patient would then be imposed with strict selection barriers. More questions will have to be asked as more fine grain information on preferences is required. To avoid such restrictions it is best to keep a small number of preference classes. Consider $C = 2$ classification policy for C_1 and C_2. Suppose C_1 is given priority over C_2 then the average waiting times of each class are,

$$W_1 = \frac{\frac{1}{2}\lambda \int_0^\infty s^2 dA_1(s)}{(1-\rho_1)} \tag{2}$$

$$W_2 = \frac{\frac{1}{2}\lambda \int_0^\infty s^2 dA_2(s)}{(1-\rho_1)(1-\rho)} \tag{3}$$

We define the preference metric of a prospective patient i as $\delta_i = B(PF_p, PF_{mc})_i$. A patient is allocated to class C_1 if the threshold level Th_1 is exceeded i.e. $\delta_i \geq Th_1$. Or classified to class C_2 if the threshold level Th_2 is not exceeded i.e. $\delta_i \leq Th_2$. If the preference metric is $Th_2 < \delta < Th_1$, then the patient is scheduled to a medical practitioner with the lowest number of queued appointments (or least loaded queue) at that time by avoiding any over-loaded conditions and congestion. Simple round robin method is selected for load balancing as it is simple to implements and provides better performance than other popular index-based methods (e.g. hashing) [12].

Load of a single queue at time t is defined as the number of occupied slots. In an outpatient clinic a service queue of a single medical practitioner would be the number of appointments waiting to be served. The waiting time for the i^{th} patient will depend on the previous delays in serving the $(i-1),(i-2),...,2,1$ patients as well as the other additional delays (e.g. delays in printing or faxing test reports or prescriptions). When the patients are allocated onto a preference score matched class, the scheduled practitioner will get those patients he can treat using his expertise. Thus it is more productive and avoids prolonged service times or time doing additional tests which might not even be useful for the final diagnosis.

IV. Simulation Results and Discussion

A. Performance Analysis with Preference Based Load Balancing

Consider the scenario when there are two preference classes. We assume the simulation model to have no misclassification errors and that the number of patients with complex medical conditions arrive less frequently than those with non-complex medical conditions. So the number of patients in class 02 will be more than in class 01. For patients of preference class 01 medical condition based preferences are more complex thus require longer service time in comparison to those in preference class 02. We modelled this scenario using the MATLAB SimEvent modules. The schematic diagram of the model is shown in Figure 4.

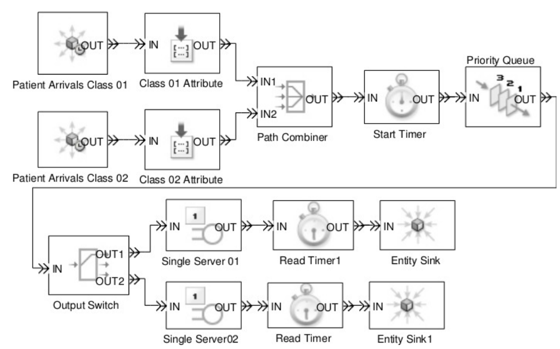


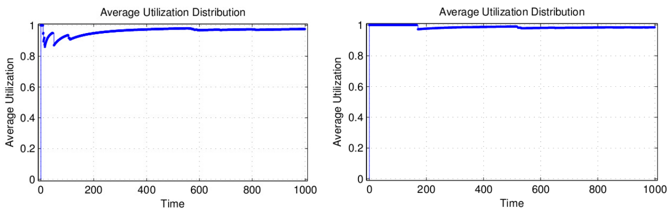
Fig. 4 Matlab Queueing Model Schematic

The results reveal that despite the small number of patients in class 01, the average system time is longer for class 01 than for the class 02 (Table I). Thus medical preference based patient classification will clearly improve the timely service facilitation for patients with more complex medical conditions. Under heavy-load conditions the service time increases for both classes.

TABLE I Service Time and Waiting Time Variation

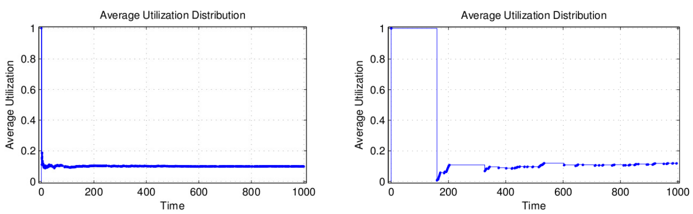
Preference Class	Class 01	Class 02
Number of Patients (low-load)	10	100
Average Service Time (seconds)	6	3.5
Number of Patients (high-load)	50	1000
Average Service Time (seconds)	8	11

Utilization for class 01 (Figure 6a.) is lesser in comparison to that of class 02 (Figure 5a.). This is mainly because the lesser number of class 01 patients. Therefore it clear that preference based classification can significantly improve the service satisfaction for patients. And it can be recommended that it is not necessary to allocate more medical practitioners to class 01 since the complex medical cases are less frequent.



(a) (b)

Fig. 5 Average Utilization Comparison- Class 01



(a) (b)

Fig. 6 Average Utilization Comparison- Class 02

When the system is heavily loaded waiting time exceeds the average expected service time. Such long waiting times can be reduced by rescheduling patients assigned for each preference class if the waiting time exceeds a threshold. We model this scenario in Matlab by rescheduling the timed-out entities with higher priority. We used the modified the model 4 to simulate the threshold detected load balanced rescheduling scenario. The results reveal that the utilization markedly improves for class 01 (Figure 6a. and Figure 6a.) when there is a redirection of timed-out entities with higher priority and the utilization remains at 100% over 90% of the time for class 02 (Figure 6b.).

V. Conclusion

In conclusion, based on the simulation of two preference classes the preference based service load balancing proves to be an efficient appointment scheduling aid by improving the utilization and reduces the average waiting time for patients when there is no classification errors and when patients with complex medical conditions are less frequent. The utilization can further be improved when threshold based rescheduling is incorporated. Since the patient is directing the clinic how to effectively schedule the most appropriate appointment there is greater service satisfaction for the patient. On the other hand since the medical practitioner gets patients to suite his expertise quality service rendering can be guaranteed. Thus, using this method can balance between the patients preferences, quality of service and service availability. Therefore we find the proposed scheduling strategy would be of great practical use for outpatient clinics in improving the quality of service facilitation.

VI. Future Work

We hope to extend this research focus by validating the proposed preference based classification for appointment scheduling in an outpatient clinic by using empirical data. We foresee that an empirical data based validation is necessary to fortify the proposed strategy to be of more practical significance.

VII. Acknowledgement

We wish to acknowledge NICTA as the main funding source for the research studentship. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

References

[1] D. Gupta and B. Denton, "Appointment scheduling in healthcare: Challenges and opportunities," *IIE Transactions*, vol. 14, pp. 800–819, 2008.

[2] L. LaGanga and S. Lawrance, "Appointment scheduling with overbooking to mitigate productivity loss from no-shows," in *Proceedings of Decision Sciences Institute Annual Conference*, November 2007, pp. 17–20.

[3] T. Cayiril and E. Veral, "Outpatient scheduling in healthcare: A review of literature," *Productions and Operations Management*, vol. 12 (4), p. 00, 2003.

[4] V. Lee, A. Ernest, M. Chen, and B. Krishnan, "Predictors of failed attendances in a multi-speciality outpatient center using electronic databases," *BMC Health Services*, vol. 5 (51), p. 00, 2005.

[5] L. Zandbelt, E. Smets, F. Oort, M. Godfried, and H. de Haes, "Satisfaction with the outpatient encounter a comparison of patients' and physicians' views," *Journal of General Internal Medicine*, vol. 19(11), pp. 1088–1095, 2004.

[6] H. Bryhni, "A comparison of load balancing techniques for scalable web servers," *IEEE Networks*, vol. 14 (4), pp. 58–64, 2000.

[7] A. Wijewickrama and S. Takakuwa, "Simulation analysis of appointment scheduling in an outpatient department of internal medicine," in *Proceedings of the Winter Simulation Conference*, 2005, pp. 2264–2273.

[8] N. Liu, S. Ziya, and V. Kulkarni, "Dynamic scheduling of outpatient appointments under no-shows and cancellations," *Manufacturing and Service Operations Management*, vol. 12 (2), pp. 347–364, 2010.

[9] D. Gupta and L. Wang, "Revenue management for a primary-care clinic in the presence of patient choice," *Operational Research*, vol. 56 (3), pp. 576–592, 2008.

[10] A. Cobham, "Priority assignment in waiting line problems," *Journal of Operational Research Society of America*, vol. 2 (1), pp. 70–76, 1954.

[11] S. von Zee and H. Theil, "Priority assignment in waiting-line problems under conditions of misclassification," *Operational Research*, vol. 9, pp. 875–885, 1961.

[12] D. DeWhitt, J. Naughton, D. Schindler, and S. Seshadri, "Practical skew handling in parallel joins," in *Proceedings of the 19th VLDB Conference*, August 1992, pp. 27–40.