Decision Support System Based on Distributed Simulation Optimization for Medical Resource Allocation in Emergency Department

Tzu-Li Chen

Department of Information Management, Fu-Jen Catholic University
510 Chung Cheng Rd , Hsinchuang, Taipei County 24205 Taiwan
chentzuli@gmail.com

Abstract. The number of emergency cases or people making emergency room visit has rapidly increased annually, leading to an imbalance in supply and demand, as well as long-term overcrowding of emergency departments (EDs) in hospitals. However, solutions targeting the increase of medical resources and improving patient needs are not practicable or feasible in the environment in Taiwan. Therefore, under the constraint of limited medical resources, EDs must optimize medical resources allocation to minimize the patient average length of stay (LOS) and medical resource wasted costs (MWCs). This study constructs a mathematical model for medical resource allocation of EDs, according to emergency flow or procedures. The proposed mathematical model is highly complex and difficult to solve because its performance value is stochastic and it considers both objectives simultaneously. Thus, this study postulates a multi-objective simulation optimization algorithm by integrating a non-dominated sorting genetic algorithm II (NSGA II) and multi-objective computing budget allocation (MOCBA), and constructs an ED simulation model to address the challenges of multi-objective medical resource allocation. Specifically, the NSGA II entails investigating plausible solutions for medical resource allocation, and the MOCBA involves identifying effective sets of feasible Pareto medical resource allocation solutions and effective allocation of simulation or computation budgets. Additionally, the discrete simulation model of EDs estimates the expected performance value. Furthermore, based on the concept of private cloud, this study presents a distributed simulation optimization framework to reduce simulation time and subsequently obtain simulation outcomes more rapidly. This framework assigns solutions to different virtual machines on separate computers to reduce simulation time, allowing rapid retrieval of simulation results and the collection of effective sets of optimal Pareto medical resource allocation solutions. Finally, this research constructs an ED simulation model based on the ED of a hospital in Taiwan, and determines the optimal ED resource allocation solution by using the simulation model and algorithm. The effectiveness and feasibility of this method are identified by conducting the experiment, and the experimental analysis proves that the proposed distributed simulation optimization framework can effectively reduce simulation time.

Keywords: Simulation optimization, Decision support, Non-dominated sorting genetic algorithm, Multi-objective computing budget allocation, Emergency department.

F.F.-H. Nah (Ed.): HCIB/HCII 2014, LNCS 8527, pp. 15–24, 2014.
© Springer International Publishing Switzerland 2014

1 Introduction

In recent years, Taiwan has gradually become an aging society. The continuous growth of the senior population annually accelerates the increase and growth rate in emergency department (ED) visits. According to statistics from the Department of Health, Executive Yuan, from 2000 to 2010, the overall number of people making emergency visits in 2000 was 6,184,031; the figure had surged rapidly to 7,229,437 in 2010, demonstrating a growth rate of approximately 16%.

People making emergency visits and the growth rate for these visits have risen rapidly in the past 11 years. Such an increase causes an imbalance between supply and demand, and ultimately creates long-term overcrowding in hospital EDs. This phenomenon is primarily caused by the sharp increase in patients (demand side), and the insufficient or non-corresponding increase in medical staffing (supply side). Consequently, medical staff capacity cannot accommodate excessive patient loads, compelling patients to wait long hours for medical procedures, thus contributing to long-term overcrowding in EDs.

The imbalance in supply and demand also prolongs patient length of stay (LOS) in the ED. According to data from the ED at Taiwan National University Hospital, Shin et al. (1999) found that, among 5,810 patients, approximately 3.6% (213 patients) had stayed over 72 hours in the ED. Of these 213 patients, some had waited for physicians or beds, whereas some had waited in the observation room until recovery or to be cleared of problems before being discharged. These issues frequently lead to long-term ED overcrowding. Based on data analysis of the case hospital examined in this research, among 43,748 patients, approximately 9% (3,883 patients) had stayed in the ED for over 12 hours, approximately 3% (1,295) had stayed over 24 hours, and approximately 1% (317 patients) had stayed in the ED for 72 hours.

Hoot and Aronky (2008) postulated three solutions to address the overcrowding of EDs: (1) Increase resources: solve supply deficiency by adding manpower, number of beds, equipment, and space. (2) Effective demand management: address problems of insufficient supply by implementing strategies, such as referrals to other departments, clinics, or hospitals. (3) Operational research: explore solutions to ED overcrowding by exploiting management skills and models developed in operational research. For instance, determining effective resource allocation solutions can improve the existing allocation methods and projects, ultimately enhancing ED efficiency, lowering patient waiting time, and alleviating ED overcrowding.

Among the previously mentioned solutions, the first solution is not attainable in Taiwan, because most hospital EDs have predetermined and fixed manpower, budget, and space; hence, resources cannot be expanded to resolve the problem. The second solution is not legally permitted in Taiwan, and is essentially not applicable. Both of the preceding solutions are seemingly inappropriate and not applicable; therefore, this study adopted the third solution, which entailed constructing an emergency flow simulation model by conducting operational research. Additionally, the simulation optimization algorithm was used to identify the optimal medical resource allocation solution under the constraint of limited medical resources to attain minimal average patient LOS and minimal MWC, subsequently ameliorating ED overcrowding.

The main purpose of this research was to determine a multi-objective simulation optimization algorithm that combines a non-dominated sorting genetic algorithm II

(NSGA II) and a multi-objective optimal computing budget allocation (MOCBA). An additional purpose was to conduct simulations of schemes and solutions by applying an ED discrete event simulation (DES) model produced using simulation software to obtain optimal resource allocation solutions.

In actual solution or scheme simulations, an enormous amount of simulation time is required to perform a large quantity of solution simulations. Therefore, a distributed simulation framework is necessary to save simulation time. This study adopted the concept of “private cloud,” and used the distributed simulation optimization framework to implement and solve this multi-objective emergency medical resource optimal allocation problem. The operation of this distributed simulation optimization framework can be categorized into two main areas: a multi-objective simulation optimization algorithm and a simulation model. During implementation and operation, NSGA II is first used to search feasible solutions and schemes. The simulation model is then used to simulate, obtain, and evaluate performance values, whereas MOCBA determines simulation frequency for the solution or scheme during simulation. For the simulation model, this study adopted a distributed framework, in which multiple virtual machines (VMs) are installed on separate computers. For solution or scheme allocation, single control logic is used to assign various resource allocation solutions to simulation models for different VMs to conduct simulation. Performance values are generated and returned after the simulation is complete. This framework is characterized by its use of distributed simulation to rapidly obtain performance values and reduce simulation time.

2 Medical Resource Allocation Model in Emergency Department

2.1 The Interfaces with Associated Tools

This study was based on the ED flow of a certain hospital as a research target. It has been established that patient arrival interval times and service times of each medical service obey specific stochastic distributions; each type of medical resource (such as staff, equipment, and emergency beds), and the presumed resource allocation at any time is deterministic or fixed and does not change dynamically according to time. Under these pre-established conditions, a multi-objective emergency medical resources optimization allocation problem in which the primary goals were minimal average LOS and minimal average MWC was sought. Under restricted medical resources, this study aimed to obtain the most viable solution for emergency medical resource allocation.

Index:

i :Index for staff type ($i = 1,\dots,I$), such as doctor and nurse etc.

j :Index for working area ($j = 1,\dots,J$), such as registration area, emergency and critical care area, treatment area and fever area etc.

k :Index of medical resources type ($k = 1,\ldots,K$), such as X-Ray machines, computer tomography (CT) machines, and lab technicians and hospital beds etc.

Parameters:

c_{ij} :Labor cost of staff type *i* in the working area *j*

c_k :Cost of medical resource type *k*

l_{ij} :Minimum number of staff type *i* in the working area *j*

l_k :Minimum number of medical resource type *k*

u_i :Maximum number of staff type *i*

u_k :Maximum number medical resource type *k*

Decision Variables:

X_{ij} :Number of staff type *i* in working area *j*

$\mathbf{X}$:Matrix of number of all staff types in all working area, $\mathbf{X} = \left(X_{ij}\right)_{I \times J}$

Y_k :Number of medical resource type k

$\mathbf{Y}$ Matrix of number of all medical resource types, $\mathbf{Y} = \left(Y_k\right)_K$

Stochastic medical resource allocation model:

$$\min \ f_1(\mathbf{X},\mathbf{Y}) = E[LOS(\mathbf{X},\mathbf{Y};\omega)] \tag{1}$$

$$\min \ f_2(\mathbf{X},\mathbf{Y}) = E[MWC(\mathbf{X},\mathbf{Y};\omega)] \tag{2}$$

Subject to

$$l_{ij} \le X_{ij} \qquad \forall i,\ j \tag{3}$$

$$l_k \le Y_k \qquad \forall k \tag{4}$$

$$\sum_j X_{ij} \le u_i \quad \forall i \tag{5}$$

$$Y_k \le u_k \qquad \forall k \tag{6}$$

$$X_{ij} \geq 0 \text{ and integer} \qquad \forall i,\ j \tag{7}$$

$$Y_k \geq 0 \text{ and integer} \qquad \forall k \tag{8}$$

Explanations of these mathematical models are as follows: Equation (1) is minimal expected patient LOS, where ω stands for the stochastic effect; Equation (2) is minimal average MWC, where ω stands for the stochastic effect. There are two levels of significance for minimal average MWC: (a) maximized resource use rate; and (b) minimized medical resource cost; Equation (3) is number of physicians and nurses in each area, which must exceed the lower limit; Equation (4) is the number of X-rays, CTs, laboratory technicians, and beds in the ED, which— must exceed the lower limit; Equation (5) is the sum of the number of physicians and nurses in each area, which must not exceed the upper limit; Equation (6) is the number of X-rays, CTs, and laboratory technicians, beds in the ED, which must not exceed the upper limit; Equation (7) is the number of physicians and nurses in each area, which must be greater than 0 and expressed as a whole number; and Equation (8) is the number of X-rays, CTs, and laboratory technicians, and beds in the ED, which must be greater than 0 and expressed as whole numbers.

3 Multi-objective Simulation Optimization

Multi-objective medical resource allocation is a stochastic optimization problem, and the ED system shows a stochastic effect. Therefore, to obtain the expected patient LOS and the expected rate of waste of each resource, the ED simulation model and the repetition of simulation are required to obtain the estimation value. However, determining the frequency of simulation repetition during the process of simulation is crucial. Excess simulation repetition improves the accuracy of the objective values, but consumes large amounts of computation resources. Therefore, this research suggests a multi-objective simulation optimization algorithm, incorporating NSGA II and MOCBA, to address the multi-objective ED resource allocation problem. The NSGA II algorithm, multi-objective population-based search algorithm, is used to identify the optimal and efficient Pareto set collected from the non-dominated medical resource allocation solutions through the evolutionary processes. However, to estimate the fitness of each chromosome (medical resource allocation solution) precisely, NSGA II needs a large number of simulation replications within the stochastic ED simulation model to find the non-dominated solution set. Moreover, the simulation replications are identical for all candidate design chromosomes to cause high simulation costs and huge computational resources. Therefore, to improve simulation efficiency, the MOCBA algorithm, new multi-objective R&S method, developed from Lee et al. (2010) is applied to reduce total simulation replications and efficiently allocate simulation replications or computation budgets for evaluating the solution quality of all candidate chromosomes to identify and select the promising non-dominated Pareto set. The algorithmic procedure for integrating NSGA II and MOCBA is demonstrated in Figure 1.

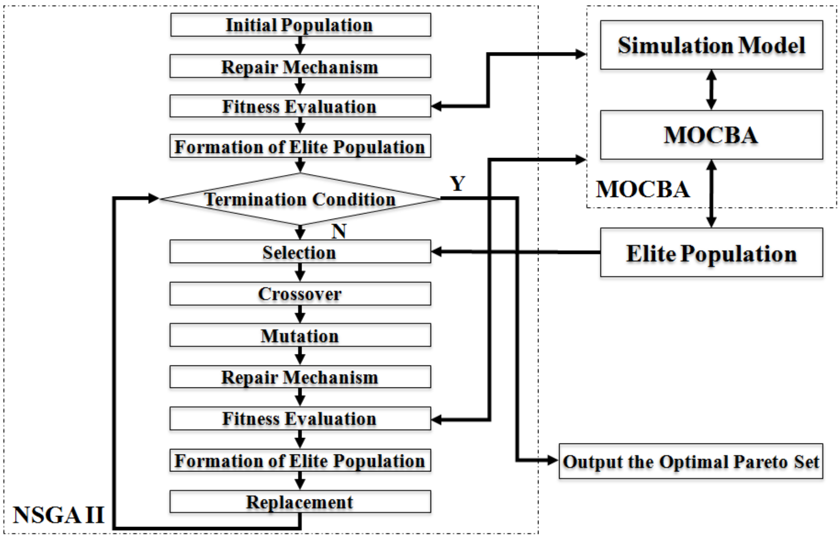

Fig. 1. The flow chart of integrating NSGA II and MOCBA algorithm

4 Distributed Simulation Optimization Framework

This study used eM-Plant 8.1 as a tool for developing the ED flow simulation model. Figure 2 illustrates the overall ED flow simulation model. In addition, a framework of distributed simulation optimization is developed to reduce the computation time by the private cloud technology. In this framework, we initially installed Microsoft

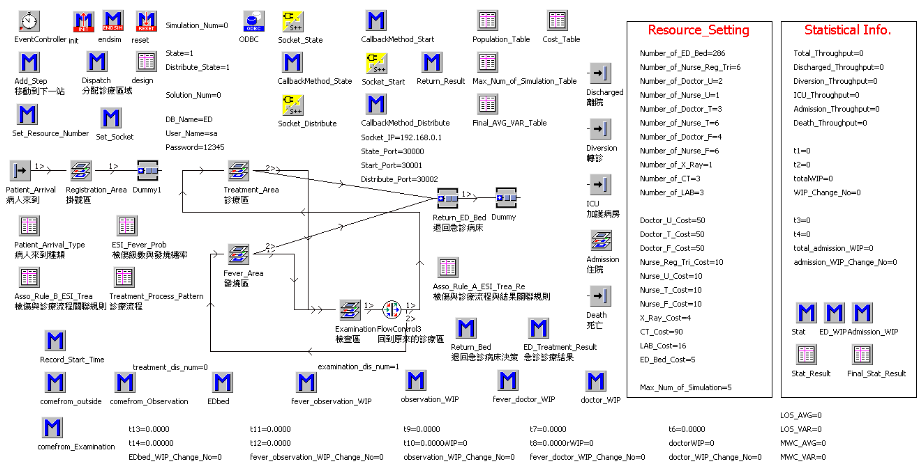

Fig. 2. Simulation model of emergency department flow

Hyper-V, a virtual operating system, on several actual servers to form a computer resource pool concept. We then established numerous virtual machines (VM) in this resource pool and assigned 1 simulation model to each VM. Emergency department procedures were subsequently simulated using these simulation models.

The distributed simulation optimization framework in Figure 3 comprised a client and a server. After the initial client parameters were set, Web services (WS) were employed to obtain the non-dominated sorting genetic algorithm-II (NSGA II) from the server via the Internet. These parameters were subsequently transferred to the NSGA II's WS. Upon receiving the HTTP request and parameter settings, the NSGA II conducts algorithmic procedures, calling WS for the multi-objective optimal budget allocation (MOCBA) algorithm when simulation is required. The MOCBA determines the number of simulation iterations required, calling WS for the simulation coordinator while simultaneously uploading the relevant simulation programs into the database. The SC's WS manages the simulation models, identifies the idle simulation models, and distributes simulation programs to the idle models to perform simulations. After identifying which model to simulate, the SC's WS commands the model to retrieve the simulation program from the database. Consequently, the simulation results are transferred to the SC's WS, which then transfers this data to the MOCBA to determine the simulation iterations required untill achieving the termination conditions. After the MOCBA is terminated, the performance results are transferred to the NSGA II's WS to again achieve the termination conditions. Following the termination of the NSGA II, the optimal program produced is transferred to the client-end.

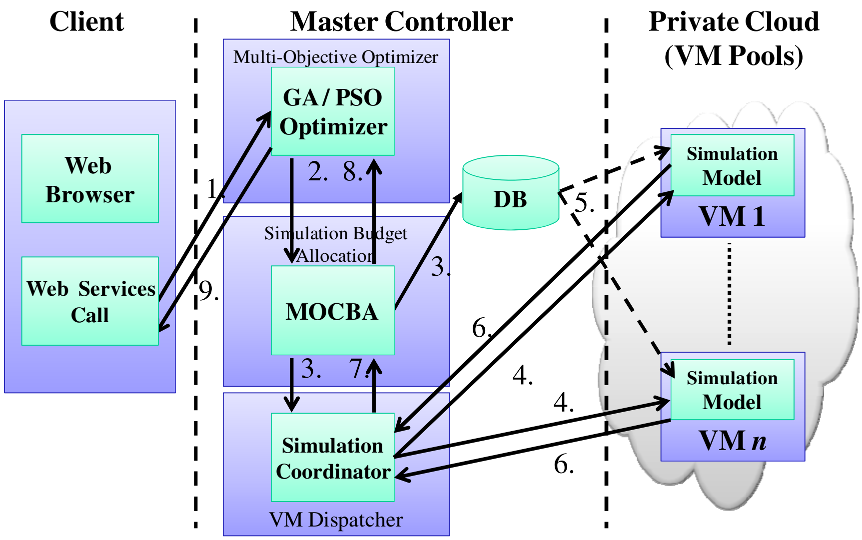

Fig. 3. Distributed simulation optimization framework

This framework was executed in the following process:

Step 1: The client calls WS for the NSGA II and transfers the parameters set by the user to the NSGA II's WS via the Internet.

Step 2: When simulation is required, the NSGA II calls WS for the MOCBA and transfers the simulation programs to the MOCBA's WS.

Step 3: When performance values are required, the MOCBA uploads the required simulation programs to the database via ADO.NET and calls WS for the SC to determine which VM simulation model to simulate.

Step 4: The SC's WS uses sockets to identify which VM is available and command the simulation model on the VM to perform a simulation.

Step 5: The simulation model uses open data connectivity to collect the simulation program data from the database after receiving the execution command from the coordinator socket.

Step 6: After executing the simulation program, the performance values are transferred to the SC's WS via the socket.

Step 7: The SC's WS transfers the performance results to the MOCBA's WS after receiving them from the simulation model.

Step 8: After receiving the performance values, the MOCBA's WS executes the MOCBA until the termination conditions are achieved. Subsequently, the performance results for algorithm termination are transferred to the NSFA II's WS.

Step 9: After receiving the performance values, the NSGA II's WS executes the NSGA II until the termination conditions are achieved. Subsequently, the produced results are transferred to the client via the Internet.

5 Experimental Analysis for the Distributed Simulation Optimization Framework

In this experiment, we primarily compared the simulation times for varying numbers of VMs to identify the differences when applying the proposed distribution simulation optimization model and the effects that the number of VMs had on the simulation times. In addition, this experiment analyzed the differences in simulation times for various allocation strategies with equal numbers of VMs.

We adopted the integrated NSGA II_MOCBA as the experimental algorithm, and employed the optimal NSGA II parameter settings determined in the previous experiments. The parameter settings were as follows: generation = 10, population size = 40, C = .7, M = .3, and the termination condition = generation (10).

The initial number of simulation iterations for the MOCBA was $n_0 = 5$, with a possible increase of $\Delta = 30$, and $P^*\{CS\} = 0.95$ for every iteration.

Regarding the number of VMs, we conducted experiments using 1, 6, 12, and 18 VMs. Table1 shows the execution times for the simulation programs with varying numbers of VMs and allocation strategies. Besides 1 VM, two methods can be used for allocating the remaining numbers of VM, specifically, including and excluding allocation of the number of simulation iterations. Excluding the allocation indicates that the simulation program is allocated to 1 VM for execution regardless of the program's number of simulation iterations, that is, the number of iterations for that program is not divided and allocated to separate VMs. Conversely, including the allocation indicates that when the number of iterations for the simulation program exceeds the initial number of iterations n_0 set by the MOCBA, the number of iterations is divided and allocated to numerous VMs for execution.

Table 1. The execution times for the simulation programs with varying numbers of VMs and allocation strategies

Number of VMs	**Allocation method**	**Number of executions**	**Execution times**
1	-	4200 executions	690.5 h (28.77 d)
6	Excluding number of runs allocation	4260 executions	112 h (4.67 d)
	Including number of runs allocation	4260 executions	105.5 h (4.40 d)
12	Excluding number of runs allocation	4290 executions	58 h (2.42 d)
	Including number of runs allocation	4230 executions	52 h (2.17 d)
18	Excluding number of runs allocation	4380 executions	52 h (2.17 d)
	Including number of runs allocation	4350 executions	40 h (1.67 d)

According to the experimental results shown in Table 1, we determined the following insights:

1. The overall execution time for 1 VM approximated a month (28 d). However, the execution time was reduced significantly to approximately 4 and 1.5 days when the number of VMs was increased to 6 and 18, respectively (Table 1). In addition, the curve exhibited a significant decline from 1 VM to 18 VMs. Thus, we can confirm from these results that the proposed distributed simulation optimization framework can effectively reduce simulation times.
2. The overall execution time was reduced from approximately 4 days to 1 day when the number of VMs increased from 6 to 18 (Table 1). In addition, the curve exhibited a decline from 6 VMs to 18 VMs. These results indicate that the simulation times can be reduced by increasing the number of VMs.
3. With a fixed number of VMs, the time required to divide and allocate simulation iterations to numerous VMs is shorter than that for allocating the entire number of iterations to 1 VM (**Error! Reference source not found.**1). Considering 6 VMs as an example, the execution time without dividing and allocating the number of simulation times was 112 h, whereas the execution time with dividing and allocating the number of iterations was 105.5 h. These results indicate that distributing the

number of simulation times among numerous VMs can reduce the overall execution time.

4. According to the experimental results, we infer that a limit exists when the number of VMs is increased to significantly reduce the simulation times. In other words, when a specific number of VMs is added to a low number of available VMs, the simulation time is significantly reduced. However, when the number of VMs increases to a specific amount, the reduction in simulation time becomes less significant, eventually reaching convergence. This indicates that after a certain number of VMs, the simulation time dos not decline with additional VMs.

6 Conclusion

This study investigated the resolution of ED overcrowding through ED medical resource optimal allocation. First, an emergency simulation model for a hospital in Taiwan was designed based on interviews and analysis regarding procedures and flow. A multi-objective simulation optimization algorithm was then designed by integrating the NSGAII algorithm and the MOCBA. To obtain simulation outcomes more rapidly by diminishing simulation time, this study proposes a distributed simulation optimization framework based on the private cloud concept to practice or implement and resolve this multi-objective emergency medical resource optimization allocation problem. In the proposed distributed simulation optimization framework, solutions or schemes are assigned to different VMs on separate computers to conduct simulations and minimize simulation time, as well as obtain simulation results more rapidly.

References

1. Ahmed, M.A., Alkhamis, T.M.: Simulation optimization for an ED healthcare unit in Kuwait. European Journal of Operational Research 198, 936–942 (2009)
2. Chen, C.H., Lee, L.H.: Stochastic simulation optimization: An Optimal Computing Budget Allocation. World Scientific Publishing Co. (2010)
3. Hoot, N.R., Aronsky, D.: Systematic Review of ED Crowding: Causes, Effects, and Solutions. Health Policy and Clinical Practice 52(2), 126–136 (2008)
4. Lee, L.H., Chew, E.P., Teng, S., Goldsman, D.: Finding the non-dominated Pareto set for multi-objective simulation models. IIE Transactions 42(9), 656–674 (2010)
5. Pitombeira Neto, A.R., Gonçalves Filho, E.V.: A simulation-based evolutionary multiobjective approach to manufacturing cell formation. Computers & Industrial Engineering 59, 64–74 (2010)