Approximate Analysis of $M/G/c/c$ State Dependent Queueing Networks

F. R. B. Cruz [a,1] and J. MacGregor Smith [b]

[a] *Department of Statistics, Federal University of Minas Gerais, 31270-901 - Belo Horizonte - MG, Brazil*

[b] *Department of Mechanical & Industrial Engineering, University of Massachusetts, Amherst MA 01003.*

Abstract

Congestion is ever present in most practical situations. We describe a methodology for approximate analysis of open state-dependent $M/G/c/c$ queueing networks in which the service rate is subject to congestion, that is, is a function of the number of customers in the system. Important performance measurements are easily computed with high accuracy, such as the blocking probability, throughput, expected number of customers in the system (known also as work-in-process), and expected waiting time. The methodology forms a basic building block useful in many practical applications and contexts. Computational results demonstrate that the methodology provides accurate results in many topological configurations as well as in the analysis of network evacuation problems in high-rise buildings.

Key words: Queueing, finite capacity, state dependent.

1 Introduction

1.1 Motivation

Waiting in line is one of the most common annoyances faced in modern society. Nobody likes to wait and no manager of the service establishments likes

[1] Corresponding author. Prof. Frederico R. B. Cruz. E-mail: fcruz@est.ufmg.br. Fax: (+55 31) 3498 5924. This research was developed while the author was a visiting professor at the Department of Mechanical & Industrial Engineering, University of Massachusetts, Amherst MA 01003.

Preprint submitted to Computers & Operations Research *10 August 2005*

impatient or lost customers since it costs them business. A line (or queue) is formed every time there is more demand than service available. In order to answer crucial questions such as "what should the level of service be to ensure satisfactory service?", one needs to know the expectations of the customers and also how to estimate performance measures for the corresponding stochastic system [1], such as for instance how many customers are expected to be waiting and for how long. Our main objective with this paper is to present a methodology suitable for answering the second question for open $M/G/c/c$ state-depedent queueing networks. Thus, an approximate analysis is proposed for open queueing networks with **M**arkovian inter-arrival time distribution, **G**eneral and state-dependent service time distribution (i.e., dependent on the number of customers in the system), c parallel servers, and total capacity c.

Probably, one of the most interesting problems that could be treated by the methodology to be presented here is related to pedestrian flows [2, 3, 4]. Applications in which the state-dependent model can be also of use include vehicular traffic modeling [5, 6], industrial engineering problems [7], and many others for which the service rate of customers decreases with increasing customer traffic [8]. In fact, Figure 1 presents experimental curves (a through f) that relate the walking speed of a pedestrian to the crowd density, based on various empirical studies [3] in which it was stated that at a mean density of 3 pedestrians per square meter (ped/m^2), walking is reduced to a shuffle, and, beyond 5 ped/m^2, forward movement is halted.

1.2 Contributions and Outline of the Paper

The main contribution of this paper is to present a new methodology suitable for approximate analysis of open $M/G/c/c$ state dependent queueing networks. Also, we will present an empirical study upon the inter-departure process between the queues of the network and will discuss an original application to network evacuation of high-rise buildings.

This paper is organized as follows. In Section 2, we present background material with the analytical congestion model for single queues. Then, in Section 3 we start by presenting the Generalized Expansion Method, a well-known method successfully used in the past to approximate performance measurements in network of queues, and ends by describing a new implementation of the method to the special case of $M/G/c/c$ state-dependent queueing networks. In Section 4, we present computational results comparing the new method with simulations. We also deal with networks of $M/G/c/c$ queues and discuss their inter-departure properties. Section 5 illustrates a relevant application to the analysis of evacuation networks in high-rise buildings. Finally, Section 6 completes the paper with conclusions and final remarks.

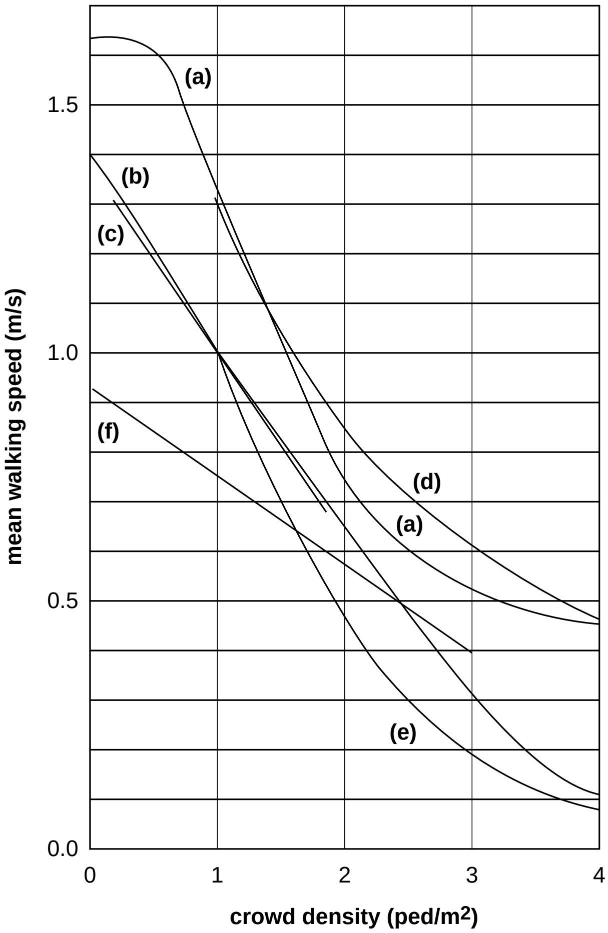

Fig. 1. Average walking speed for pedestrians.

2 Analytical Model for Flows with Congestion

In the $M/G/c/c$ state-dependent queueing model of some circulation space in a building or environment (for pedestrians, vehicles, parts etc.) the circulation space behaves as the server to its occupants. The number of servers is equal to the capacity of the circulation space, c. Entities that arrive to the circulation space when it is at capacity may not enter it. The queue consists entirely of the circulation space without any additional buffer floor space. Therefore, the total number of entities that are allowed in the entire queueing system is also equal to the capacity of the circulation space.

The limiting probabilities for the random number of entities N in a $M/G/c/c$ queueing model, $p_n \equiv \Pr[N = n]$, are as follows [4]

$$p_n = \left\{ \frac{\left[\lambda E[T_1]\right]^n}{n! f(n) f(n-1) \cdots f(2) f(1)} \right\} p_0, \;\; n = 1, 2, \ldots, c, \tag{1}$$

in which p_0 is the empty system probability, given by

$$p_0^{-1} = 1 + \sum_{i=1}^{c} \left\{ \frac{\left[\lambda E[T_1]\right]^i}{i! f(i) f(i-1) \cdots f(2) f(1)} \right\}, \tag{2}$$

λ is the arrival rate, $E[T_1] = l/V_1$ is the expected service time of a lone entity in a circulation space of length l, considering that V_1 is the speed of a lone entity, and c is the capacity of the circulation space.

For our convenience, from now on we will consider only the particular application of the model to the indoor traffic of pedestrians. Thus, we have that a corridor may be seen as the circulation space, the speed of a lone occupant is $V_1 \approx 1.5$ m/s, and the capacity of a corridor may be assumed to be

$$c = \lfloor klw \rfloor, \tag{3}$$

in which $\lfloor x \rfloor$ is the largest integer not superior to x, l is the length, w is the width, and k is the capacity of the corridor per square-unit. Considering pedestrian related applications and realizing that k represents the crowd density, a reasonable value for k would be 5, because beyond this density, movement ceases, in accordance to experimental studies [3].

Notice that in (1) and (2), $f(n) = V_n/V_1$ is the service rate, considered to be the ratio of the average walking speed of n people in the corridor to that of a lone occupant V_1. Basically, what one wants is that the congestion model represents the effect depicted in Figure 1, in which the service rate depends on the number of customers in the system. Successful in the past was an exponential model in which the service rate decays following the expression [4]

$$f(n) = \exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma}\right], \tag{4}$$

with $\gamma = \log\left[\frac{\log(V_a/V_1)}{\log(V_b/V_1)}\right] / \log\left(\frac{a-1}{b-1}\right)$, and $\beta = \frac{a-1}{[\log(V_1/V_a)]^{1/\gamma}} = \frac{b-1}{[\log(V_1/V_b)]^{1/\gamma}}$.

The values a and b are arbitrary points used to adjust the exponential curve. In pedestrian related applications, commonly used values are $a = 2lw$ and $b = 4lw$ corresponding to crowd densities of 2 and 4 ped/m^2 respectively. Looking at the curves presented in Figure 1, reasonable values for such points are $V_a = 0.64$ and $V_b = 0.25$.

From (1), important performance measures can be derived

$$\begin{cases} p_c &= \Pr[N = c], \\ \theta &= \lambda(1 - p_c), \\ L &= E[N] = \sum_{n=1}^{c} n p_n, \\ W &= E[T] = L/\theta, \end{cases} \tag{5}$$

in which p_c is the blocking probability, θ is the throughput in ped/s, L is the expected number of customers in the corridor (also known as work-in-process, WIP), and W, here derived from Little's formula, is the expected service time in seconds. A interesting study on the effect of an increase of the node capacity, c, in the network performance is presented by Mitchell & MacGregor Smith [9]. As a final remark, we hope it is clear that the methodology is general. Although illustrated here for the case of pedestrians, the model could be adjusted to other applications or to different data sets [6, 7].

3 Generalized Expansion Method for Networks of $M/G/c/c$ Queues

3.1 Background

The Generalized Expansion Method (GEM) was proposed by Kerbache & MacGregor Smith [10] back in the eighties and has a long research tradition on the subject. Explained recently in its most basic features [11], the GEM is a combination of repeated trials and node-by-node decomposition approximation methods, with a key characteristic that an artificial holding node is added preceding each finite queue in the network in order to register blocked customer that attempt to enter the finite node when it is at capacity (see Figure 2). By the addition of holding nodes, the queueing network is expanded into an equivalent Jackson network, in which each node can then be decomposed and analyzed separately. Now, we shall describe briefly the GEM.

Stage 1: Network Reconfiguration

For each node with finite capacity, an artificial node is added directly preceding it, as shown in Figure 2. Customers that attempt to move to the new node but are unable since it is at capacity are re-routed to the artificial node. The probability that an arriving customer is blocked by node j equals p_{c_j}. Thus, with probability $(1 - p_{c_j})$, it will enter node j, and with probability p_{c_j} it will

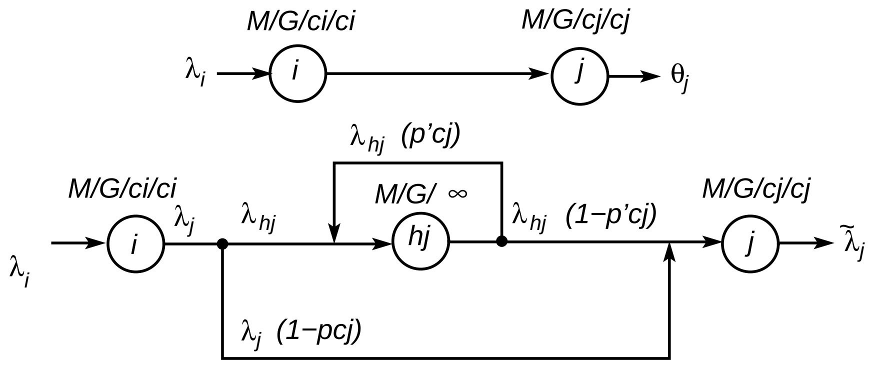

Fig. 2. Generalized expansion method.

enter holding node h_j. The holding node is modeled as an $M/G/\infty$ queue, so that there will be no waiting to enter this node.

After service at the holding node, the customer will be blocked with a new probability, p'_{c_j}. With probability $(1-p'_{c_j})$, it will proceed to the following node. Otherwise, it must retrace its path through the feedback loop into artificial node h_j again.

Stage 2: Parameter Estimation

The value of p_{c_j} can be determined from known analytical results, i.e, directly by (1), for $M/G/c/c$ state-dependent queues, $p_{c_j} \equiv \Pr[N = c_j]$. The value of p'_{c_j} is determined from approximation results that use diffusion techniques [12].

Stage 3: Feedback Elimination and Service Update

A reconfiguration of the holding node is performed, so that the strong dependencies in the arrival process caused by the repeated visits (feedback) to the artificial node are removed. The feedback arc is removed from the holding node by recomputing the service rate at this node. Finally, it can be shown that the average service time that a customer spends at node i preceding node j is given by

$$\tilde{\mu}_i^{-1} = \mu_i^{-1} + p_{c_j}(\mu'_h)^{-1}. \tag{6}$$

Actually any general network topology could also be handled by the GEM. The interested reader should check details on the paper of Kerbache & MacGregor Smith [11]. Now, we shall describe a new algorithm to approximately compute (6), especially developed for $M/G/c/c$ state dependent queueing networks.

3.2 New Algorithm

We remind the reader that (6) represents the last step of the GEM, whose ultimate goal is to provide an approximation scheme to update the service rates of upstream nodes that takes into account all blocking after service in there, caused by downstream nodes. The GEM give us the rationale to the approximation we are about to present. Because of the state-dependent service times in $M/G/c/c$ queues, the computation of (6) is not straightforward. Thus, an iterative algorithm to compute (6) will be proposed.

First, let us define the network under study as a graph $G(V, A)$, in which V is a finite set of nodes (corridors) and A, a finite set of arcs (connections between pair of corridors). The proposed algorithm, Figure 3, is based on the fact that blocking in the downstream nodes cause blocking after service at the upstream nodes. As a result, the effective service rates at the upstream nodes are reduced in a similar fashion as in Markovian systems, in accordance to (6). In $M/G/c/c$ state-dependent queueing networks, a similar effect is present, regardless of the dynamic changes in the service rates as users joint or leave the system. Thus, the algorithm progressively adjusts the state-dependent service rate curve, as blocking increases in the downstream nodes. The adjustments are made by considering that flow conservation throughout the networks is a constraint that must be satisfied, as explained in the following paragraphs.

The performance evaluation algorithm is comprised of two parts. First, Figure 3-a, a pre-evaluation is performed to every node in the network disregarding any possible inter-node blocking effect. In other words, the pre-evaluation procedure approximates the actual network to a Jackson network, in which each node can be analyzed separately. Notice that the pre-evaluation procedure is a variant of Dijkstra's labeling algorithm for determination of shortest paths in graphs [13]. For instance, in the network illustrated in Figure 4, a possible valid sequence to perform pre-evaluations is $1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 5 \rightarrow 6$. The reason that a node could be pre-evaluated only after all of its predecessors is that one needs to know the overall arrival rate at this node before proceeding with any calculation and, clearly, the total arrival rate is a function of the outputs of all predecessor nodes. Under very light traffic, with negligible inter-node blocking effects, the network is actually a Jackson network and the pre-evaluation procedure alone will do just fine for an accurate approximation of all performance measures.

The second part of the algorithm, Figure 3-b, seeks flow conservation, that is $\theta_j \leq \lambda_j + \sum_{\forall\, i|(i,j)\in A} \theta_i p_{ij}$, for all $j \in V$. The evaluation algorithm is a labeling algorithm working in reverse. For the network presented in Figure 4, a possible valid sequence to perform the evaluations is $6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$, once a node can be evaluated only if all of its successors are already evaluated

```
algorithm
    read graph, G(V, A)
    read arrival rates and lengths, λ_i, l_i, ∀ i ∈ V
    read routing probabilities, p_ij, ∀ (i, j) ∈ A
    initialize set of labeled nodes, P ← ∅
    while P ≠ V
        choose j such that (j ∈ V) and (j ∉ P)
        if {i| (i, j) ∈ A} ⊆ P then
            /* compute performance measures */
            E[T_1]_j ← l_j/V_1
            compute Pr[N = c_j], θ_j, L_j, W_j
            /* forward information to successors */
            for ∀ k ∈ {k'| (j, k') ∈ A} then
                λ_k ← λ_k + θ_j p_jk
            end for
            /* label node as pre-evaluated */
            P ← P ∪ {j}
        end if
    end while
end algorithm
```

a) pre-evaluation algorithm

```
algorithm
    initialize set of labeled nodes, P ← ∅
    initialize maximum throughput, θ_i^max ← ∞, ∀i ∈ V
    while P ≠ V
        choose i such that (i ∈ V) and (i ∉ P)
        if {j| (i, j) ∈ A} ⊆ P then
            /* update performance measures */
            E[T_1]*_i ← min E[T_1]_i
                        s.t.: θ_i ≤ θ_i^max,
                              E[T_1]_i ≥ l_i/V_1
            compute Pr[N = c_i], θ_i, L_i, W_i
            /* backpropagate to predecessors */
            for ∀ k ∈ {k'| (k', i) ∈ A} then
                update θ_k^max
            end for
            /* label node as evaluated */
            P ← P ∪ {k}
        end if
    end while
end algorithm
```

b) evaluation algorithm

Fig. 3. Performance evaluation algorithm.

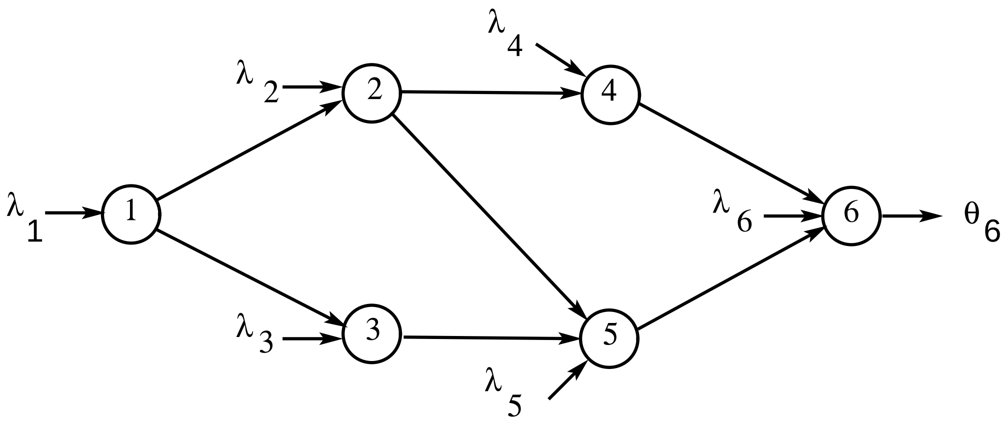

Fig. 4. $M/G/c/c$ queues in a mixed topology.

as explained ahead. The evaluation algorithm is designed to correct estimates mainly in those cases with bottlenecks in the downstream nodes. The estimates are improved by means of adjustments in the expected service time for lone occupant of each node i, $E[T_1]_i$, as explained in the following.

We define $\theta_i^{\max}$ as the maximum throughput of node i, i.e. the maximum flow that can be forwarded to the set of successors of i. Thus, if i is an end node (that is, without any successors), obviously $\theta_i^{\max} = \infty$. Otherwise, the update of $\theta_i^{\max}$ will depend on the knowledge of the throughput of all successors of i. Initially, all nodes are assumed to have no limits and $\theta_i^{\max}$ is set to infinity. Starting from the end nodes, $E[T_1]_i$ is minimized, subject to providing a throughput less or equal to $\theta_i^{\max}$ and not being inferior to l_i/V_1, in which l_i is the length of the ith corridor and V_1, the walking speed for lone occupant ($V_1 \approx 1.5$ m/s, for the pedestrian application). Then, all performance measures of i are updated, taking into account this new $E[T_1]_i$. Finally, $\theta_i^{\max}$ is updated for the set of predecessors of i. Updating is done by assigning to predecessors an equal part of θ_i. If some predecessor is not able to provide its share, the flow is rerouted among the other predecessor to ensure flow conservation.

Assuming there are no circuits in $G(V, A)$ then:

Lemma 1 *The new performance evaluation algorithm has running time complexity $\mathcal{O}(V^2)$.*

Proof: *Follows from Dijkstra's algorithm.*

■

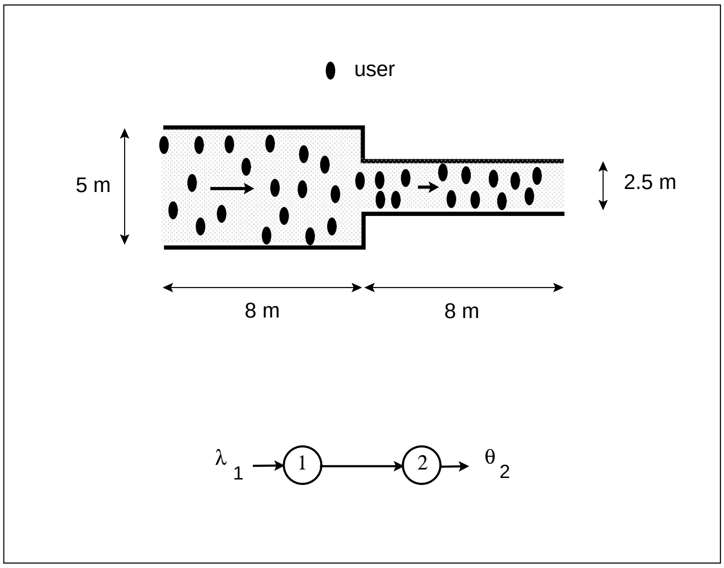

Fig. 5. 2-node system under analysis.

4 Computational Experiments

The performance evaluation algorithm was coded in C++ and is available upon request from the authors. All computational experiments were carried out on a PC, CPU Pentium II 400 MHz, 256 MB RAM, running Microsoft Windows NT 4.0 operating system. In order to confirm the accuracy of the estimates, a discrete-event digital simulation model [14] was used. We performed 30 replications to compute 95% confidence intervals. For all cases tested, the performance evaluation algorithm runs within a single minute.

4.1 Remarks on Basic Topologies

Experiments (not shown) were conducted for basic configurations. Assuming homogeneous (same length) queues, simulation results confirmed the accuracy of the estimates, being the analytical approximate results mostly within the 95% confidence interval. The interested reader may check details in the paper of Cruz et al. [14].

A question that is critically relevant at this juncture is whether corridors in various topologies would be accurately modeled as an $M/G/c/c$ network. In order to gain insight into this question, we will analyze the system presented in Figure 5, composed by two corridors, node #1: 8.0×5.0 m^2, and node #2: 8.0×2.5 m^2, modeled as a 2-node $M/G/c/c$ state dependent queueing networks in tandem. Someone may argue that the problem is rather simple but as we shall see it is general enough for our purposes.

We remark that the bottleneck at node #2 is the most difficult case to be analyzed because of the strong dependency it may cause among the two queues, as a function of the arrival rate. Notice that, under heavy traffic and the bottleneck being otherwise in node #1, part of the incoming flow would be forced to balk, consequently decoupling the two queues. Thus, one would have a Jackson network whose queues could be analyzed independently, given that the output of an $M/G/c/c$ queue is well-known to be Poisson [15]. Soon, we shall demonstrate corroborating experimental evidence on these two remarks.

4.1.1 Analysis for $\lambda = 2.0$

First, we assumed an arrival rate $\lambda = 2.0$ ped/s (7200 ped/h) and 2,000 s of total time simulated. Figure 6 shows the service times at nodes #1 and #2. We can also draw histograms for the service time as seen in Figure 7. Notice that a departure from node #1 may be subject to blocking after service but in this case there will be none, since the arrival rate is too light and easily handled by node #2 without any blocking.

In this case of $\lambda = 2.0$ ped/s, the two queues are in fact decoupled and it is well known that under such circumstances the departure process in each of them must be Poisson [15]. In Figure 8 and Table 1, one can see the inter-departure times produced by the simulation model. Indeed, the histogram is visually identical to an exponential distribution, data have roughly same mean and standard deviation, an a high p-value > 0.30 obtained for the χ^2 test does not reject the null hypotheses of exponential distribution. Additionally, no autocorrelation was observed.

4.1.2 Analysis for $\lambda = 3.0$

A more interesting case occurs when we consider $\lambda = 3.0$ ped/s (10,800 ped/h) and 2,000 s of total time simulated, also presented in Figure 6. Notice that there will be many users blocked, since now the arrival rate is too high to be handled by node #2 without any blocking, as seen in Figure 6, that shows a waiting time above 50 seconds.

The first important fact to be observed here is that it is necessary to discard a *warm-up* period before making any steady-state analysis. So, ignoring the first half of the 2,000 s simulated, we can draw histograms for the service time as seen in Figure 7. Notice that the system is so overloaded that the service time in node #2 is almost 10 times larger than the lone occupant service time ($8.0/1.5 \approx 5.3$ s). Also, notice how differently the service times histograms are. Indeed, node #2 has virtually a deterministic service time. Here, we can convince ourselves that a more flexible analytical model than simply an M/M must be used if more accuracy in modeling congestion is needed.

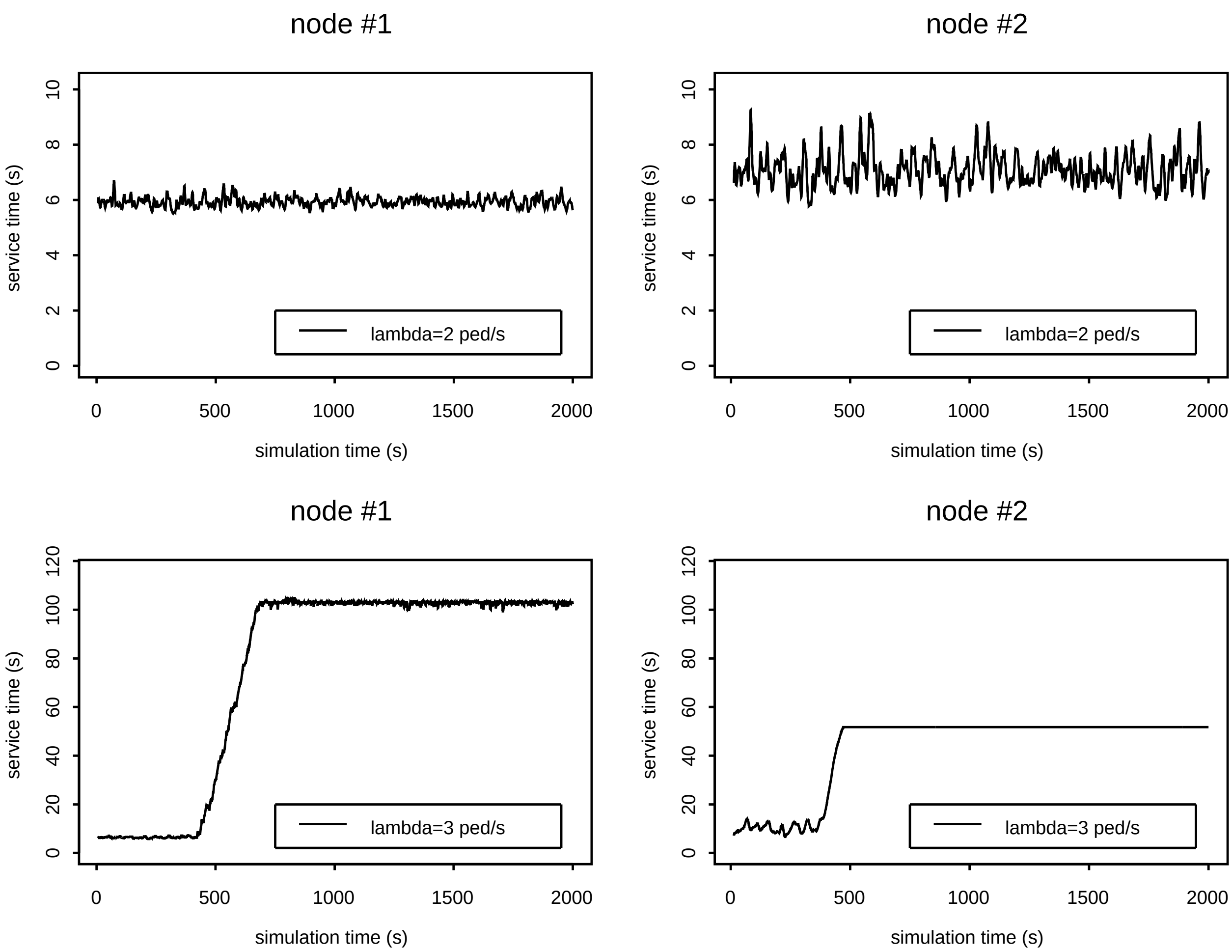

Fig. 6. Service times along the simulation.

Studying the inter-departure times for this system, and, again, discarding the first half of the observations in order to avoid transient perturbations, we can obtain the histogram presented in Figure 8 and find out that the departure process from node #1 is not far from the exponential distribution. The descriptive statistics presented in Table 1 reinforce this conclusion. However, some care should be take as some autocorrelation is seen in data [16].

Let us try and explain the results from the simulation model and relate them to the known results of a single node $M/G/c/c$ model and generalize these results. In the particular 2-node network simulated, let us hypothesize that there is an equivalent single-node representation of the 2-node network. This equivalence is possible since there is no queue (waiting room) for $M/G/c/c$ nodes.

If we take the 2-node network and convert it to a single node of dimensions 24 m $\times$ 2.5 m then the performance results of the single node network should match the simulation results, since the single node has the same capacity as

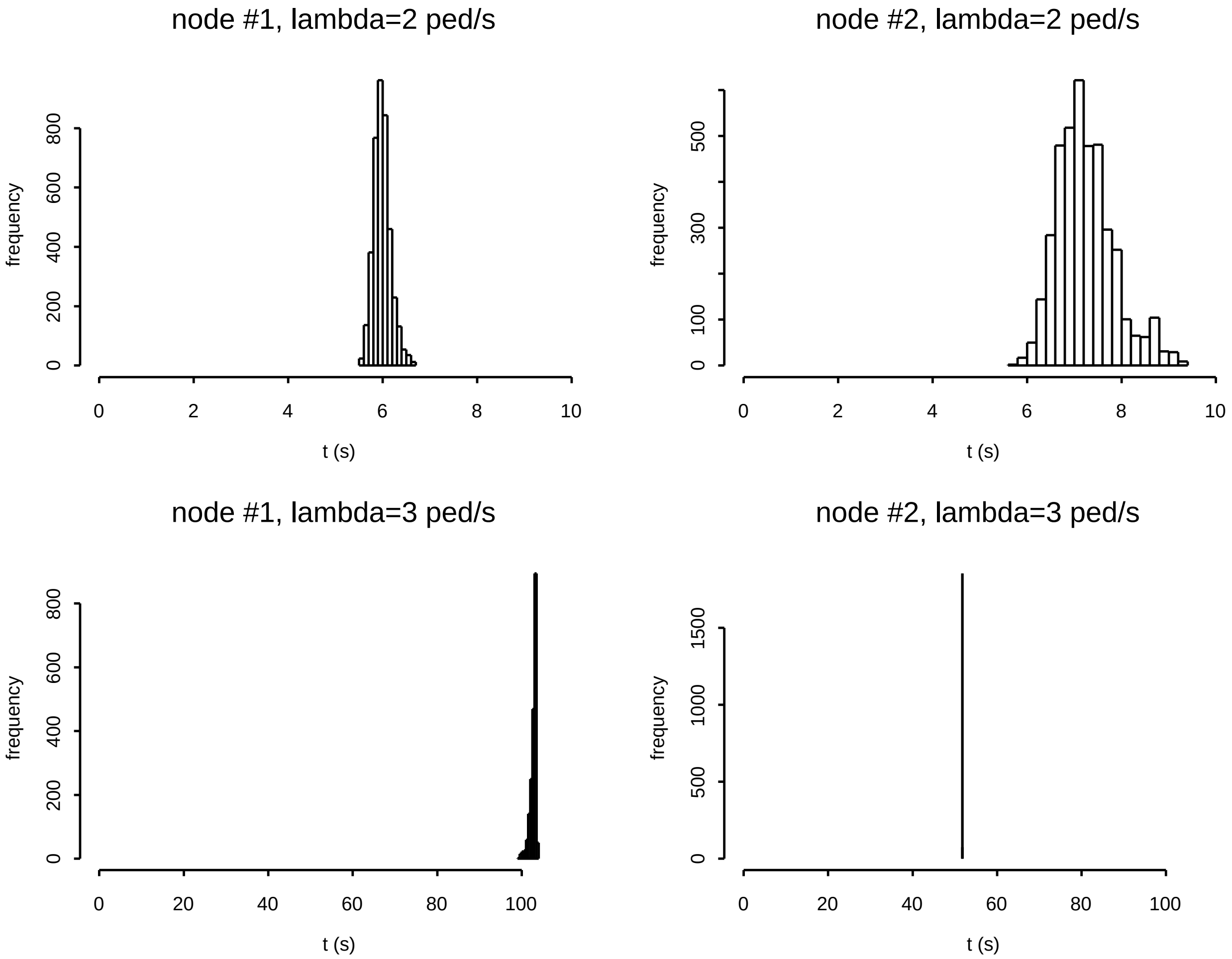

Fig. 7. Service time histograms.

Table 1
Descriptive statistics for the inter-departure time and model adjustment.

									exponential distribution		
λ	Node	Min.	Q1	Median	Mean	Std. Dev.	Q3	Max.	λ	MSE	p-value
2	1	0.000122	0.144	0.341	0.490	0.475	0.686	3.51	0.490	0.000473	0.303
	2	0.000122	0.147	0.338	0.492	0.484	0.687	4.41	0.492	0.000497	0.366
3	1	0.00214	0.160	0.367	0.518	0.488	0.797	2.15	**0.518**	0.0128	**0.005**
	2	0.00214	0.160	0.367	0.518	0.488	0.797	2.15	**0.518**	0.0128	**0.005**

the 2-node network and the performance results essentially match the inter-departure rate from the simulation model for node #2, Table 1, which is

$$\theta^{-1} = 0.518 \Rightarrow \theta = 1.931 \text{ ped/s}.$$

In the transformation process from 2 nodes to 1 node, there is a critical dimension that must be maintained, leaving the other dimension variable. In the

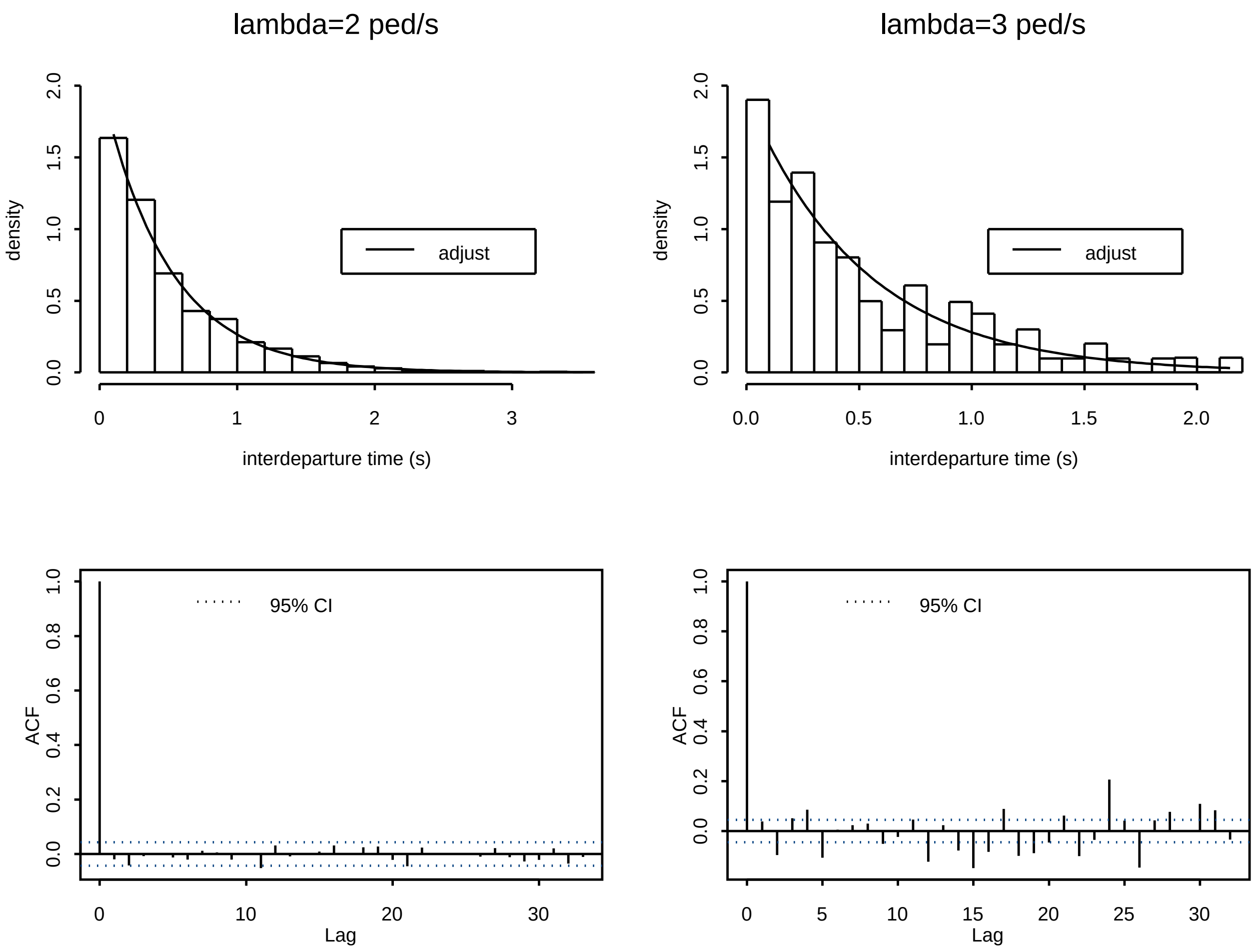

Fig. 8. Inter-departure time histograms and autocorrelation functions for node #1.

following output we see the results from the single node performance measures of a $M/G/c/c$ queue of 24 m × 2.5 m where the corridor width was fixed.

```
LINEAR (1), OR EXPONENTIAL (2), VELOCITY MODEL.
2
INPUT LAMBDA, LENGTH, AND WIDTH RESPECTIVELY
3 24 2.5
System Capacity = 300.000 # in System = 298.106
Effective Lambda= 1.945   Waiting Time= 153.295
Blocking Probability = 0.35178318619728
```

These results match the output results of the simulation model, i.e.,

$$\theta = 1.931 \approx \texttt{Effective Lambda} = 1.945 \text{ ped/s}.$$

Since we already know that the inter-departure time distribution from a single node $M/G/c/c$ is exponential [15], then:

Corollary 1 *Any 2-node $M/G/c/c$ network that has an equivalent single-queue representation will have an output (including those that are lost) which is Poisson.*

The proof is by virtue of the results from the single node network and the quasi-reversible properties of $M/G/c/c$ queues (for more information, see [17]).

This result can be generalized to more than 2-nodes for certain tree-topologies, since the transformation process should be straightforward. For more general corridor shapes, the transformation process would be a bit trickier and this process is reserved for another paper.

While the above property is a very important result for 2-node networks, one may not always want to find the equivalent single-node representation for complex series, merge, and split networks. One would like to treat the network of $M/G/c/c$ queues algorithmically. Therefore, we need an accurate approximation tool such as the one proposed in the previous section of the paper.

5 Analysis of Evacuation Networks

Here, we intend to illustrate a relevant application for the proposed algorithm. One possible use is for the analysis of evacuation networks in high-rise buildings. In case of emergencies, the network of corridors and stair-wells are crucial links for safe evacuation of the occupants. The objective is to identify possible bottlenecks constricting the flow of occupants out of the area.

5.1 Example

Figure 9 illustrates a 10-story building. Stairwells 8.5 m long $\times$ 1.2 m wide are assumed to interconnect each floor and each one of them is accessible by a corridor also 8.5×1.2, forming a 20-node series-merge network, as seen in Figure 9. We also assume an arrival rate λ equally assigned for each of the ten floors. Arrival rates of 0.125, 0.25, 0.5, and 1.0 ped/s were examined.

5.2 Computational Results

Figure 10 shows the performance measures as a function of the node. Nodes #20 to #11 represent the access corridors to the stairwell. Nodes #10 to #1 represents the stairwells between tenth and first floor. A better way to see

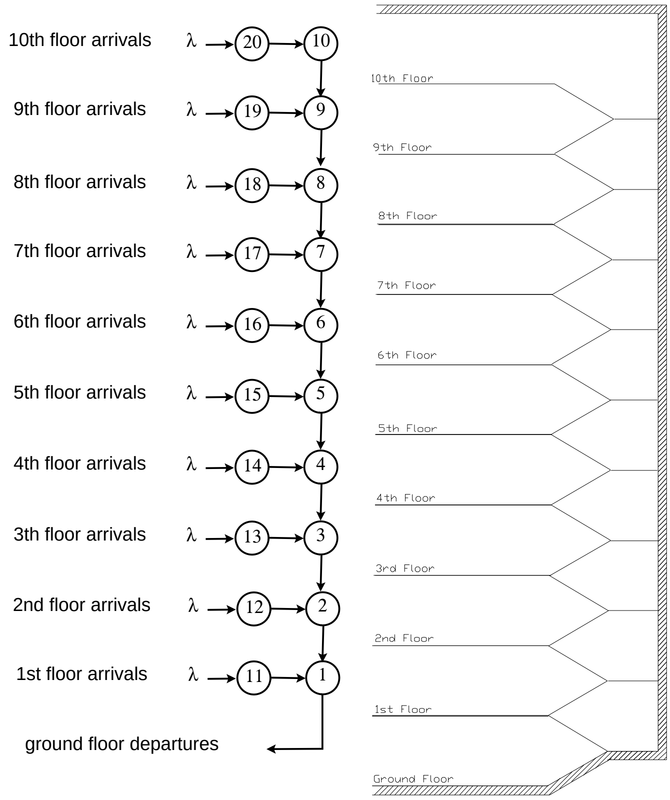

Fig. 9. A 10-story building section and queueing network.

the pattern of each performance measure is through graphs, as illustrated by Figure 10. Since nodes #20 to #11 illustrate the corridors and node #10 to #1, the stairwells, the results in Figure 10 are kept separate. We could not use simulation to verify directly the accuracy of these results because of time overflow.

The blocking probability p_c remains almost constant and insignificant under low traffic but rises exponentially towards the upper stories as the arrival rate increases. The same pattern is verified both for the corridor nodes (#20 to #11) and for the stairwell nodes (#10 to #1).

Under a low arrival rate, the throughput θ will decrease linearly towards the upper stories of the stairwell but remains constant along all of access corridors. However, under high arrival rates the throughput will decrease exponentially towards the top being quite small in the upper story. This is a clear indication that the upper stories will be the ones to be most severely affected by the blocking effect caused by an undersized stairwell.

There is not much to say about the expected number of users L. In the corridors and stairwells, L can reach its limit ($5 \times 8.5 \times 1.2 = 51$) even under

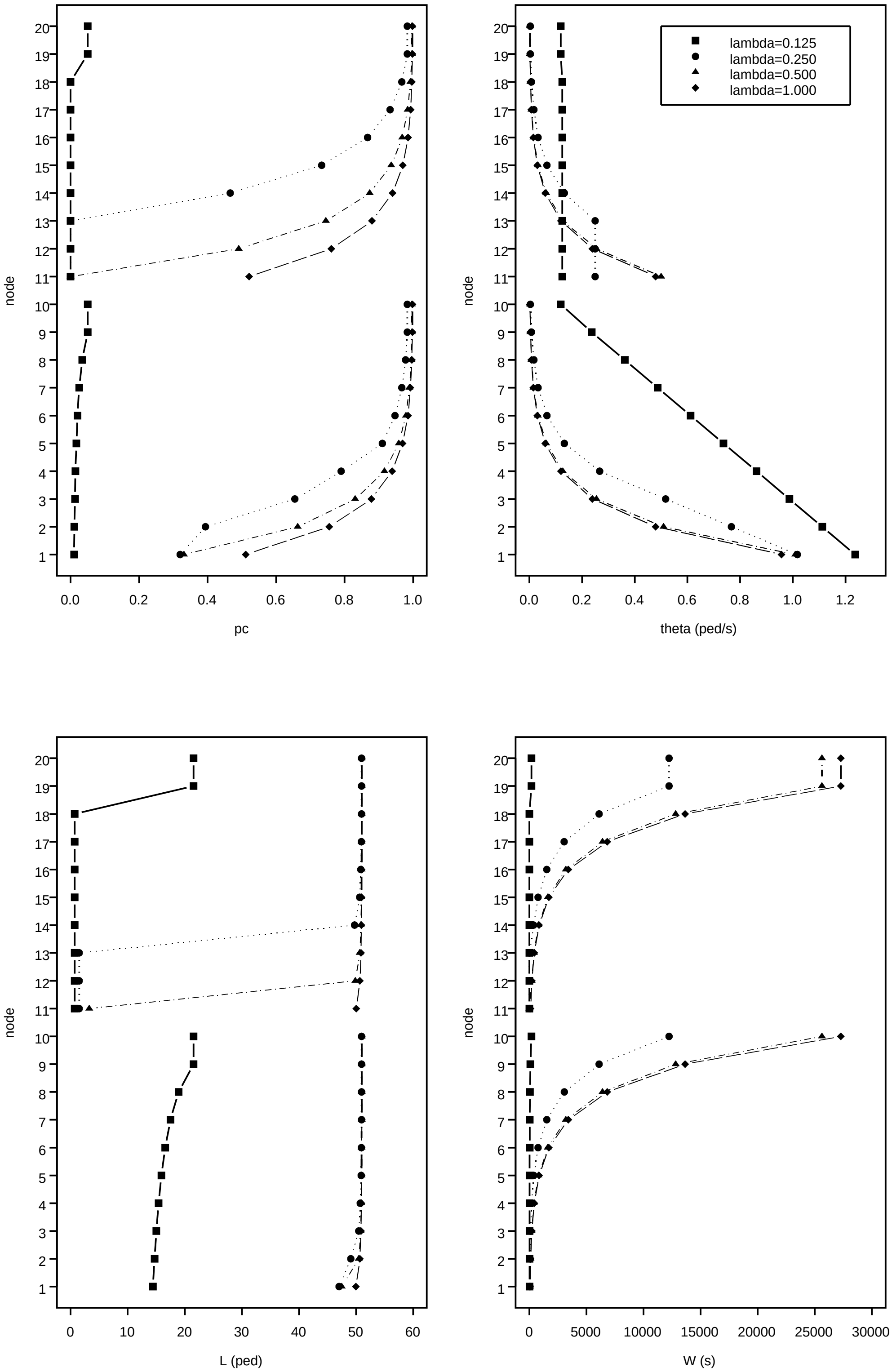

Fig. 10. Performance measures for a 10-story building.

the relatively small arrival rate of $\lambda = 0.25$ ped/s. However, this performance measure may help one to better see that something must be inappropriate with corridors and stairwells that fill up so easily.

The expected service time, W, perhaps one of the most highly regarded performance measure in evacuation plans, is interesting to be analyzed closely. Here we can see that the service time of each node increases exponentially towards the top of the building even under light traffic so that one can expect also exponentially increasing evacuation times. Additionally, these times can vary significantly even under slight changes in the arrival rate leading us to a very unstable stochastic system whose behavior is very hard to predict. As a final remark, the evacuation problem was analyzed closely by Cruz et al. [18] when they proposed an algorithm to optimize the stairwell case and found a solution that led to linear increasing evacuation times towards the upper stories.

6 Concluding Remarks

We have presented a general methodology suitable for approximate analysis of $M/G/c/c$ queueuing networks with state dependent service rates. The importance of this model was stressed and a brief review of recent results in the area was presented. We have also discussed in detail the application of the model to pedestrian traffic flow problems. Simulation results indicated that the inter-departure times may be confidently assumed to be exponentially distributed. However, a complete understanding of the stochastic process involved still presents a formidable challenge to analytical analysis and should be the focus of future research efforts. Finally, we have presented in detail an application to the analysis of evacuation in high-rise buildings. Computational results demonstrated that the methodology is accurate and may be considered as a basic building block for more complex approaches, such as those including optimization problems that embed congestion effects.

Acknowledgments

The research of Frederico R. B. Cruz has been partially funded by the CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 201046/1994-6, 301809/1996-8, 307702/2004-9, and 472066/2004-8, the FAPEMIG (*Fundação de Amparo à Pesquisa do Estado de Minas Gerais*), grants CEX-289/98 and CEX-855/98, and PRPq-UFMG, grant 4081-UFMG/RTR/FUNDO/PRPq/99.

References

[1] Gross D, Harris CM. *Fundamentals of Queueing Theory*. John Wiley & Sons, New York, NY, 3rd edition, 1998.

[2] Fruin JJ. *Pedestrian Planning and Design*. Metropolitan Association of Urban Designers and Environmental Planners, Inc., New York, NY, 1971.

[3] Tregenza PR. *The Design of Interior Circulation*. Van Nostrand Reinhold Company, New York, NY, 1976.

[4] Yuhaski SJ, MacGregor Smith J. Modeling circulation systems in buildings using state dependent models. *Queueing Systems*, 4:319–338, 1989.

[5] Alfa AS, Neuts MF. Modeling vehicular traffic using the discrete time Markovian arrival process. *Transportation Science*, 29(2):109–117, 1995.

[6] Jain R, MacGregor Smith J. Modeling vehicular traffic flow using $M/G/C/C$ state dependent queueing models. *Transportation Science*, 31(4):324–336, 1997.

[7] MacGregor Smith J, Daskalaki S. Buffer space allocation in automated assembly lines. *Operations Research*, 36(2):343–358, 1988.

[8] MacGregor Smith J, Chikhale N. Buffer allocation for a class of nonlinear stochastic knapsack problem. *Annals of Operations Research*, 58:323–360, 1995.

[9] Mitchell D, MacGregor Smith J. Topological network design of pedestrian networks. *Transportation Research Part B*, 35:107–135, 2001.

[10] Kerbache L, MacGregor Smith J. Asymptotic behavior of the expansion method for open finite queueing networks. *Computers & Operations Research*, 15(2):157–169, 1988.

[11] Kerbache L, MacGregor Smith J.. Multi-objective routing within large scale facilities using open finite queueing networks. *European Journal of Operational Research*, 121(1):105–123, 2000.

[12] Labetoulle L, Pujolle G. Isolation method in a network of queues. *IEEE Transactions on Software Engineering*, SE-6(4):373–381, 1980.

[13] Dijkstra EW. A note on two problems in connection with graphs. *Numerical Mathematics*, 1:269–271, 1959.

[14] Cruz FRB, MacGregor Smith J, Medeiros, RO. An $M/G/C/C$ state dependent network simulation model. *Computers & Operations Research*, 32(4):919–941, 2005.

[15] Cheah J, MacGregor Smith J. Generalized $M/G/C/C$ state dependent queueing models and pedestrian traffic flows. *Queueing Systems*, 15:365–386, 1994.

[16] Kim S, Muralidharan R. Taking account of correlations between streams in queueing network approximations. *Queueing Systems*, 49:261–281, 2005.

[17] Kelly FP. *Reversibility and Stochastic Networks*. John Wiley & Sons, New York, NY, 1979.

[18] Cruz FRB, MacGregor Smith J, Queiroz DC. Service and capacity allocation in $M/G/C/C$ state dependent queueing networks. *Computers & Operations Research*, 32(6):1545–1563, 2005.