



## WestminsterResearch

<http://www.wmin.ac.uk/westminsterresearch>

### **A semi-open queueing network approach to the analysis of patient flow in healthcare systems.**

**Haifeng Xie**  
**Thierry Chausalet**  
**Mike Rees**

School of Informatics

Copyright © [2007] IEEE. Reprinted from the proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems. IEEE CBMS 2007, Maribor, Slovenia, 20-22 June 2007. IEEE, Los Alamitos, USA, pp. 719-724. ISBN 0769529054.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Westminster's products or services. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

---

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners. Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

---

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of the University of Westminster Eprints (<http://www.wmin.ac.uk/westminsterresearch>).

In case of abuse or copyright appearing without permission e-mail [wattsn@wmin.ac.uk](mailto:wattsn@wmin.ac.uk).

# A Semi-open Queueing Network Approach to the Analysis of Patient Flow in Healthcare Systems

Haifeng Xie, Thierry Chausalet, Mike Rees  
Health and Social Care Modelling Group  
School of Informatics  
University of Westminster, London, UK  
{xieh, chausst, reesm}@wmin.ac.uk

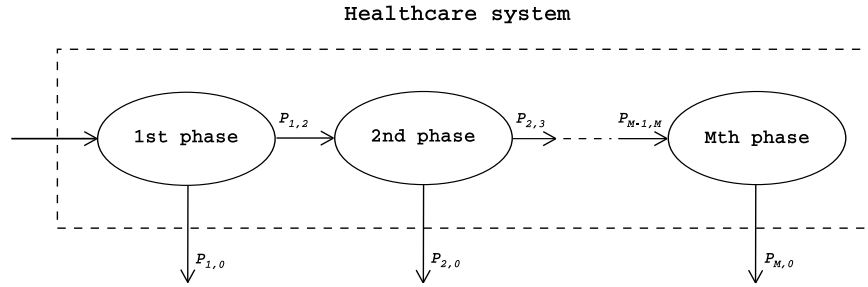
## Abstract

*In this paper, we present a modelling framework for patient flow in a healthcare system using semi-open queueing network models, which introduces a total bed constraint, above which new patients will be refused admission. Hence this model provides a realistic representation of a real system. This approach enables us to have access to a range of established methods that deals with queueing network models. We demonstrate the usefulness of the model in the context of a geriatric department and show that hospital managers can use this model to gain better understanding of the dynamics of patient flow and to study potential long-term impacts of policy changes.*

## 1 Introduction

Patient flow has been recognised as a major factor in improving the efficiency of a healthcare system [1]. In a healthcare system, patients can be regarded as progressing through a set of logical stages that resemble the process of care, such as assessment, diagnosis, rehabilitation and long-stay care, etc [2, 3]. Adopting this view, compartmental models have been developed to explain patient flow through a geriatric department [2]. These models show different phases of care interact to influence admission and discharge of patients in a department of geriatric medicine. Most studies focus on deriving steady-state solutions for systems without bed capacity constraints. That is a system assuming to have as many beds available as required at any time. Clearly this assumption is generally untrue in reality. Therefore, an introduction of bed capacity constraint in such models will provide a more realistic representation of the real world situation.

In a previous study [5], we introduced a modelling framework for patient flow in a healthcare system using a closed queueing network model [6]. This approach enabled us to have access to a range of established methods that deals with queueing network models. It also provides a natural platform for incorporating a capacity constraint on the system, e.g. total number of beds available in a hospital department. In this paper, we present an extension to this approach, which relaxes the assumption in our previous study [5] that the system is always full. This approach, which is based on semi-open queueing network models, introduces a total bed constraint (above which new patients will be refused admission), which relaxes the assumption that the system is always full (i.e. a closed network) and hence provides a more realistic representation of a real system. We show that our results are applicable to both the closed network model and open network model (as discussed by Irvine *et al* [4]) since they are simply special cases of the semi-open network model.



**Figure 1. Health care system modelled with  $M$  phases.**  $p_{i,i+1}$  ( $i = 1, \dots, M - 1$ ) represent the transfer probabilities and  $p_{i,0}$  ( $i = 1, \dots, M$ ) the discharge probabilities. Subscript 0 represents the state of being outside the healthcare system.

In the following section, we introduce the queueing network approach to the analysis of the healthcare system. In Section 3, we derive the corresponding performance measures, such as loss probability of the system and expected occupancy level. In Section 4, a case study is presented to illustrate the application of the proposed model to a hospital geriatric department. Conclusions and directions for further research are given in Section 5.

## 2 Queueing network models for healthcare systems

The progression of patients in a healthcare system is modelled as consisting of a set of conceptual phases. On admission, patients enter the 1st phase, which could correspond to assessment, diagnosis, etc. After staying a period of time, they are either discharged (including by death), or transferred into the 2nd phase, which could be some form of rehabilitation. The process continues until at the final phase all patients are eventually discharged. A diagrammatic representation of the system with  $M$  phases is shown in Fig. 1. The transfer probability  $p_{i,i+1}$  ( $i = 1, \dots, M - 1$ ) represents the probability that a patient will move to phase  $i + 1$  upon completion of phase  $i$ ; and  $p_{i,0}$  ( $i = 1, \dots, M$ ) is the discharge probability upon leaving phase  $i$ . We assume new patients arrive at a constant rate  $\lambda$  per unit time (taken to be days without loss of generality). Further, we assume that there are a total of  $K$  beds in the system. That is there are at most  $K$  patients in the system at any one time. Under this setting, the system will admit patients for as long as there is available beds, and patients will be turned away when there is no empty beds left in the system. A service node of the queueing network is one of the phases. The servers at a service node are all hospital beds providing the corresponding phase of care at the time. The number of servers at each service node varies with time depending on the dynamic use of each bed, as patients move through care phases.

More specifically, the number of servers at node  $i$ , denoted by random variable  $L_i$  ( $i = 1, \dots, M$ ), varies between 0 and  $K$  and satisfies the condition  $\sum_{i=1}^M L_i \leq K$ . We further assume that, for a server of service node  $i$ , its service time follows an exponential distribution with mean  $1/\mu_i$ , where  $\mu_i$  is the constant service rate per bed, i.e. the average number of patients treated per day per bed.

We solve the homogeneous linear system  $e_i = \sum_{j=0}^M e_j p_{j,i}$  ( $i = 0, 1, \dots, M$ ), for the relative arrival rates  $e_i$  ( $i = 1, \dots, M$ ) to within a multiplicative constant [6]. The relative arrival rates are the effective rates of patients entering each phase per day. Assuming  $e_0 = 1$ , the relative arrival

rate  $e_i$  becomes

$$e_i = \prod_{j=1}^{i-1} p_{j,j+1}, \quad (i = 2, \dots, M) \quad (1)$$

which is the proportion of admitted patients who will enter phase  $i$ .

It can be shown that the joint probability distribution of the number of patients in each phase is given by [6]

$$\pi(l_1, l_2, \dots, l_M) = \frac{1}{H(K)} \prod_{i=1}^M \frac{\rho_i^{l_i}}{l_i!}, \quad \text{for } \sum_{i=1}^M l_i \leq K, \quad (2)$$

where  $l_i$  is the number of patients in the  $i$ th phase; and  $\rho_i = \lambda e_i / \mu_i$  ( $i = 1, \dots, M$ ), in which  $e_i$  is the relative arrival rate to phase  $i$ ,  $\mu_i$  is the constant service rate per bed of phase  $i$ , and  $H(K)$  is the normalisation constant. By the use of the multinomial theorem,  $H(K)$  can be computed in the following closed-form

$$H(K) = \sum_{k=0}^K \frac{1}{k!} \left( \sum_{i=1}^M \rho_i \right)^k. \quad (3)$$

## 2.1 Special case: closed network

A closed network is a special case of the semi-open network model where the system is always full. In other words, a discharge from the system is immediately replaced by an admission. Such condition is reasonable for hospital departments that usually run at full occupancy level. Under this condition, the normalisation constant is given as  $H(K) = (\sum_{i=1}^M \rho_i)^K / K!$ , and therefore (2) becomes

$$\pi(l_1, l_2, \dots, l_M) = \frac{K!}{l_1! \dots l_M!} \prod_{i=1}^M r_i, \quad \text{where } r_i = \frac{e_i / \mu_i}{\sum_{j=1}^M e_j / \mu_j}. \quad (4)$$

This suggests that the number of patients in each phase jointly follow a multinomial distribution with parameters  $K$  and  $r_i$  ( $i = 1, \dots, M$ ).

## 2.2 Special case: open network

An open network is another special case of the semi-open network model where the system is assumed to have infinite capacity to admit patients. Under this condition, we can show that the normalisation constant reduces to  $H(K) = \exp(-\sum_{i=1}^M \rho_i)$ , and (2) becomes

$$\pi(l_1, l_2, \dots, l_M) = \prod_{i=1}^M \frac{\rho_i^{l_i} e^{-\rho_i}}{l_i!}, \quad (5)$$

which implies that the number of patients in phase  $i$  ( $i = 1, \dots, M$ ) are independent Poisson distributed random variables with rate  $\rho_i$ .

Although it is rarely true in reality that a hospital department will have almost infinite bed capacity, this special case can still be useful in representing an ideal scenario for comparison purpose.

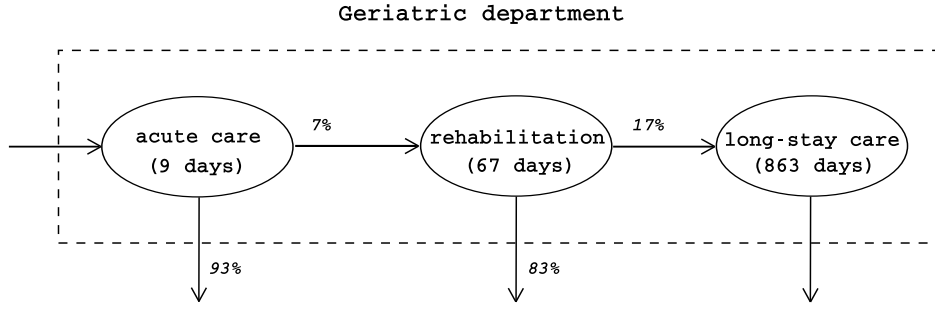


Figure 2. Parameters reported in Parry [7] for a geriatric department.

### 3 Performance measures

Measures of system performance, such as the loss probability of the system (i.e. the probability that a patient has to be turned away because the system is full), expected occupancy level, effective admission rate, etc. can be easily computed.

Since the probability that there are  $l$  ( $l \leq K$ ) patients in the system is given by

$$\pi_l = \sum_{\sum_{i=1}^M l_i=l} \pi(l_1, l_2, \dots, l_M) = \frac{1}{H(K)} \frac{\left(\sum_{i=1}^M \rho_i\right)^l}{l!}, \quad (6)$$

therefore, the loss probability of the system, i.e. the probability that a patient has to be turned away because the system is full, is the probability that there are exactly  $K$  patients in the system, which is given by

$$P_{loss} = \pi_K = \frac{1}{H(K)} \frac{\left(\sum_{i=1}^M \rho_i\right)^K}{K!}. \quad (7)$$

The expected number of patients in the system is given by

$$EL = \sum_{k=1}^K k\pi_k = (1 - P_{loss}) \sum_{i=1}^M \rho_i, \quad (8)$$

with the corresponding variance as

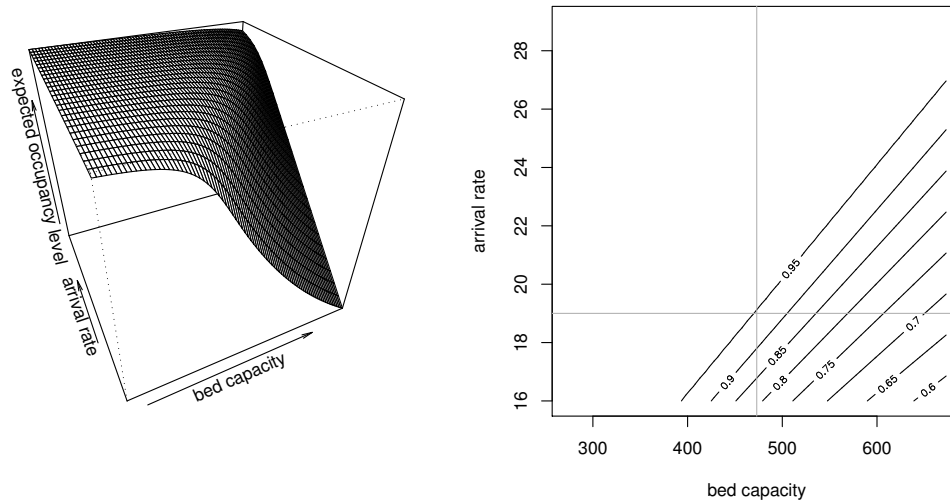
$$\text{Var}L = (\pi_K - \pi_{K-1} - \pi_K^2) \left(\sum_{i=1}^M \rho_i\right)^2 + (1 - \pi_K) \sum_{i=1}^M \rho_i. \quad (9)$$

Similarly, the expected number of patients at the  $i$ th phase ( $i = 1, \dots, M$ ) of the system is given as

$$EL_i = (1 - P_{loss}) \rho_i, \quad (10)$$

with the corresponding variance as

$$\text{Var}L_i = (\pi_K - \pi_{K-1} - \pi_K^2) \rho_i^2 + (1 - \pi_K) \rho_i. \quad (11)$$



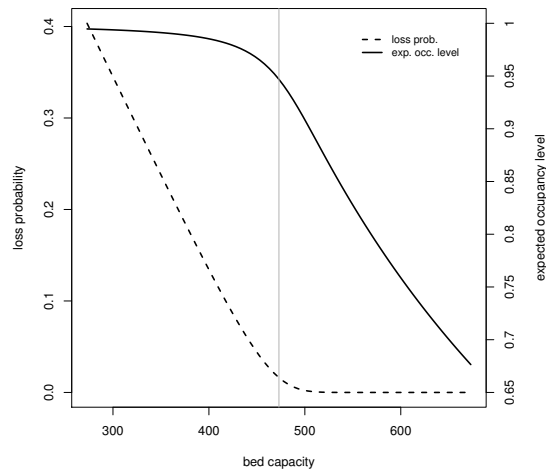
**Figure 3. Expected occupancy level as a function of bed capacity and arrival rate. The graph on the right is the contour plot of the 3D version on the left. Current values of bed capacity and arrival rate are indicated by the gray lines.**

#### 4 Application of the model

We demonstrate the use of the model using parameters reported by Parry [7], who applied the methodology proposed in [2] to analyse the data of a geriatric department, which had 473 beds. Three phases were suggested - acute care (1st phase), rehabilitation (2nd phase) and long-stay care (3rd phase) with average length of stay being 9 days, 67 days and 863 days, respectively. Upon leaving the 1st phase, 7% of the patients move to the 2nd phase, of which 17% move to the 3rd phase (see Fig. 2). New patients arrived at an estimated rate of 19 patients per day.

Using equations derived in the previous sections, we compute that the number of patients in the department is expected to be 447.9 (s.d = 16.2), and the expected number of patients in each phase is 168.2 (s.d = 11.9), 87.7 (s.d = 9.0), 192.0 (s.d = 12.6), respectively. This suggests that more than 40% of the beds are expected to be occupied by long-stay patients. The expected occupancy level is 94.7%, and the probability that a patient will be refused admission because the department is full is expected to be 0.016.

In addition, we can use the model to study how the performance measures change in relations to changes in system parameters, such as bed capacity, arrival rate, etc. Fig. 3 shows how changes in bed capacity and patient arrival can have a large impact on expected occupancy level in the department. Hospital managers can use the contour plot to estimate the number of extra beds required to cope with increased demand if the department is simply to maintain the same level of occupancy level. Generally, the number of beds in a department represents a trade-off between the desire to lower the risk of turning patient away because the department is full, and the preference of having high occupancy level. Fig. 4 shows the impact of changing bed capacity on loss probability and expected occupancy level. This plot allows hospital managers to visualise and balance these conflicting goals when deciding on the number of beds.



**Figure 4. The effect on loss probability and expected occupancy level due to changes in bed capacity. Current bed capacity is indicated by the gray line.**

## 5 Conclusion

In this paper, we introduce a novel approach to the development of a realistic and generic model of patient flow in a health care system. By incorporating a total bed capacity constraint, above which new patients will be refused admission, this model provides a realistic representation of a healthcare system. Health care policy makers will be able to use this model to gain better understanding of the dynamics of patient flow and to study potential long-term impacts of policy changes, which will contribute to better care for patients.

## References

- [1] M. J. Côte, “Understanding patient flow”, *Decis Line*, 2000, 31:8–10.
- [2] G. W. Harrison, and P. H. Millard, “Balancing acute and long-term care: The mathematics of throughput in departments of geriatric medicine”, *Methods Inf Med*, 1991, 30:221–228.
- [3] M. J. Faddy, and S. I. McClean, “Markov chain modelling for geriatric patient care”, *Methods Inf Med*, 2005, 44:369–373.
- [4] V. Irvine, S. I. McClean, and P. H. Millard, “Stochastic models for geriatric in-patient behaviour”, *IMA J Math Appl Med Biol*, 1994, 11:207–216.
- [5] T. J. Chausalet, H. Xie, and P. H. Millard, “A closed queueing network approach to the analysis of patient flow in health care systems”, *Methods Inf Med*, 2006, 45:492–497.
- [6] H. Chen, and D. D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization*, Springer-Verlag, New York, 2001.
- [7] A. Parry, “An age-related service revisited”, In: P. H. Millard, and S. I. McClean (editors), *Go with the Flow: A system approach to healthcare planning*, pp. 127-130, Royal Society of Medicine Press, London, 1996.