8-2013

Improving Patient Length-of-Stay in Emergency Department Through Dynamic Resource Allocation Policies

Kar Way TAN
Singapore Management University, karway.tan.2007@smu.edu.sg

Wei Hao TAN
Singapore Management University, weihao.tan.2010@smu.edu.sg

Hoong Chuin LAU
Singapore Management University, hclau@smu.edu.sg

DOI: https://doi.org/10.1109/CoASE.2013.6653988

Citation

TAN, Kar Way; TAN, Wei Hao; and LAU, Hoong Chuin. Improving Patient Length-of-Stay in Emergency Department Through Dynamic Resource Allocation Policies. (2013). *IEEE International Conference on Automation Science and Engineering (CASE)*. 984-989. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/1934

Improving patient length-of-stay in emergency department through dynamic resource allocation policies*

Kar Way Tan, Wei Hao Tan and Hoong Chuin Lau[1]

Abstract— **In this work, we consider the problem of allocating doctors in the ambulatory area of a hospital's emergency department (ED) based on a set of policies. Traditional staffing methods are static, hence do not react well to surges in patient demands. We study strategies that intelligently adjust the number of doctors based on current and historical information about the patient arrival. Our main contribution is our proposed data-driven online approach that performs adaptive allocation by utilizing historical as well as current arrivals by running symbiotic simulation in real-time. We build a simulation prototype that models ED process that is close to real-world with time-varying demand and re-entrant patients. The experimental results show that our approach allows the ED to better cope with demand surges and to meet a service level desired by the hospital.**

I. INTRODUCTION

The Emergency Department (ED) is often seen as a place with long waiting time and lack of doctors to serve the patients. However, it is one of the most important departments in a hospital that is required to efficiently serve patients with critical medical needs. Since patients arrive at the ED without an appointment, it is a challenge to plan the number of resources (e.g., doctors and nurses) and effectively utilize the resources in the ED that best meet the unpredictable demand and satisfy a desired service quality level (e.g., length-of-stay (LOS) in the ED).

In this paper, we consider a motivating case based on a real-life study in a local hospital where a challenging service quality is presented and a static planned doctor's schedule is used in the ED. The hospital's ED has two sections of operations - the *resuscitation and critical care* area (also referred to as the back room in this paper) serves critical patients of acuity levels 1 and 2; the *ambulatory* area (also referred to as the front room) serves less critical patients with acuity levels 3 and 4. The hospital we are studying has a desired service level of serving the patients in ambulatory area within a specific time. Static doctor's schedules for both rooms are planned manually based on perceived understanding about the demand at various hours of a day, over the entire week or month. Such schedules generally do not react well to uncertainties such as surges in patient arrival. The hospital also reported its challenge to meet the desired service level. The doctors between two rooms are at times flexibly moved between the rooms in an ad-hoc and unplanned manner, without the ability to measure the impact on the service level measure such as the LOS of the patients in both rooms. If the ED experiences a surge in demand (usually in the front room), additional doctors may be deployed but there is lack of information about how long the additional doctors are required.

To address the challenges, we design dynamic resource allocation strategies with the aim to achieve the following objectives: (1) ability to respond to uncertainties (e.g., surges in demand) by leveraging on real-time data and effective use of simulation interacting directly with the physical systems to find optimal resource requirement; (2) ability to ensure that the quality of the related facility (such as the ED's back room) is fulfilled. Our proposed dynamic resource adjustment strategies are data-driven and are to provide invaluable real-time decision support to the ED operations. The term *strategy* is used interchangeably with *policy* in this paper.

We model the ED process as dual time-varying ($G_b(t)/M_b/S_b(t)$, $G_f(t)/M_f/S_f(t)$) queues for the back room (subscript b) and front room (subscript f) respectively. The front room is modeled as a queue with re-entrant patients. The patients who arrive at the hospital is required to first undergo a registration process and triage by a nurse before waiting at a queue to consult the doctor. As our aim is to study the doctor's requirements at the consultation, we model consultation service as our first service station (station 1 as shown in Figure 1). With a probability of b, the patient may be asked to do investigative tests or treatment at service station 2 (see Figure 1). Upon completion and test results are ready, patient re-enter the same queue at station 1 to be reviewed by the *same* doctor. After review, patient is either discharged or be admitted as an in-patient. More details about the process can be found in our previous work [1].

The scope of this paper is to provide dynamic resource allocation strategies for managing the queue to doctor (station 1). We present a model that produce heuristically-optimized doctor's schedule that better cope with demand surges in real-life dynamic situation. To our best knowledge, this is the first work to consider queue design and real-time queue control under time-varying demand with re-entrant customers for a dual-facility service center.

II. RELATED WORK

In this section, we review related works in three aspects. Firstly, the work on managing queue design and queue control; secondly, work on managing staffing for time-varying

*This research/project is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

[1] K.W. Tan, W.H. Tan and H.C Lau are with School of Information Systems, Singapore Management University, 80 Stamford Road, Singapore 178902. {kwtan, weihao.tan.2010, hclau} at smu.edu.sg

arrival with re-entrant customer (patient); and finally work on simulation approaches for handling dynamic ED processes.

For queue design and control that manages facilities with back and front rooms, Berman et al. [2] presents a scenario where the front room has shoppers and check-out stations and the back room has other indirect non-customer-facing work. They presented a staff-switching policy, where resources in both rooms are shared and can be switched to front room such that customers' wait-time in queue is within an acceptable value. They present a heuristic solution to find the optimal switching policy while fulfilling the back room minimum requirement. Terekhov et al. [3] builds on [2] by using a constraint-programming approach. They claim to be the first work in finding proven optimality of the switching problem. However, they remove the concept of shoppers and restricted it to a ”two-dimensional” queueing model. None of these works addresses the situation where the arrival is time-varying.

To address the staffing problem under time-varying arrivals, there are two major approaches. The first approach is suitable for short service-time processes. The methods include steady-state approximations such as in PSA (Piecewise Stationary Analysis), RCCP (Rough Cut Capacity Planning)[4], Stationary Independent Period by Period (SIPP), or lag-SIPP [5][6]. The second approach is applicable to long service-time processes. The methods are MOL (Modified Offered Load) [6] and ISA (Infinite Server Approximation) [7]. None of these addresses the re-entrance property of the ED process. Yom-Tov [8] considered both time-varying arrival and re-entrance process modeled using Erlang-R queue model. The method provides proactive (pre-planned) staffing plans, which is not ideal in handling uncertain demand. Time-varying arrival is handled using MOL approach and staffing policy is then based on square-root staffing formula.

The staff-switching problem under time-varying arrivals and re-entrant patients is a complex problem beyond analytical models. In this work, we model this problem as an optimization problem and solve it via a combination of simulation and heuristics. We refer to [8] that provides a benchmark and also an initial solution to our local search algorithm, which we then apply an improved solution for our problem. In addition, we propose the use of simulation in real-time to make prediction on future demands upon which our dynamic allocation policy is based. Using simulation to address some problems in ED is not new, many existing literature [9][10][11] used discrete-event simulation to enable complex problems to be analysed. In these work, the simulator is used offline as a tool to verify the proposed models but is not used in real-time for decision support. The work in [12] used a simulator as a tool to test which staffing method is suitable in real-time. Another idea is use of *symbiotic simulation* [13][14] where a simulator is used in real-time interacting with the physical systems for decision-support. Yet another related work is concerned with the optimization of an objective function using simulation, generally termed *simulation optimization* (For details, we refer the reader to the comprehensive survey on simulation optimization in [15].) In our approach, offline simulation is used to validate and evaluate the staffing strategies and symbiotic simulation with *simulation optimization* is used in one of the strategies to provide real-time staffing plans.

III. BACKGROUND

We will be adopting some findings from [8] in our resource allocation strategies. In [8], the goal is to identify the staffing required at the ED that provide stable performance as measured by the probability of wait $(P(W > 0))$, i.e., regardless of the time of arrival of the patient, the probability of waiting in the queue remains relatively constant. This is the QED (Quality and Efficiency Driven) balance that it is trying to achieve. Based on the MOL approach, the time-varying offered load R_i at station i [16] is given by

$$R_i(t) = E[\lambda_i^+(t - S_{i,e})]E[S_i] \tag{1}$$

where λ_i^+ is the aggregated-arrival-rate function to node i, S_i represents the service time at node i and $S_{i,e}$ is a random variable representing the excess service time at node i.

The doctor's staffing $s_1(t)$ at station 1 at time t is determined by substituting the time-varying offered load formula into the square-root staffing formula [6][7] where β is chosen according to the steady-state Halfin-Whitt formula [17]:

$$s_1(t) = R_1(t) + \beta\sqrt{R_1(t)}; \quad \forall t > 0 \tag{2}$$

Based on our data analysis using 6-months data from the local hospital, the arrival rates function is shown in Figure 2. We observe that the arrival rates for Sundays and Mondays are similar and higher while arrival rates for Tuesday to Saturday are similar and lower than the former. The arrivals are observed to be non-homogeneous Poisson processes. For the purpose of comparison of our proposed strategies with existing work with an analytical model, the service rates at stations 1 and 2 are simplified to single exponential distributions of μ and δ respectively. Based on [8], we are to solve the following ordinary differential equations (O.D.E) via numerical approximation for general rate of arrival and exponential service rates.

$$\begin{aligned} \frac{d}{dt}R_1(t) &= \lambda_t + \delta R_2(t) - \mu R_1(t) \\ \frac{d}{dt}R_2(t) &= p\mu R_1(t) - \delta R_2(t) \end{aligned} \tag{3}$$

IV. THE MODEL

Figure 1 shows the overview of our queueing model. Our aim is to intelligently determine or forecast the optimal number of doctors required at station 1 (in the front room of the ED) in order to meet the front room service quality as desired by the hospital while maintaining the back room quality. We focus on station 1 because we learnt that it is the bottleneck of the ED process (through an on-site survey). We measure the performance of our strategies using the average length-of-stay (LOS) of the patients. In our study, the LOS of a patient is the duration that a patient spends at the ED from registration to end of consultation or end of review consultation if investigative tests or treatment are required. With reference to Figure 1, we define the following notations used in our model:

- $\lambda_b(t)$ - The time-varying arrival rate of new patient at the back room. Each $\lambda_b(t)$ is defined per hourly over a week's horizon.
- $\lambda_f(t)$ - The time-varying arrival rate of new patient at the front room as shown in Figure 2. Each $\lambda_f(t)$ is defined per hourly over a week's horizon.
- $\overline{LOS}(t)$ - The average LOS of the patients who leave the ED within the time interval of $[t, t+1)$.
- LOS_{max} - Hospital's desired service quality in terms of LOS
- μ - The service rate of the doctors at station 1. We assume a homogeneous service rate for all doctors.
- δ - The service rate for investigative tests or treatment at station 2.
- $room_{max}$ - The physical constraint in the ED at the front room. This corresponds to the maximum number of consultation rooms in the real-life set up of the ED.
- $S_{max}(t)$ - The maximum number of doctors that can be deployed at ED (both front and back rooms combined) at time t. This corresponds to the resource capacity in the real-life.
- $S_b(t)$ - The number of doctors required at the back room to serve the demand the back room. The QED regime as per equation 2 is used to ensure that most patients (critically ill patients) do not have to wait for a doctor at the back room.
- $S_f(t)$ - The number of doctors to be placed at the front room.
- b - The probability that a patient will go for treatment/tests.

The following are the constraints for our problem:

$$S_f(t) \le S_{max}(t) - S_b(t) \tag{4}$$

$$S_f(t) \le room_{max} \tag{5}$$

$$\overline{LOS}(t) \le LOS_{max} \quad (soft \quad constraint) \tag{6}$$

Constraint (4) specifies the back room service level guarantee, (5) specifies the physical front room constraint and (6) specifies the service level constraint.

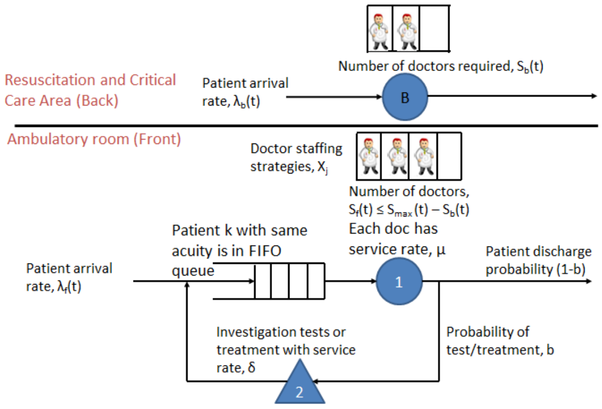

Number of doctors required, $S_b(t)$
Resuscitation and Critical Care Area (Back)
Patient arrival rate, $\lambda_b(t)$
B
Ambulatory room (Front)
Doctor staffing strategies, X_j
Number of doctors, $S_f(t) \le S_{max}(t) - S_b(t)$
Each doc has service rate, μ
Patient k with same acuity is in FIFO queue
Patient arrival rate, $\lambda_f(t)$
1
Patient discharge probability (1-b)
Investigation tests or treatment with service rate, δ
Probability of test/treatment, b
2

Fig. 1. The queueing model

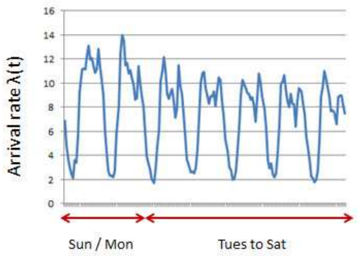


Fig. 2. Our Time-Varying Arrival to ED

V. Resource allocation strategies

The resource allocation strategies in which doctor's schedule in the front room for station 1 can be determined either in a proactive manner or dynamically depending on the actual arrival conditions. The doctor's schedule is designed to adjust at the minimum of an hour interval. This is to prevent a schedule that is too nervous with many changes within a small time-interval. We also assume that the back room requirement is fairly predictable based on historical data and hence back room staffing is pre-processed and is available before allocation of doctors to the front room.

A. Strategies using Square-root Staffing Rule

1) Strategy HIST - Resource allocation using historical trends: The HIST strategy is a proactive strategy that considers only historical trends taken from the actual data of a real-world hospital. This method is based on findings in [8] and is used as the benchmark for our other strategies. We apply the Erlang-R method as presented in Section III for arrival to both the back room and front room. We first solve the ODE (Equation 3) and apply square-root staffing rule (Equation 2) using numerical approximation. The results from staffing rule for front room requirements are subjected to the *back room service level guarantee* and *physical front room constraint*. Without the back room and physical constraints, this is proven to provide stable queue wait-time (not LOS) over time as shown in [8]. However, with the addition of the constraints, the number of doctors that can be assigned to the front room may be less than the output from the staffing rule and the wait-time stability can no longer be guaranteed. This provides the opportunity to develop better strategies.

2) Strategy DYN - Dynamic resource allocation using real-time data: With the support of enterprise systems and real-time monitoring devices today, it is possible to track the time-varying arrival rates in real-time. With these arrival rates, we can then calculate the *actual* time-varying offered load for the previous time intervals. One challenge to apply this method is that the arrival rates must be recorded in a significantly large time frame such as hourly. However, to approximate the offered loads, one must use a small time-interval such as per minute time frame in order to compute reasonably-sensible offered load for the next time period.

The DYN strategy is designed to be reactive to real-time observation of the arrivals and perform short-term forecast (such as hourly basis). Let $R_1'(t)$ and $R_2'(t)$ be the real-time offered load based on *actual* real-time arrival rates at time

t. The doctors' requirement $S_f(t+1)$ for the next hour is determined by square-root staffing rule $S_f(t+1) = R_1(t+1) + \beta\sqrt{R_1(t+1)}$, where $R_1(t+1) = R_1'(t) + \frac{d}{dt}R_1'(t)$. With considerations of constraints 4 and 5, $S_f(t+1)$ is set to $min[S_f(t+1), S_{max}(t+1) - S_b(t+1), room_{max}]$.

B. Strategies using Heuristic Optimization Model with Symbiotic Simulation

By far, both HIST and DYN strategies rely heavily on the staffing rule. One major limitation with the use of staffing rule is that service level quality constraint may be violated when we conform to the back room quality constraint and physical front room constraint as the number of doctors deployed at the front room is limited to the $min[S_{max}(t) - S_b(t+1), room_{max}]$. To remove or minimize the violations, an intuition is that one could allocate more doctors in the previous hour(s) before violation occurs; or to allocate more doctors later to clear the queue. We propose our next 2 strategies (namely HIST-OPT and DYN-OPT) that will use heuristic methods to find optimized resource allocation.

1) Strategy HIST-OPT - Optimized resource allocation using historical trends: The idea for HIST-OPT strategy is to provide an option for an optimized resource allocation by using analytics data in the case that real-time is not available. Let C_l be the cost of labour of deploying a doctor for a single unit time t and C_d be the cost if the number doctors in time t deviates from $t-1$. The deviation cost is included in the model is for stability of the schedule as we do not wish to have a schedule where the number of doctors changes too frequently from hour to hour. We have the following objective function that is subjected to constraints (4) to (6).

$$min \quad C_l \sum_t S_f(t) + C_d \sum_t [S_f(t) - S_f(t-1)]^2 \qquad (7)$$

We deploy a local search algorithm to search for a schedule. The search algorithm finds a time t where service quality constraint is violated. In an attempt to remove the violation, the algorithm searches for a time t_1 nearest to t when $(0 \leq t_1 < t)$ and constraints (4) and (5) are not violated and increase resource by 1 unit at t_1. Next, it finds a time t_2 nearest to t when ($t \leq t_2 \leq$ length of simulation) and constraints (4) and (5) are not violated and increase resource by 1 unit at t_2. The algorithm selects a schedule that meets the constraints or one with a lower cost among the 2 solutions and repeats the search until either there is no violation at t or no further solution for the violation at t. The search then moves on to the next violation. This search will terminate when the first schedule without a violation is found or when maximum number of searches has been reached. If latter, an infeasible schedule with the least violation is returned.

2) Strategy DYN-OPT - Optimized Dynamic resource allocation using real-time data: The design of DYN-OPT is to attempt to address the limitations of DYN which has short planning time (hence being too nervous to react to changes every hour) and HIST-OPT which is not unable to react to demand changes. The intuition behind DYN-OPT is to make use of the known real-time information as well as the forecast to the future based on historical information about the arrival. The aim here is to plan a number of hours ahead such that resources in the ED can be better informed or additional resources can be better arranged.

We define a vector (L, H), where L denotes the lead-time and H denotes the time horizon. Between the current time t and lead-time L, there will be no change in staffing. This is to address the fact that most ED cannot add or remove a doctor instantaneously. The variable H defines the horizon of re-planning, e.g., plan for the horizon of 8 hours. The next planning period is then $t+H$. Our optimization model becomes:

$$min \quad C_l \sum_{t=t+L}^{t+L+H} S(t) + C_d \sum_{t=t+L}^{t+L+H} [S(t) - S(t-1)]^2 \qquad (8)$$

We deploy a local search algorithm that is similar to HIST-OPT, with the exception that real-time arrival rates are used and planning is short-termed. We use the *actual* arrival rates to calculate $R_1'(t)$ and $R_2'(t)$ up to time t, then we use historical arrival rates to compute the staffing required for the period $(t+L, t+L+H)$. The optimization search algorithm to manage any violation within the planning horizon H is the same as HIST-OPT.

VI. Experimental Evaluation

A. Experimental Setup

We design the prototype of our model and named it as the Dynamic Queue Management (DQM). It is capable of receiving both the historical data and real-time data (from the live systems). The prototype is developed as an event-driven simulator to act as the physical systems in the actual deployment. We termed this as the *DQM simulator* where patient arrival, consultation, treatment and investigative test events are simulated using the resources as determined by the strategies. An optimization model is built within DQM simulator to perform the local searches. We also include a *symbiotic simulation system* in the DQM simulator to evaluate the performance of schedules produced by the optimization model. Both the DQM simulator and the symbiotic simulation system are implemented in Java$^{\text{TM}}$. In terms of implementation, HIST requires only historical data, DYN requires real-time data. HIST-OPT requires historical data and the optimization model, while DYN-OPT requires historical, real-time data and the symbiotic simulation system.

Each experiment is run over 100 replications and the result is an average over the replications. In each optimization for HIST-OPT and DYN-OPT, 50 replications are used and average is taken to evaluate the solution. The maximum search iteration is set to 300. Evaluation of DYN-OPT candidate solutions is done in the symbiotic simulation system. In DYN-OPT, the lead-time is set to 0 (plan for next hour) and the planning horizon is set to 8 hours. We derive the parameters $\lambda_b(t)$, $\lambda_f(t)$, μ, δ and b from historical data using a commercial business analytics software, SAS$^{®}$. The time-varying arrival rates ($\lambda_b(t)$, $\lambda_f(t)$) are recorded over a week, each day has its own time-varying arrival rates that change every hour, we refer to this as the *historical arrival rates* (see Figure 2). The probability of re-entrance b is found to be 0.4. The average service rates of doctors μ is 4 per hour. The registration and triage service time is assumed to be exponentially distributed with the mean of 14.2 minutes. The hospital's desired service quality, LOS_{max} is set to 60 minutes.

To determine the hourly back room minimum requirement. Since patients with high severity level must be attended almost immediately, we applied the MOL and square-root staffing under QED regime. The QED staffing will assure that on the average, the number of patient is less than the number of doctors and hence ensure minimal wait time. We used that to determine the maximum number of doctors possible at the front-room based on constraint as stated in Equation 4.

For each set of experiments (for all strategies), a simulation over 9 days is taken and the first and last day are discarded in order to remove the inaccurate results from simulation start-up and winding down. The remaining 7 days represents the 7 days of the week with different patterns of arrivals as observed in real-life. To simulate various real-life scenarios, we run the experiments under three conditions: Normal, Low and High load. The normal load condition assumes the arrival rates to be exactly the same as historical rates. The low load condition, the arrival rates of on the days Thursdays, Fridays, Saturdays and Sundays are halved. In the high load condition, the arrival rates of the same affected days are doubled.

B. Experimental Results

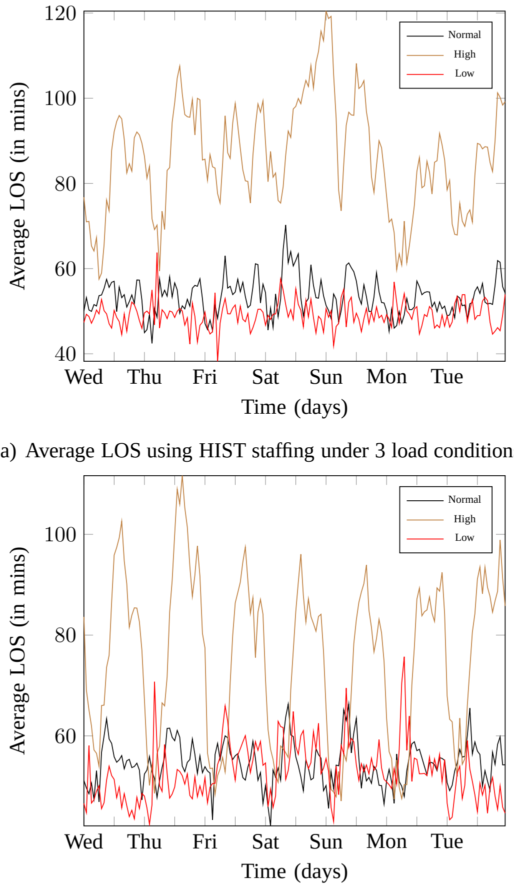
The first experiment is to compare HIST and DYN (which use the staffing rule) to illustrate the value of using real-time information. As shown in Figure 3, we observe that DYN is more capable of reacting to surges. The doctor staffing schedule generated by DYN is able to keep the average LOS stable under the conditions of low load and normal. Although the average LOS is longer under high load conditions, we can see the doctor's schedule generated by DYN allows the queue to clear. In the case of HIST, if the load is high, the average LOS remains higher than desired service quality over the period of demand surges (Thursday to Sunday).

Table I shows the average number of doctor's hours required to be deployed over a week under the given strategy and demand conditions. Supposed if an ED, over time, has equal probability of experiencing high load, normal load and low load, then the average cost of using a dynamic strategy (computed to be average of 650 doctor's hours) is lower than that of a static one which based on historical data (static at 681 doctor's hours).

HIST			DYN		
Normal	High	Low	Normal	High	Low
681	681	681	678	743	530

TABLE I

AVERAGE NUMBER OF DOCTOR'S HOURS PER WEEK



(a) Average LOS using HIST staffing under 3 load conditions

(b) Average LOS using DYN staffing under 3 load conditions

Fig. 3. Results of varying demand for the strategies using staffing rule

We can see that the staffing rule does provide reasonably good solutions. However, both HIST and DYN do not guarantee that the solution can satisfy the service level constraint as in Equation 6. We provide an example of use of the optimization model in DQM to find a solution that satisfies the service level constraint. Figure 4 shows that ED may be able to achieve a performance within the desired service level with the use of an optimization model such as HIST-OPT. This is observable under normal and low load conditions. The HIST-OPT strategy requires additional increase of only 8 doctor's hours per week compared to HIST. However, we need to understand that this performance is not guaranteed because a patient may require more time to be monitored and be treated with quality care in the real-life situation. The results in this figure also show that HIST-OPT is a static method and hence is not able to react to high load conditions.

Experimental results that appear to be counter-intuitive to us are the results of DYN-OPT. DYN-OPT, having the features of using real-time data as well as historical intuitively suggest that it will be one that is more flexible and react well to surges. However, we observed that the performance of DYN-OPT is not ideal. In particular, under the high-load test for demand surges over 4 days (Thursday to Sunday) as shown in Figure 5, we can see that at every x-axis ticks

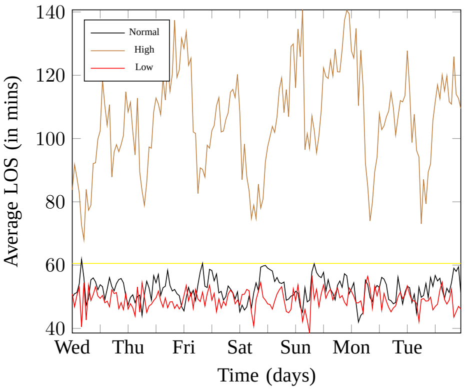


Fig. 4. Performance of HIST-OPT under 3 load conditions

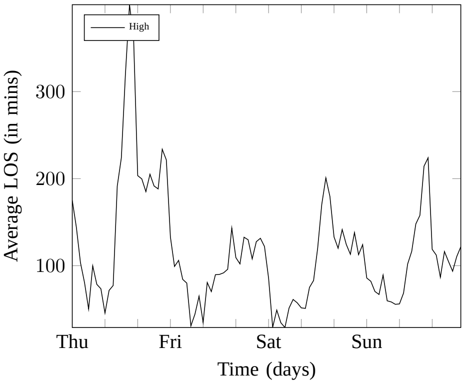


Fig. 5. Performance of DYN-OPT over the 4 days of demand surges in the high-load test

(which is the 8-hourly DYN-OPT's horizon), there is a drop in the average LOS at the start of each horizon because the method will indicate that more doctors are required to serve the patients under unexpected high load. The entire horizon is then planned based on historical arrival which is only half of the actual arrival. As such, the queue builds up again during this period when there exists no resource planning until the next horizon. Although, DYN-OPT is able to react to demand surges by lowering the LOS at the start of each horizon, we see that strategy DYN is better at handling surges. This is because DYN has the opportunity to adjust the resources at every hour. However, one may find DYN a strategy too reactive and hospital may not be able to find doctors within a short notice. As such, we see more opportunity to further improve DYN-OPT strategy or investigate how the DYN-OPT parameters such as horizon can affect or improve the performance. DYN-OPT is potentially a good short-term planning strategy as we recognize through DYN and HIST-OPT that there are benefits in using real-time information and an optimization method.

VII. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed several strategies for resource allocation at the ambulatory area of an ED, both proactive and dynamic methods, while maintaining the service quality at the back room. We found in our experiments that real-time information can be leveraged to manage demand surges or to release doctors to the backroom operations during low-peak period. In today's technology where real-time data is easily accessible (e.g., patient real-time movements to be tracked and data can be fed quickly into live systems), we believe that our proposed concept of dynamic resource allocation is implementable. Our simulation prototype with optimization model and symbiotic simulation system allows modeling and evaluation of intelligent staffing strategies in real-world complex ED process that provide results which are helpful to healthcare decision makers. Our strategies provide alternatives for the decision makers to select an effective method based on each hospital's appetite for performance, cost and implementation complexity.

REFERENCES

[1] K. W. Tan, C. Wang, and H. C. Lau, "Improving patient flow in emergency department through dynamic priority queue," *IEEE International Conference on Automation Science and Engineering*, 2012.

[2] O. Berman and R. C. Larson, "A queueing control model for retail services having back room operations and cross-trained workers," *Computers & Operations Research*, vol. 31, no. 2, pp. 201–222, Feb. 2004.

[3] Terekhov, Daria, Beck, and J. Christopher, "A constraint programming approach for solving a queueing control problem," *J. Artif. Int. Res.*, vol. 32, no. 1, pp. 123–167, May 2008.

[4] T. Vollmann, "Capacity planning: The missing link," *Production and Inventory Management (1st Qtr. 1973)*, pp. 61–74, 1973.

[5] L. V. Green, P. J. Kolesar, and W. Whitt, "Coping with time-varying demand when setting staffing requirements for a service system," *Production and Operations Management*, vol. 16, no. 1, pp. 13–39, 2009.

[6] O. B. Jennings, A. Mandelbaum, W. A. Massey, and W. Whitt, "Server staffing to meet time-varying demand," *Management Science*, vol. 42, no. 10, pp. 1383–1394, 1996.

[7] Z. Feldman, A. Mandelbaum, W. A. Massey, and W. Whitt, "Staffing of time-varying queues to achieve time-stable performance," *Management Science*, vol. 54, no. 2, pp. 324–338, 2008.

[8] G. B. Yom-Tov, "Queues in hospitals: Queueing networks with reentering customers in the qed regime," *The Technion - Israel Institute of Technology Ph.D. Thesis*, 2010.

[9] A. Komashie and A. Mousavi, "Modeling emergency departments using discrete event simulation techniques," *Winter Simulation Conference*, 2005.

[10] F. Pajouh and M. Kamath, "Applications of queueing models in hospitals," *Midwest Association for Information Systems*, 2010.

[11] W. Samaha, S.; Armel, "The use of simulation to reduce the length of stay in an emergency department." *Winter Simulation Conference*, 2003.

[12] S. Zeltyn, Y. N. Marmor, A. Mandelbaum, B. Carmeli, O. Greenshpan, Y. Mesika, S. Wasserkrug, P. Vortman, A. Shtub, T. Lauterman, D. Schwartz, K. Moskovitch, S. Tzafrir, and F. Basis, "Simulation-based models of emergency departments:: Operational, tactical, and strategic staffing," *ACM Trans. Model. Comput. Simul.*, vol. 21, no. 4, pp. 24:1–24:25, Sept. 2011.

[13] M. Y. Low, S. J. Turner, D. Ling, H. L. Peng, P. Lendermann, L. Chan, and S. Buckley, "Symbiotic simulation for business process re-engineering in high-tech manufacturing and service networks," in *Winter Simulation Conference*. IEEE, 2007, pp. 568–576.

[14] Y. N. Marmor, S. Wasserkrug, S. Zeltyn, Y. Mesika, O. Greenshpan, B. Carmeli, A. Shtub, and A. Mandelbaum, "Toward simulation-based real-time decision-support systems for emergency departments," in *Simulation Conference (WSC), Proceedings of the 2009 Winter*. IEEE, 2009, pp. 2042–2053.

[15] M. C. Fu, "Optimization for simulation: Theory vs. practice," *INFORMS Journal on Computing*, vol. 14, no. 3, pp. 192–215, 2002.

[16] W. Massey and W. Whitt, "Networks of infinite-server queues with nonstationary poisson input," *Queueing Systems*, vol. 13, pp. 183–250, 1993.

[17] S. Halfin and W. Whitt, "Heavy-traffic limits for queues with many exponential servers," *Operations research*, pp. 567–588, 1981.