Finite-Source Queueing Systems and their Applications

János Sztrik
University of Debrecen
Institute of Mathematics and Informatics
Department of Information Technology
jsztrik@math.klte.hu

August 7, 2001

Finite-Source Queueing Systems

- Introduction
  - ⋆ Queueing systems
  - ⋆ Performance measures for finite-source systems
- Analytical Results
  - ⋆ Homogeneous $M/M/r$ systems, the classical model
- Numerical Methods
  - ⋆ A recursive method for the $M/G/1/$ system
- Asymptotic Methods
  - ⋆ Preliminary results
  - ⋆ Heterogeneous multiprocessor systems

Introduction

- Queueing systems
  - ⋆ Kendall's notations
  - ⋆ Performance measures
- Performance measures for finite-source systems
  - ⋆ Homogeneous systems
  - ⋆ Asymptotic properties
  - ⋆ Heterogeneous systems

Queueing systems

A single station queueing system consists of a queueing buffer of finite or infinite size and one or more identical servers. Such an elementary queueing system is also referred to as a service station or, simply, as a node. First we start with a short description of queueing systems, see for example, [9, 15, 29, 79].

A server can only serve one customer at a time and hence, it is either in a "busy" or an "idle" state. If all servers are busy upon the arrival of a customer, the newly arriving customer is buffered, assuming that buffer space is available, and waits for its turn. When the customer currently in service departs, one of the waiting customers is selected for service according to a *queueing (or scheduling) discipline*. An elementary queueing system is further described by an arrival process, which can be characterized by its sequence of interarrival time random variables $\{A_1, A_2, \dots\}$. It is common to assume that the sequence of interarrival times is independent and identically distributed, leading to an arrival process that is known as a renewal process. The distribution function of interarrival times can be continuous or discrete.

The average interarrival time is denoted by $E[A] = \overline{T}_A$ and its reciprocal by the average arrival rate λ:

$$\lambda = \frac{1}{\overline{T}_A}. \tag{1}$$

The most common interarrival time distribution is the exponential, in which case the arrival process is Poisson. The sequence $\{B_1, B_2, \cdots\}$ of service times of successive jobs also needs to be specified. We assume that this sequence is also a set of independent random variables with a common distribution function.

The mean service time $E[B]$ is denoted by $\overline{T}_B$ and its reciprocal by the service rate μ:

$$\mu = \frac{1}{\overline{T}_B}. \tag{2}$$

However, there are many practical situations when the request's arrivals do not form a renewal process, that is the arrivals may depend on the number of customers, request, jobs ect staying at the service facility. This happens in the case of **finite-source queueing systems** .

Let us consider some specific examples following in order of their appearance in practice, see for example [2, 12, 15, 33, 58, 77]

Example 1

Consider a set of N machines that operate independently of each other. After a random time they may break down and need repair by one or several operatives (repairmen) for a random time. The repair is carried out by a specific discipline and after having been served each machine renew his operation. It is assumed that the server can handle only one machine at a time. Besides the usual main characteristics in reliability theory we would like to know the **distribution of the failre-free operation time** of the whole system.

Example 2

Suppose a single unloader system at which trains arrive which bring coal from various mines. There are N trains involved in the coal transport. The coal unloader can handle only one train at a time and the unloading time per train has an exponential distribution with mean $1/\mu$. The unloader is subject to breakdowns when the unloader is in operation. The operating time of an unloader has an exponential distribution with mean $1/\eta$ and the time to repair a broken unloader is exponentially distributed with mean $1/\xi$. The unloading of the train that is in service when the unloader breaks down is resumed as soon as the repair of the unloader is completed. An unloaded train returns to the mines for another trainload of coal. The time for a train to complete a trip from the unloader to the mines and back is assumed to have an exponential distribution with mean $1/\lambda$.

Example 3

N terminals request to use of a computer (server) to process transactions. The length of time that the terminal takes to generate a request for the computer is called “think” time. The length of time from the instant a terminal generates a transaction until the computer completes the transaction (and instantaneously responds by communicating this fact to the user at the terminal) is called “response time”. We would like to know, for example, the rate at which transactions are processed (which in steady-state equals the rate at which they are generated) is called “throughput”, which is one the most important performance measures showing the system’s processing power.

As we could see all the above mentioned examples have a common characteristic: We have a queueing system in which requests for service are generated by a finite number N of identical or heterogeneous sources and the requests are handled by a single or multiple server(s). The service times of the requests generated by the sources are random variables. It is assumed that the server can handle only one request at a time and uses specified service discipline. New requests for service can be generated only by idle sources, which are sources having no previous request waiting or being served at the server. A source idle at the present time will generate a request independently of the states of the other sources.

Depending on the assumptions on source, service times of the requests and the service disciplines applied at the service facility, there are many queueing models at different level to get the main steady-state performance measures. It is also easy to see, that depending on the application we can use the terms *request, customer, machine, message, job* equivalently. The above mentioned models (problems) are referred to as *machine repair, machine repairmen, machine interference, unloader problem, terminal model, or quasirandom input processes, finite population models*, respectively.

Kendall's notation

The following notation, known as Kendall's notation, is widely used to describe elementary queueing systems:

$$A/B/m/K/N \text{ – queueing discipline,}$$

where A indicates the distribution of the interarrival times, B denotes the distribution of the service times, m is the number of servers, K is the capacity of the system, that is the maximum nunber of customers staying at the facility (sometimes in the queue), and N denotes the number of sources.

The following symbols are normally used for A and B:

M	Exponential distribution (Markovian or memoryless property)
E_k	Erlang distribution with k phases
H_k	Hyperexponential distribution with k phases
C_k	Cox distribution with k phases
D	Deterministic distribution, i.e., the interarrival time or service time is constant
G	General distribution
GI	General distribution with independent interarrival times

However, due to the complexity of high-speed networks, there is considerable interest in traffic arrival processes where successive arrivals are correlated. Such non-GI arrival processes include Markov-modulated Poisson process (MMPP) , or batch Markovian arrival process (BMAP), [16, 24, 33, 40, 49].

The *queueing discipline or service strategy* determines which job is selected from the queue for processing when a server becomes available. Some commonly used queueing disciplines are:

FCFS (First-Come-First-Served): If no queueing discipline is given in the

Kendall notation, then the default is assumed to be the FCFS discipline. The jobs are served in the order of their arrival.

LCFS (Last-Come-First-Served): The job that arrived last is served next.

SIRO (Service-In-Random-Order): The job to be served next is selected at random.

RR (Round Robin): If the servicing of a job is not completed at the end of a time slice of specified length, the job is preempted and returns to the queue, which is served according to FCFS. This action is repeated until the job service is completed.

PS (Processor Sharing): This strategy corresponds to round robin with infinitesimally small time slices. It is as if all jobs are served simultaneously and the service time is increased correspondingly.

IS (Infinite Server): There is an ample number of servers so that no queue ever forms.

Static Priorities: The selection depends on priorities that are permanently assigned to the job. Within a class of jobs with the same priority, FCFS is used to select the next job to be processed.

Dynamic Priorities: The selection depends on dynamic priorities that alter with the passing of time.

Preemption: If priority or LCFS discipline is used, then the job currently being processed is interrupted and preempted if there is a job in the queue with a higher priority.

As an example of Kendall's notation, the expression

$$M/G/1 \text{ – LCFS preemptive resume (PR)}$$

describes an elementary queueing system with exponentially distributed interarrival times, arbitrarily distributed service times, and a single server. The queueing discipline is LCFS where a newly arriving job interrupts the job currently being processed and replaces it in the server. The servicing of the job that was interrupted is resumed only after all jobs that arrived after it have completed service.

$$M/G/1/K/N$$

describes a finite-source queueing system with exponentially distributed source times, arbitrarily distributed service times, and a single server. There are N request in the system and they are accepted for service iff the number of requests staying at the server is less than K. The rejected customers return to the source and start a new source time with the same distibution. It should be noted that as a special case of this situation the $M/G/1/N/N$ system could be considered. However, in this case we use the traditional $M/G/1//N$, or $\langle N/M/G/1\rangle$ notation.

It is natural to extend this notation to heterogeneous requests, too. The case when we have different requests is denotes by $\rightarrow$. So, the

$$\vec{M}/\vec{G}/1/K/N$$

denotes the above system with different rates and service times.

Performance measures

Because a queueing model represents a dynamic system, the values of the performance measures vary with time. Normally, however, we are content with the results in the steady-state. The system is said to be in *steady state* when all transient behavior has ended, the system has settled down, and the values of the performance measures are independent of time. The system is then said to be in *statistical equilibrium*, i.e., the rate at which jobs enter the system is equal to the rate at which jobs leave the system. Such a system is also called a *stable system*. Transient solutions of simple queueing systems are available in closed-form, but for more general cases, we need different techniques as described in for example, [9, 34].

The most important performance measures are:

Probability of the number of requests in the system P_k: It is often possible to describe the behaviour of a queueing system by means of the probability vector of the number of jobs in the system P_k. The mean values of most of the other interesting performance measures can be deduced from P_k:

$$P_k = P[\text{there are } k \text{ jobs in the system}]\,.$$

Utilization, or carried load ρ': If the queueing system consists of a single server, then the utilization ρ' is the fraction of the time in which the server is busy, i.e., occupied.

In case when the source is infinite and there is no limit on the number of jobs in the single server queue, the server utilization is given by:

$$\rho' = \rho = \frac{\text{mean service time}}{\text{mean inter-arrival time}} = \frac{\text{arrival rate}}{\text{service rate}} = \frac{\lambda}{\mu}\,. \tag{3}$$

The utilization of a service station with multiple servers is the mean fraction of active (or busy) servers.

In the above mentioned case since $m\mu$ is the overall service rate:

$$\rho = \frac{\lambda}{m\mu}\,, \tag{4}$$

and ρ can be used to formulate the condition for stationary behavior mentioned previously. The condition for stability is:

$$\rho < 1\,, \tag{5}$$

i.e., on average the number of jobs that arrive in a unit of time must be less than the number of jobs that can be processed.

Throughput γ: The throughput of an elementary queueing system is defined as the mean number of jobs whose processing is completed in a single unit of time, i.e., the departure rate. Since the departure rate is equal to the arrival rate λ for a queueing system in statistical equilibrium, the throughput is given by:

$$\gamma = m \cdot \rho \cdot \mu \tag{6}$$

in accordance with Eq. (4). We note that in the case of finite buffer or finite-source queueing system, throughput is usually different from the external arrival rate.

Response time T: The response time, also known as the sojourn time, is the total time that a job spends in the queueing system.

Waiting time W: The waiting time is the time that a job spends in a queue waiting to be serviced. Therefore we have:

$$\text{Response time} = \text{waiting time} + \text{service time}\,.$$

Since W and T are usually random variables, their mean should be calculated. Then:

$$\overline{T} = \overline{W} + \frac{1}{\mu}\,. \tag{7}$$

The distribution functions of the waiting time, $F_W(x)$, and the response time, $F_T(x)$, sometimes are also required.

Queue length Q: The queue length, Q, is the number of jobs in the queue.

Number of jobs in the system L: The number of jobs in the queueing system is represented by L. Then:

$$\overline{L} = E(L) = \sum_{k=1}^{\infty} kP_k \, . \tag{8}$$

The mean number of jobs in the queueing system $\overline{L}, E(L)$ and the mean queue length $\overline{Q}, E(Q)$ can be calculated using one of the most important theorems of queueing theory, **Little's theorem, (law)**:

$$\overline{L} = \gamma \overline{T} \qquad , \overline{Q} = \gamma \overline{W} \, .$$

Little's theorem is valid for all queueing disciplines and arbitrary $GI/G/m$, and $GI/G/K/N$ systems.

Performance measures for finite-source systems

- Homogeneous systems
- Asymptotic properties
- Heterogeneous systems

Homogeneous systems

For the better understanding let consider an $M/G/1$ system without server vacations treated in details in [76]. One of the performance measures in our system is the *mean message response time* $E[T]$ defined as the mean time from the arrival of a new message to its service completion, that is, the mean time a message spends in the service facility. Since the mean time that each message takes to complete cycle of staying in the source and staying in the service facility is $E[T] + 1/\lambda$, the *throughput* γ of the system, which is defined as the mean number of messages served per unit time in the whole system, is given by $N/(E[T] + 1/\lambda)$. On the other hand, if P_0 is the probability that the server is idle at an arbitary time, then $\rho' = 1 - P_0$ is the *carried load* or *server utilization*, namely, the long run fraction of the time that the server is busy. Thus, the throughput is also given by $(1 - P_0)/b$. By equating these two expressions for the throughput, we get

$$\gamma = \frac{N}{E[T] + 1/\lambda} = \frac{1 - P_0}{b} = \frac{\rho'}{b} \tag{9}$$

Hence we have

$$E[T] = \frac{Nb}{1 - P_0} - \frac{1}{\lambda} \tag{10}$$

If $E[L]$ denotes the mean number of messages in the service facility at an arbitary time, we also have the relationship

$$\gamma = \lambda(N - E[L]) \tag{11}$$

that equates the throughput to the mean number of messages arriving per unit of time. Thus we get

$$E[L] = N - \frac{1 - P_0}{\lambda b} = \gamma E[T] \tag{12}$$

which is an example of **Little's theorem** applied to those messages that are accepted by the service facility. The ratio

$$E = \frac{N - E[L]}{N} = \frac{\gamma}{N\lambda} = \frac{1 - P_0}{N\lambda b} \tag{13}$$

is called the *machine availability* in machine interference models, since it represents the expected fraction of the time that a machine remains in working condition, E is the the *machine efficiency*, because it is the ratio of the total actual production to what would have been achived had no stoppage taken place. From (9) through (11, 12), it is clear that performance measures such as ρ' , γ, $E[T]$, and $E[L]$ can be obtained once we have evaulated P_0.

Let $E[\Theta]$ be the mean length of a busy period. Since the state of the system repeats regenerative cycles of a busy period of mean length $E[\Theta]$ and an idle period of mean length $E[I] = 1/(N\lambda)$, the probability P_0 that the server is idle at an arbitary time is given by

$$P_0 = \frac{E[I]}{E[\Theta] + E[I]} = \frac{1/(N\lambda)}{E[\Theta] + 1/(N\lambda)} \tag{14}$$

If π_0 denotes the probability that the service facility is empty after a service completion, $1/\pi_0$ is the mean number of messages that are served during each busy period. This can be seen by considering a long period of time during wich a large number of (say $\mathbf{N}$) messages are served. Such a period will include $\mathbf{N}\pi_0$ busy periods on the average, because π_0 is the probability that a busy period is terminated after a service completion. Therefore, on the average $1/\pi_0$ messages are served per busy period. Hence, the mean length of a busy period is given by

$$E[\Theta] = \frac{b}{\pi_0} \tag{15}$$

From (14) and (15), we get

$$P_0 = \frac{\pi_0}{\pi_0 + N\lambda b} \tag{16}$$

Substituting (16) into (9),(10), and (12) we can express the throughput γ, the mean message response time $E[T]$, and the mean number $E[L]$ of messages in the service facility at an arbitary time in terms of π_0, too, as

$$\gamma = \frac{N\lambda}{\pi_0 + N\lambda b} \qquad ; \qquad E = \frac{1}{\pi_0 + N\lambda b} \tag{17}$$

$$E[T] = Nb - \frac{1-\pi_0}{\lambda} \tag{18}$$

$$E[L] = N\left(1 - \frac{1}{\pi_0 + N\lambda b}\right) \tag{19}$$

We can find π_0 by analyzing a Markov chain of the queue size embedded at service completion times, or the method of supplementary variables can be applied to obtain P_0.

Asymptotic properties

We can discuss some asymptotic properties of these performance measures without recourse to detailed analysis of the system state. When N is fixed, for $\lambda \approx 0$ we have almos no congestion at the service facility, which means that $\pi_0 \approx 1, P_0 \approx 1, \gamma \approx N\lambda, E \approx 1, E[T] \approx b$ and $E[L] \approx N\lambda b$. As $\lambda \to \infty$, every message whose service has just been completed returns to the facility almost immediately.
Therefore

$$\pi_0 \to 0, \quad P_0 \to 0, \quad \gamma \to 1/b,$$
$$E \to 0, \quad E[T] \to Nb, \quad E[L] \to N.$$

We note that $E[T]$ in (10) or (18) as a function of N has simple asymptotic forms. When $N = 1$ (which is equivalent to a *loss system* M/G/I/I), we obviously have $\pi_0 = 1$ and $E[T] = b$. As $N \to \infty$, we have $\pi_0 \to 0$ and so

$$E[T] \approx Nb - \frac{1}{\lambda} \qquad \text{as} \qquad N \to \infty \tag{20}$$

The value of N, denoted by N^*, at which two straight lines $E[T] = b$ and the one in (20) as a function of N intersect each other is called the *saturation number* by [42] (sec.4.12). It is given by

$$N^* = 1 + \frac{1}{\lambda b} \tag{21}$$

Note that this can be written as $N^* = (b + 1/\lambda)/b$. Therefore, if nature were kind and all messages required exactly b service time and exactly $1/\lambda$ generation time(a deterministic system), then N^* would be the maximum number of messages that could be scheduled without causing mutual interference [42] page 209.

Heterogeneous systems

In this section, we study $M/G/1/N$ systems with a heterogeneous population; that is, we assume that messages can be distinguished according to their arrival rates and service time distributions. We consider three models that differ with respect to the population constraint: an individual message model, a multiple finite-source model, and a single finite-source model. In the *individual message model*, each message has a distinct arrival rate and a distinct service time distribution. It is also called a *singel buffer model* because of its equivalence to a system of multiple classes of messages in wich each class is allotted a single buffer. In the *multiple finite-source model*, see [37] (sec. III.1), there are P classes of messages and the population size of class p is fixed at $N_p(< \infty)$ such that $N = \sum_{p=1}^{P} N_p$. The individual message model is a special case of the multiple finite-source model in which $P = N$ and $N_p = 1$ for $i \leq p \leq N$. In the *single finite source model* the total number of messages in the system is fixed at N, and each message becomes a message of one of P classes with given probability when it leaves the source.

The multiple finite-source model and the singel finite-source model may be associated with *flow control* and *congestion avoidance* mechanisms in computer communication networks. Namely, the multiple finite-source model in wich the population size is fixed for each class corresponds to the *window flow control* Let us first assume that each of N messages has different characteristics. In terms of machine interference problems, each machine is assumed to have a different breakdown rate and a different repair time ditribution. Specifically, let λ_i be the rate at wich message i in the source arrives at the service facility, and let $B_i(x)$ be the distribution function (DF) for the service time of message i, where $i = 1, 2, \ldots, N$. We also denote by b_i and $B_i^*(s)$ the mean and Laplace-Stieltjes transform (LST) of $B_i(x)$, respectively. We call this system an *individual message model*. The total arrival rate when all messages are in the source is denoted by

$$\Lambda = \sum_{i=1}^{N} \lambda_i \tag{22}$$

We denote by $E[T_i]$ the mean response time of message i, and by γ_i the throughput of message i, that is, the mean number of times that message i is served per unit time, where $i = 1, 2, \ldots, N$. These are related by

$$\gamma_i = \frac{1}{E[T_i] + 1/\lambda_i} \qquad 1 \leq i \leq N \tag{23}$$

If $\Gamma(i)$ denotes the mean number of times that message i is served in a busy period of length Θ, the throughput γ_i can also be expressed as

$$\gamma_i = \frac{\Gamma(i)}{E[\Theta] + E[I]} \qquad 1 \leq i \leq N \tag{24}$$

where

$$E[I] = \frac{1}{\Lambda} \tag{25}$$

is the mean lenght of an idle period I, and

$$E[\Theta] = \sum_{j=1}^{N} b_j \Gamma(j) \tag{26}$$

is the mean lenght of an busy period Θ. The carried load (total server utilization) ρ' is given by

$$\rho' = \frac{E[\Theta]}{E[\Theta] + E[I]} = 1 - P_0 \tag{27}$$

where P_0 is the probability that the service facility is empty at an arbitary time. The total throughput γ of the system is given by

$$\gamma = \sum_{i=1}^{N} \gamma_i = \frac{\sum_{i=1}^{N} \Gamma(i)}{E[\Theta] + E[I]} \tag{28}$$

Hence we can obtain the throughput γ_i and the mean response time $E[T_i]$ once we have calculated $\{\Gamma(j);\ 1 \leq j \leq N\}$, where $i = 1, 2, \ldots, N$. The mean waiting time of message i is given by

$$E[W_i] = E[T_i] - b_i = \frac{1}{\gamma_i} - \frac{1}{\lambda_i} - b_i \qquad 1 \leq i \leq N \tag{29}$$

If $P^{(i)}$ denotes the probability that message i is present in the service facility at an arbitary time, we have

$$P^{(i)} = \frac{E[T_i]}{E[T_i] + 1/\lambda_i} = \gamma_i E[T_i] = 1 - \frac{\gamma_i}{\lambda_i} \qquad 1 \le i \le N \tag{30}$$

which represents **Little's theorem** for message i in the service facility. In terms of machine repairman problems, $P^{(i)}$ is the probability that machine i is down at an arbitrary time. Alternatively, we can express the mean response time $E[T_i]$ and the throughput γ_i for message i in terms of $P^{(i)}$ as

$$E[T_i] = \frac{P^{(i)}}{\lambda_i(1 - P^{(i)})} \qquad ; \qquad \gamma_i = \lambda_i(1 - P^{(i)}) \tag{31}$$

In the following some important references are listed concerning **finite-source queueing models and their applications** In some of the books one can find terms, like *machine repair, machine repairmen, machine interference, unloader problem, terminal model, quasirandom input processes, finite population models*, etc.

Comprehensive books and papers: [2, 5, 9, 15, 17, 18, 19, 28, 29, 35, 37, 39, 44, 46, 48, 52, 53, 74, 75, 76, 77, 81].

Computer and communication systems: [1, 8, 16, 22, 23, 24, 34, 33, 36, 40, 41, 42, 43, 49, 51, 54, 55, 56, 59, 60, 57, 79].

Manufacturing processes: [32],.

Reliability theory: [25, 26, 27, 47, 80]

Construction and mining: [12]

Data management: [14].

It should be noted that there are many papers on **machine interference and related problems,** but our aim is to refer only the most important ones, which are closely connected to the problems or results presented in this work. Here they are: [6, 7, 10, 11, 13, 20, 30, 31, 38, 45, 50, 58, 78, 82, 83].

The main aim of the following chapters is to show how different methods can be applied in the investigation of finite-source queueing systems. Thus, analytical, numerical and asymptotic approaches are presented and in most cases numerical results illustrate the problem in question. Furthermore, the most important sources of information are listed to draw attention of the interested readers. Finally, some of the works of the author is either presented or cited.

Analytical Results

Homogeneous $M/M/r$ systems, the classical model

This section presents the classical queuing theory approach to solving a machine interference problem. It should be noted that this system is analyzed by many authors in different books. It is a classical example for queueing systems with state-dependent arrival rates and it can be treated in the framework of the so-called *birth-and-death processes*. The present problem is descibed in several classical books on queueing systems, for example [2, 12, 15, 41, 29, 33, 79] suct to mention the basic ones. Our aim is to show the form of the steady-state probabilities of stopped machines. In the above mentioned works one can find the detailed analysis of waiting time, down time distibution of machines, too. Several numerical examples from real life situations illustrates this interesting system.

It is also proved that in steady-state the arriving machines's distribution in system containing N machines is the same as the outside observer's distribution for the corresponding system with $N-1$ machines, or other words in arrival epochs the distribution is the same as the time-average distribution of system with one less machine.

The assumptions of the model are as follows:

Suppose that there are N machines and r operators and

1. The time between breakdowns (or production time) of any one of the machines is a sample from a *negative exponential* probability distribution with mean $1/\lambda$, (or mean rate λ). A breakdown is random and is independent of the operating behavior of the other machines. Then, when there are n machines not working at time t,

   Prob (one of the $N-n$ machines goes down in the interval $(t, t+\Delta t)) =$ $(N-n)\lambda\Delta t + o(\Delta t)$,

   where Δt is a small increment of time.

2. Any one of the n down machines requires only one of the r operators to fix it. The service time distribution is *negative exponential* with mean

$1/\mu$ for each machine and each operator. The service times are mutually independent and also independent of the number of down machines.

Then

Prob [one of the n down machines is fixed in an interval Δt]

$$= \begin{cases} n\mu\Delta t + o(\Delta t), & \text{for } 1 \leq n \leq r \\ r\mu\Delta t + o(\Delta t), & \text{for } r < n \leq N \end{cases}$$

3. The machines are *served in the order od their beakdowns*.

Let

$$\text{L(t)} = \text{the number of down machines at time t}$$

and

$$P_n(t) = Prob(L(t) = n | L(0) = i), \quad n = 0, \ldots, N.$$

Then the stochastic process, $(L(t), t \geq 0)$, is a birth-and-death process, with rates

$$\lambda_n = \begin{cases} (N-n)\lambda, & n = 0, 1, \ldots, N \\ 0, & n > N \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & n = 1, 2, \ldots, r \\ r\mu, & n = r+1, \ldots, N \end{cases}$$

The forward Kolmogorov-equations of the birth-death process are

$$P_0'(t) = N\lambda P_0(t) + \mu P_1(t)$$

$$P_n'(t) = -\{(N-n)\lambda + n\mu\}P_n(t) + (N-n+1)\lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t),$$
$$1 \leq n < r$$

$$P_n'(t) = -\{(N-n)\lambda + r\mu\}P_n(t) + (N-n+1)\lambda P_{n-1}(t) + r\mu P_{n+1}(t),$$
$$r \leq n < N$$

$$P_N'(t) = -r\mu P_N(t) + \lambda P_{n-1}(t)$$

This finite system of ordinary differential equations can be solved and we get the transient probabiliies.

For the equilibrium values of P_n by setting these derivatives equal to zero while noting that the equilibrium (or stationary or steady state) values are

$$P_n = \lim_{t \to \infty} P_n(t)$$

The flow balance equations (steady-state equations) become

$$\begin{aligned}
N\lambda P_0 &= \mu P_1 \\
\{(N-n)\lambda + n\mu\} P_0 &= (N-n+1)\lambda P_{n-1} + (n+1)\mu P_{n+1}, \quad 1 < n < r \\
\{(N-n)\lambda + r\mu\} P_0 &= (N-n+1)\lambda P_{n-1} + r\mu P_{n+1}, \quad r \le n < N \\
r\mu P_N &= \lambda P_{N-1}
\end{aligned}$$

These equations are solved recursively using the relationship

$$(N-n)\lambda P_n = (n+1)\mu P_{n+1}, \quad 0 \le n < r$$
$$(N-n)\lambda P_n = r\mu P_{n+1}, \quad r \le n < N.$$

Letting $\rho = \lambda/\mu$ (the servicing factor), the steady-state probabilities are

$$P_n = \binom{N}{n}\left(\frac{\lambda}{\mu}\right)^k \hat{P}_0 \qquad \text{for } 0 \le n \le r, \tag{32}$$
$$P_n = \frac{N!}{(N-n)!r!r^{n-r}}\left(\frac{\lambda}{\mu}\right)^n P_0 \qquad \text{for } r \le n \le N.$$

where P_0 is obtained by solving $\sum_{n=0}^{N} P_n = 1$ to get

$$P_0 = \left(\sum_{n=0}^{r} \binom{N}{n} \rho^n + \sum_{n=r+1}^{N} \binom{N}{n} \frac{n!}{r! r^{n-r}} \rho^n \right)^{-1}$$

In the following only the **main performace measures** characteristic of machine interference problem are mentioned.

1. **The expected (average) number of down machines** is

$$E[L] = \sum_{n=0}^{N} n P_n$$

   It can be seen that there is no closed-form expression for $E(L)$ in general, but for a particular problem (system), $E[L]$ is easily computed. There is a closed-form expression for single-server (only one operator) systems. In this case

$$E[L] = N + \frac{\lambda + \mu}{\lambda}(1 - P_0)$$

2. **Machine efficiency or machine utilization**

$$U_m = \frac{N - E[L]}{N}$$

or percentage of average production obtained (or the fraction of total production time on all machines).

3. **Average operator utilization**

$$U_s = \sum_{n=0}^{N} \frac{nP_n}{r} + \sum_{n=r+1}^{N} P_n$$

or fraction of time an operator would be working.

4. **Average number of idle operators**

$$r - rU_s = \sum_{n=0}^{r} (r-n)P_n$$

5. **Average number of machines waiting**

$$\overline{Q} = \sum_{n=r+1}^{N} (n-r)P_n$$

6. **Average down time of machines**

$$\overline{T} = \frac{E(L)}{\lambda(N - E(L))}$$

7. **Mean waiting time of machines**

$$\overline{W} = \frac{\overline{Q}}{\lambda(N - E(L))}$$

By dividing measure 4 by the number of operators, r, and measure 5 by the number of machines, N, some related measures are

- **Coefficient of loss for operator**

$$\frac{\sum_{n=0}^{N}(r-n)P_n}{r}$$

  or percentage of idle operators.

- **Coefficient of loss for machines**

$$\frac{\sum_{n=r+1}^{N}(n-r)P_n}{N}$$

or percentage of interference time.

The purpose of the following example is to show the advantages obtained in system performances and productivity from the pooling of operators. In this case several operators have the same assignment of machines.

Table 1 has values for operator utilization for pairs of (N, r) parameters that have the same machine per operator ratio $(N/r = 4$ and then 15).

Notice that the operator utilization is increasing for a given ρ even though the ratio of the number of machines per operator stays the same. This is an indication that it is better, when feasible, to pool operators rather than to assign a particular number of machines to each operator individually. The example considers two cases: (1) 6 machines serviced by one operator and (2) 20 machines serviced by three operators. The results show that, even though the workload per operator increased from system 1 (6 machines/operator) to system 2 ($6\frac{2}{3}$ machines/operator), the machines were serviced more efficiently in system 2. The advantages of pooling are well known.

ρ	N	r	U_s
0.45	4	1	0.881
	8	2	0.934
	16	4	0.994
0.05	15	1	0.656
	30	2	0.682
	60	4	0.705

Table 1: Operator utilizations for proportional parameters

Notice in Table 1 that, for a given number of machines per operator (assuming N/r is integer), as r (and therefore N) increases, the operator utilization slowly increases. Likewise, under the same conditions, the machine efficiency will slowly increase.

Using the method of supplementary variables similar problems were treated in [63, 64, 65, 66, 67].

Numerical Methods

- A recursive method for the $M/G/1/$ system
  - ⋆ The mathematical model
  - ⋆ Numerical results

A recursive method for the $M/G/1$ system

Closed-form solutions for the steady-state probabilities are very seldom. Different analytical methods are used to investigate the involved processes and related numerical problems. For the most common procedures and tools the interested reader is referred to [8, 24, 31, 33, 34, 49, 54, 60, 77, 81]

In the following the results of [31] are introduced and some numerical examples are demonstrated. For ready use of numerical results a book of tables has been produced by [53] for $M/M/r$ model. A collection of various theoretical results along with numerical work can also be found in [12] where he has discussed in detail its application in construction and mining. Takács [74] gives the explicit expression for the distribution of the number of working (up) machines of the $(M/G/1)$ model. However, for a large number of machines the computation of probabilities using Theorem 2 in [74] (p. 195) may pose problem as it involves many factorials. Even for the simple model $M/M/C$, Gross and Harris [29] (p. 108) makes similar comments and proposed a recursive method for computing probabilities. To obtain the steady state probability distribution of the number of down machines at arbitrary time epoch $P_n(0 \le n \le N)$ one can also use the embedded Markov chain technique, see [76]. The objective of this section is to provide an alternative method, using the supplementary variable technique and considering the supplementary variable as the remaining repair time, to obtain $P_n(0 \le n \le N)$ for $M/G/1$ model which is used to obtain the various system performance measures such as average number of down machines, average waiting time and operator utilization etc. The method is recursive and can be used for any repair time distribution such as mixed generalized Erlang (MGE_h), generalized Erlang (GE_h), hyperexponential (HE_h), generalized hyperexponential (GH_h) and uniform $U(a,b)$ etc. The only input required for efficient evaluation of state probabilies is the LST of the repair time distribution.

The mathematical model

Consider a machine repairman problem with a single repairman and a set of N working machines Let us assume that the running times of the machines between breakdowns have an exponential distribution with mean $1/\lambda$ and the repair (service) time of the machines are independently identically distributed random variables (i.i.d.r.v's) having distribution function $B(u)$, a probability density function (p.d.f.) $b(u)$ and a mean repair time b_1. The state of the system at time t is given by

N(t)=Number of down machines, and
U(t)=Remaining repair time for the machine under repair.

Let us define

$$P_0(t) = P(N(t) = 0), \tag{33}$$

and

$$P_n(u,t)du = P\{N(t)=n, u < U(t) \le u+du\}, \quad u \ge 0, \quad n = 1,2,\dots,N. \tag{34}$$

$$P_n(t) = P(N(t)=n) = \int_0^\infty P_n(u,t)du, \quad n = 1,2,\dots,N. \tag{35}$$

Relating the states of the system at time t and $t+dt$, we obtain

$$\frac{\partial}{\partial t}P_0(t) = -N\lambda P_0(t) + P_1(0,t), \tag{36}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right)P_1(u,t) = -(N-1)\lambda P_1(u,t) + N\lambda P_0(t)b(u) + + P_2(0,t)b(u), \tag{37}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right)P_r(u,t) = -(N-r)\lambda P_r(u,t) + (N-r+1)\lambda P_{r-1}(u,t) +$$

$$+ P_{r-1}(0,t)b(u), \qquad 2 \le r \le N-1 \tag{38}$$

$$\left(\frac{\partial}{\partial t} - \frac{\partial}{\partial u}\right) P_N(u,t) \quad = \quad \lambda P_{N-1}(u,t). \tag{39}$$

Since we discuss the model in steady state, we let $t \to \infty$ in equations (36)-(39).

Further define

$$P_n = \lim_{t\to\infty} P_n(t), \quad 0 \le n \le N \tag{40}$$

$$P_n(u) = \lim_{t\to\infty} P_n(u,t), \quad 1 \le n \le N. \tag{41}$$

$$B^*(s) = \int_0^\infty e^{-su} dB(u) = \int_0^\infty e^{-su} b(u)du, \tag{42}$$

$$\left.\begin{array}{ll} P_n^*(s) = \int_0^\infty e^{-su} P_n(u)du & 1 \le n \le N \\[2ex] P_n = P_n^*(0) = \int_0^\infty P_n(u)du, & 1 \le n \le N. \end{array}\right\} \tag{43}$$

and

$$\int_0^\infty e^{-su} \frac{\partial}{\partial u} P_n(u) du = s P_n^*(s) - P_n(0). \quad (44)$$

From (36)-(44) and the fact that all derivatives with respect to t are zero, it follows that

$$N\lambda P_0 = P_1(0), \quad (45)$$

$$[(N-1)\lambda - s]P_1^*(s) = N\lambda P_0 B^*(s) + P_2(0)B^*(s) - P_1(0), \quad (46)$$

$$[(N-r)\lambda - s]P_r^*(s) = (N-r+1)\lambda P_{r-1}^*(s) + P_{r+1}(0)B^*(s) - P_r(0),$$

$$2 \le r \le N-1 \quad (47)$$

$$-sP_N^*(s) = \lambda P_{N-1}^*(s) - P_N(0). \quad (48)$$

Using (45) in (46) and then adding (46) to (48), we obtain

$$\sum_{r=1}^{N} P_r^*(s) = \frac{1 - B^*(s)}{s} \sum_{r=1}^{N} P_r(0). \tag{49}$$

Taking $s \to 0$ in (49), we get

$$\sum_{r=1}^{N} P_r^*(0) = b_1 \sum_{r=1}^{N} P_r(0) \tag{50}$$

where $b_1 = -B^{*(1)}(0)$ is mean repair time.

Our main objective is to obtain $P_n \equiv P_n^*(0)(1 \leq n \leq N)$ from (45 -(48). To achieve it, our strategy will be to obtain first $P_n(0)(1 \leq n \leq N)$ and then using it we finally evaluate $P_n^*(0)(1 \leq n \leq N)$.

Using (45) in (46) and then setting $s = (N-1)\lambda$ and $s = 0$ respectively in

(46), we get

$$P_2(0) = \frac{N\lambda[1 - B^*((N-1)\lambda)]}{B^*((N-1)\lambda)} P_0, \tag{51}$$

and

$$P_1^*(0) = \frac{1}{(N-1)\lambda} P_2(0). \tag{52}$$

Now setting $s = (N-r)\lambda$, in (47), we obtain

$$\begin{aligned} P_{r+1}(0) \quad &= \quad \frac{1}{B^*((N-r)\lambda)}[P_r(0) - (N-r+1)\lambda P_{r-1}^*((N-r)\lambda)], \\ &\quad\; 2 \leq r \leq N-1. \end{aligned} \tag{53}$$

Setting $s = (N-r)\lambda$ in (46) for $r = 2, 3, \ldots, N-1$, we get

$$\begin{aligned} P_1^*((N-r)\lambda) &= \frac{1}{(r-1)\lambda}[N\lambda P_0\{B^*((N-r)\lambda)-1\}+ \\ &\quad + P_2(0)B^*((N-r)\lambda)]. \end{aligned} \tag{54}$$

From equation (47) for $r = 3, 4, \ldots, N-1$, we get

$$\begin{aligned} P_j^*((N-r)\lambda) &= \frac{1}{(r-j)\lambda}[(N-j+1)\lambda P_{j-1}^*((N-r)\lambda)+ \\ &\quad + P_{j+1}(0)B^*((N-r)\lambda)-P_j(0)], \\ &\quad 2 \le j \le r-1. \end{aligned} \tag{55}$$

Hence $P_3(0), P_4(0), \ldots, P_N(0)$ can be obtained recursively using (51), (54), (55) and (53) in terms of P_0.

Now setting $s = 0$ in (47), we get

$$\begin{aligned} P_r^*(0) &= \frac{1}{(N-r)\lambda}[(N-r+1)\lambda P_{r-1}^*(0) + P_{r+1}(0) - P_r(0)], \\ & 2 \leq r \leq N-1. \end{aligned} \quad (56)$$

As $P_2(0), P_3(0), \ldots, P_N(0)$ are known, $P_2^*(0), P_3^*(0), \ldots, P_{N-1}(0)$ can be determined recursively using (52) and (56) in terms of P_0.

Now the only unknown quantity is $P_N^*(0)$ which can be obtained from equation (48). To obtain it, differentiate equation (48) with respect to s and set $s = 0$, we get

$$P_N^*(0) = -\lambda P_{N-1}^{*(1)}(0). \quad (57)$$

To get $P_{N-1}^{*(1)}(0)$, differentiate (47) and (46) with respect to s and set $s=0$.

$$\begin{aligned} P_r^{*(1)}(0) &= \frac{1}{(N-r)\lambda}[(N-r+1)\lambda P_{r-1}^{*(1)}(0) \\ &\quad + P_{r+1}(0)B^{*(1)}(0) + P_r^*(0)], \qquad 2 \le r \le N-1 \qquad (58) \\ P_1^{*(1)}(0) &= \frac{1}{(N-r)\lambda}[N\lambda P_0 B^{*(1)}(0) + P_2(0)B^{*(1)}(0) + P_1^*(0)]. \quad (59) \end{aligned}$$

As $P_1^{*(1)}(0)$ is known completely from (59), $P_r^{*(1)}(0), (2 \le r \le N-1)$ can be determined recursively from (58) and hence $P_N^*(0)$ is known from (57). So $P_n^*(0)(1 \le n \le N)$ is known in terms of P_0, which can be determined using the normalizing condition

$$P_0 + \sum_{n=1}^{N} P_n^*(0) = 1. \qquad (60)$$

The steady state probability distribution of the number of down machines at service completion or departure epoch $\pi_n (0 \leq n \leq N-1)$ can also be obtained from $P_r(0) (1 \leq r \leq N)$ and is given by

$$\pi_n = \frac{P_{n+1}(0)}{\sum_{r=1}^{N} P_r(0)}, \quad n = 0, 1, \ldots, N-1. \tag{61}$$

To demonstrate the working of the method we consider analytically a simple example where the repair time distribution is exponential and number of machines (N) is four i.e. $M/M/1//4$ model. In this case

$$B^*(s) = \frac{\mu}{\mu + s}$$

From (51), we get

$$P_2(0) = \frac{4\lambda[1 - B^*(3\lambda)]}{B^*(3\lambda)} P_0.$$

Now from (53), we have

$$P_3(0) = \frac{1}{B^*(2\lambda)}[P_2(0) - 3\lambda P_1^*(2\lambda)],$$

where $P_1^*(2\lambda)$ is obtained from (54)

$$P_1^*(2\lambda) = \frac{1}{\lambda}[4\lambda P_0\{B^*(2\lambda) - 1\} + P_2(0)B^*(2\lambda)].$$

Now again from (53), we have

$$P_4(0) = \frac{1}{B^*(\lambda)}[P_3(0) - 2\lambda P_2^*(\lambda)].$$

where $P_2^*(\lambda)$ is obtained from (55)

$$P_2^*(\lambda) = \frac{1}{\lambda}[3\lambda P_1^*(\lambda) + P_3(0)B^*(\lambda) - P_2(0)].$$

To know $P_2^*(\lambda)$ we require $P_1^*(\lambda)$ which can be obtained from (54)

$$P_1^*(\lambda) = \frac{1}{2\lambda}[4\lambda P_0\{B^*(\lambda) - 1\} + P_2(0)B^*(\lambda)].$$

From above we get

$$P_2(0) = 12\tfrac{\lambda^2}{\mu}P_0,$$

$$P_1^*(2\lambda) = \tfrac{4\lambda}{\mu+2\lambda}P_0, \quad P_3(0) = 24\tfrac{\lambda^3}{\mu^2}P_0, \quad P_1^*(\lambda) = \tfrac{4\lambda}{\mu+\lambda}P_0,$$

$$P_2^*(\lambda) = 12\tfrac{\lambda^2}{\mu(\mu+\lambda)}P_0, \quad P_4(0) = 24\tfrac{\lambda^4}{\mu^3}P_0.$$

Hence from (52) and (56), we get

$$P_1^*(0) = \frac{1}{3\lambda} P_2(0) = \frac{4\lambda}{\mu} P_0,$$

$$P_2^*(0) = \frac{1}{2\lambda} P_3(0) = 12 \frac{\lambda^2}{\mu^2} P_0,$$

$$P_3^*(0) = \frac{1}{\lambda} P_4(0) = 24 \frac{\lambda^3}{\mu^3} P_0.$$

Finally to determine $P_4^*(0)$ we have from (57)

$$P_4(0) = -\lambda {P_3^*}^{(1)}(0),$$

where ${P_3^*}^{(1)}(0)$ can be obtained from (58)

$${P_3^*}^{(1)}(0) = \frac{1}{\lambda} [2\lambda {P_2^*}^{(1)}(0) + P_4(0) B^{*(1)}(0) + P_3^*(0)],$$

again $P_2^{*(1)}(0)$ an be obtained from (58)

$$P_2^{*(1)}(0) = \frac{1}{2\lambda}[3\lambda P_1^{*(1)}(0) + P_3(0)B^{*(1)}(0) + P_2^*(0)],$$

To know $P_2^{*(1)}(0)$ we require $P_1^{*(1)}(0)$ which can be obtained from (59)

$$P_1^{*(1)}(0) = \frac{1}{3\lambda}[4\lambda P_0 B^{*(1)}(0) + P_2(0)B^{*(1)}(0) + P_1^*(0)].$$

From above we get

$$P_1^{*(1)}(0) = -4\frac{\lambda}{\mu^2}P_0, P_2^{*(1)}(0) = -12\frac{\lambda^2}{\mu^3}P_0, P_3^{*(1)}(0) = -24\frac{\lambda^3}{\mu^4}P_0,$$

and hence

$$P_4^*(0) = 24\frac{\lambda^4}{\mu^4}P_0$$

or

$$P_1^*(0) = 4\rho P_0, P_2^*(0) = 12\rho^2 P_0, P_3^*(0) = 24\rho^3 P_0, P_4^*(0) = 24\rho^4 P_0.$$

where $\rho = \frac{\lambda}{\mu}$.

Since $P_0 + P_1^*(0) + P_2^*(0) + P_3^*(0) + P_4^*(0) = 1$, we get

$$P_0 = \frac{1}{1 + 4\rho + 12\rho^2 + 24\rho^3 + 24\rho^4}.$$

It can be easily seen that this result matches with the expression given in [29] p. 105.

The system performance measures such as average and standard deviation of the number of down machines in the system $(\overline{N})$ and (SDL), average number of machines waiting for repair in queue (L_q), average waiting time in the system $(\overline{T})$, average waiting time in queue $(\overline{W})$, operator utilization (U) and the average number of operating machines (AOP) can be obtained from the probability distribution of the number of down machines at arbitrary time epoch and are given by:

$$\overline{N} = \sum_{n=0}^{N} n.P_n, \; SDL = \sqrt{\sum_{n=0}^{N} n^2.P_n - L^2},$$

$$L_q = \sum_{n=1}^{N} (n-1).P_n, \; \overline{T} = \frac{\overline{N}}{\lambda'}, \; \overline{W} = \frac{L_q}{\lambda'},$$

Here $\lambda' = \lambda(N - \overline{N})$ is an *effective arrival rate* into the system, sometimes called *throughput* denoted by γ

$$U = 1 - P_0 \quad and \quad AOP = N - \overline{N}.$$

Numerical results

To demonstrate the working of the method proposed above carried out extensive numerical work on "CYBER 180/840A computer system" for variety of repair time distributions such as M, E_h, D, HE_2, GE_4 and uniform $U(a,b)$ but only a few are presented here. The probability distribution of the number of down machines (P_n) along with system performance measures have been presented in the self explanatory Table 2 for $N = 5$ and different values of traffic intensity $\rho(=\lambda b_1)$. Table 3 shows the probability distribution of the number of down machines at an arbitrary time epoch P_n and the departure time epoch π_n for E_{10} when $\rho = 0.5$ and $N = 5$. In the fourth column of this Table, P_n is again obtained from π_n using a relation given in [38] by Jeyachandra and Shanthikumar A computer program has been written based on a method of [74] for comparisons. It should be noted that Takács obtains the distribution of the number of up machines i.e. $Q_n(0 \le n \le N)$, whereas we obtain the distribution of the number of down machines. Both are equal if P_n is compared with Q_n in reverse order i.e.
$P_0 = Q_N, P_1 = Q_{N-1} \dots P_N = Q_0$ The difficulty encountered in using

Takács result is that if fails for large N whereas the proposed method works without any difficulty. The results has also been tested with those given [12] for operator utilization in case of E_5, E_{10} and D with varying ρ and N. Effect of ρ on U and $\overline{W}$ for fixed $N = 5$ have been shown in Tables 4 and 5, respectively. It is seen that as ρ increases both U and $\overline{W}$ also increase. But for the same value of ρ, U for H_2 is less than U of M, E_{10} and D. It is also observed that $\overline{W}$ for H_2 is more than $\overline{W}$ of M, E_{10} and D. However for $\rho \geq 0.5$, $\overline{W}$ is almost same for all the repair time distributions. The effect of the number of machines (N) on U and $\overline{W}$ for the fixed value of $\rho = 0.3$ is given in Table 6. As N increases both U and $\overline{W}$ also increases irrespective of the repair time distributions but for $N \geq 12, U$ remains same for all the repair time distributions. The same is also true for $\overline{W}$.

P(n)	M ρ=0.5	E_{10} $\rho = 0.7$	D ρ=0.7	H_2 ($\mu_1 = 1.41$, ($\mu_2 = 5.26$) ($\alpha_1 = 0.21$, $\alpha_2 = 0.79$) $\rho = 0.3$)	0 (a = 0.01, b = 0.09) $\rho = 0.9$	GE_4 ($\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$, $\mu_4 = 4$) $\rho = 0.3$
P(0)	0.36697E-01	0.11763E-02	0.62725E-03	0.16141E+00	0.50140E-03	0.10867E+00
P(1)	0.91743E-01	0.15888E-01	0.12110E-01	0.19939E+00	0.68133E-02	0.24016E+00
P(2)	0.18349E+00	0.10269E+00	0.97234E-01	0.21607E+00	0.53685E-01	0.30048E+00
P(3)	0.27523E+00	0.31946E+00	0.33197E+00	0.19791E+00	0.23300E+00	0.23267E+00
P(4)	0.27523E+00	0.41045E+00	0.42045E+00	0.14666E+00	0.45373E+00	0.10034E+00
P(5)	0.13671E+00	0.15033E+00	0.13760E+00	0.78559E-01	0.25227E+00	0.17680E-01
Sum	0.10000E+01	0.10000E+01	0.10000E+01	0.10000E+01	0.10000E+01	0.10000E+01
$\overline{N}$	3.07339	3.57310	3.57230	2.20470	3.88940	2.02890
SDL	1.30424	0.92744	0.89142	1.51480	0.87103	1.21130
$\underline{Lq}$	2.11009	2.57430	2.57300	1.36610	2.88990	1.13760
$\overline{T}$	1.59520	2.50410	2.50220	0.78872	0.19457	4.74220
$\overline{W}$	1.09520	1.80410	1.80220	0.48872	0.14457	2.65890
U	0.96330	0.99882	0.99937	0.83859	0.99950	0.89133
AOP	1.92660	1.42690	1.42770	2.79530	1.11060	2.97110

Table 2: The probability distribution of the number of down machines and system performance measures for $N = 5$

n	P_n GR method	π_n GR method	P_n (Using Jeyachandra & Shanthikumar relation)
0	$0.10530E{-}01$	$0.26605E{-}01$	$0.10530E{-}01$
1	$0.68335E{-}01$	$0.13812E{+}00$	$0.68335E{-}01$
2	$0.21794E{+}00$	$0.33040E{+}00$	$0.21794E{+}00$
3	$0.36235E{+}00$	$0.36621E{+}00$	$0.36235E{+}00$
4	$0.27442E{+}00$	$0.13867E{+}00$	$0.27442E{+}00$
5	$0.66432E{-}01$		$0.66423E{-}01$
Sum	$0.10000E{+}01$	$0.10000E{+}01$	$0.10000E{+}01$

Table 3: The probability distribution of the number of down machines at arbitrary time epoch and departure time epoch for $N = 5, \rho = 0.5E_{10}$

ρ	M	E_{10}	D	H_2
0.10	0.43605	0.44309	0.44398	0.43049
0.20	0.71513	0.74594	0.75042	0.69688
0.30	0.86079	0.90401	0.91046	0.83859
0.40	0.93027	0.96763	0.97254	0.91112
0.50	0.96330	0.98947	0.99216	0.94889
0.60	0.97961	0.99654	0.99779	0.96929
0.70	0.98808	0.99882	0.99937	0.98080
0.80	0.99270	0.99958	0.99982	0.98755
0.90	0.99535	0.99985	0.99995	0.99167
1.00	0.99693	0.99994	0.99998	0.99427
1.10	0.99791	0.99998	0.99999	0.99596
1.20	0.99854	0.99999	1.00000	0.99708

Table 4: Effect of ρ on $U(N = 5)$

ρ	M	E_{10}	D	H_2
0.10	0.04666	0.02843	0.02618	0.06173
0.20	0.19834	0.14059	0.13258	0.23496
0.30	0.44258	0.35928	0.34753	0.48872
0.40	0.74992	0.66690	0.65648	0.79509
0.50	1.10952	1.10266	1.10198	1.11347
0.60	1.14624	1.14104	1.14066	1.14950
0.70	1.18422	1.18041	1.18022	1.18685
0.80	1.22294	1.22017	1.22007	1.22504
0.90	1.26210	1.26007	1.26002	1.26378
1.00	1.30154	1.30003	1.30001	1.30288
1.10	1.34115	1.34001	1.34000	1.34223
1.20	1.38080	1.38001	1.38000	1.38175

Table 5: Effect of ρ on $\overline{W}(N=5)$

N		M	E_{10}	D	H_2
2	U	0.42830	0.44640	0.447490	0.432830
	$\overline{W}$	0.06923	0.04409	0.040818	0.086241
4	U	0.75742	0.79439	0.799950	0.737780
	$\overline{W}$	0.28432	0.21060	0.200100	0.326500
6	U	0.92821	0.96531	0.970200	0.908050
	$\overline{W}$	0.63921	0.56469	0.555290	0.682270
8	U	0.98641	0.99818	0.998880	0.976830
	$\overline{W}$	1.11331	1.11044	1.110270	1.115690
10	U	0.99833	0.99997	0.999990	0.995860
	$\overline{W}$	1.17050	1.17000	1.170000	1.171250
12	U	0.99986	1.00000	1.000000	0.999470
	$\overline{W}$	1.23005	1.23000	1.230000	1.230190
14	U	0.99999	1.00000	1.000000	0.999950
	$\overline{W}$	1.29000	1.29000	1.290000	1.290020
16	U	1.00000	1.00000	1.000000	1.000000
	$\overline{W}$	1.35000	1.35000	1.350000	1.350000
18	U	1.00000	1.00000	1.000000	1.000000
	$\overline{W}$	1.41000	1.41000	1.410000	1.410000
20	U	1.00000	1.00000	1.000000	1.000000
	$\overline{W}$	1.47000	1.47000	1.470000	1.470000

Table 6: Effect of number of machines on operator utilization and average waiting time in queue $\rho = 0.3$

This system has been generalized to $\vec{M}/\vec{G}/1/FIFO$ system which can be found in [61, 62].

Asymptotic Methods

- Preliminary results
- Heterogeneous multiprocessor systems
  - ⋆ The queueing model
  - ⋆ Performance measures
  - ⋆ Numerical results

Preliminary results

In this section a brief survey is given of the most related theoretical results due to Anisimov [3, 5], to be applied later on.

Let $(X_\epsilon(k), k \geq 0)$ be a Markov chain with state space

$$\bigcup_{q=0}^{m+1} X_q, \qquad X_i \cap X_j = 0,\ i \neq j,$$

defined by the transition matrix $\left\|p(i,j)\right\|$ satisfying the following conditions:

1. $p_\epsilon(i^{(0)}, j^{(0)}) \to p_0(i^{(0)}, j^{(0)})$, as $\epsilon \to 0$,$i^{(0)}, j^{(0)} \in X_0$, and $P_0 = \left\|p_0(i^{(0)}, j^{(0)})\right\|$ is irreducible;
2. $p_\epsilon(i^{(q)}, j^{(q+1)}) = \epsilon\alpha^{(q)}(i^{(q)}, j^{(q+1)}) + o(\epsilon), i^{(q)} \in X_q,\ j^{(q+1)} \in X_{q+1}$

3. $p_\epsilon(i^{(q)}, f^{(q)}) \longrightarrow 0$, as $\epsilon \longrightarrow 0, i^{(q)}, f^{(q)} \in X_q,\ q \geq 1$;

4. $p_\epsilon(i^{(q)}, f^{(z)}) \equiv 0,\ i^{(q)} \in X_q, f^{(x)} \in X_z,\ z - q \geq 2$

In the sequel the set of states X_q is called the q-th level of the chain, $q = 1, \ldots, m+1$. Let us single out the subset of states

$$\langle \alpha_m \rangle = \bigcup_{q=0}^{m} X_q$$

Denote by $\{\pi_\epsilon(i^{(q)}), i^{(q)} \in X_q\},\ q = 1, ..., m$ the stationary distribution of a chain with transition matrix

$$\left\| \frac{p_\epsilon(i^{(q)}, j^{(z)})}{1 - \sum_{k^{(m+1)} \in X_{m+1}} p_\epsilon(i^{(q)}, k^{(m+1)})} \right\|, i^{(q)} \in X_q, j^{(z)} \in X_z,\ q, z \leq m,$$

Furthermore denote by $g_\epsilon(\langle \alpha_m \rangle)$ the steady state probability of exit from

$\langle \alpha_m \rangle$, that is

$$g_\epsilon(\langle \alpha_m \rangle) = \sum_{i^{(m)} \in X_m} \pi_\epsilon(i^{(m)}) \sum_{j^{(m+1)} \in X_{m+1}} p_\epsilon(i^{(m)}, j^{(m+1)})).$$

Denote by $\{\pi_0(i^{(0)}), i^{(0)} \in X_0\}$ the stationary distribution corresponding to P_0 and let

$$\overline{\pi_0} = \{\pi_0(i^{(0)}), i^{(0)} \in X_0\}, \ \overline{\pi_\epsilon}^{(q)} = \{\pi_\epsilon(i^{(q)}), i^{(q)} \in X_q\}$$

be row vectors. Finally, let

$$A^{(q)} = \left\| \alpha^{(q)}(i^{(q)}, j^{(q+1)}) \right\|, \ i^{(q)} \in X_q, \ j^{(q+1)} \in X_{q+1}, q = 0, \ldots, m$$

defined by Condition 2.

Conditions (1)-(4) enables us to compute the main terms of the asymptotic expression for $\overline{\pi_\epsilon}^{(q)}$ and $g_\epsilon(\langle\alpha_m\rangle)$. Namely, we obtain

$$\begin{aligned} \overline{\pi_\epsilon}^{(q)} &= \epsilon^q \overline{\pi_0} A^{(0)} A^{(1)} \dots A^{(q-1)} + o(\epsilon^q) \; q = 1, \dots, m, \\ g_\epsilon(\langle\alpha_m\rangle) &= \epsilon^{m+1} \overline{\pi}_0 A^{(0)} A^{(1)} \dots A^{(m)} \underline{1} + o(\epsilon^{m+1}), \end{aligned} \tag{62}$$

where $\underline{1} = (1, ..., 1)^*$ is a column vector, see Anisimov *et al.* [5] pp. 141-153. Let $(\eta_\epsilon(t), t \geq 0)$ be a Semi-Markov Process (SMP) given by the embedded Markov chain $(X_\epsilon(k), k \geq 0)$ satisfying conditions (1)-(4). Let the times $\tau_\epsilon(j^{(s)}, k^{(z)})$ – transition times from state $j^{(s)}$ to state $k^{(z)}$ – fulfill the condition

$$\mathbf{E} \exp\{i\Theta\beta_\epsilon\tau_\epsilon(j^{(s)}, k^{(z)})\} = 1 + a_{jk}(s, z, \Theta)\epsilon^{m+1} + o(\epsilon^{m+1}), \; (i^2 = -1)$$

where β_ϵ is some normalizing factor. Denote by $\Omega_\epsilon(m)$ the instant at which the SMP reaches the $(m+1)$-th level for the first time, exit time from $\langle\alpha_m\rangle$ provided $\eta_\epsilon(0) \in \langle\alpha_m\rangle$.Then we have:

Theorem 1. *[cf. [5] pp. 153] If the above conditions are satisfied then*

$$\lim_{\epsilon \longrightarrow 0} \mathbf{E} \exp\{i\Theta\beta_\epsilon\Omega_\epsilon(m)\} = (1 - A(\Theta))^{-1},$$

where

$$A(\Theta) = \frac{\sum\limits_{j^{(0)},k^{(0)} \in X_0} \pi_0(j^{(0)})p_0(j^{(0)},k^{(0)})a_{jk}(0,0,\Theta)}{\overline{\pi}_0 A^{(0)} A^{(1)} \ldots A^{(m)} \underline{1}}$$

Corollary 1. *In particular, if* $\alpha_{jk}(s,z,\Theta) = i\Theta m_{jk}(s,z)$ *then the limit is an exponentially distributed random variable with mean*

$$\frac{\sum\limits_{j^{(0)},k^{(0)} \in X_0} \pi_0(j^{(0)})p_0(j^{(0)},k^{(0)})m_{jk}(0,0)}{\overline{\pi_0} A^{(0)} A^{(1)} \ldots A^{(m)} \underline{1}}$$

Heterogeneous multiprocessor systems

Performance evaluation and quantitative analysis of multiprocessor systems is of immense importance due to the multiplicity of the component parts and the complexity of their functioning. Several works have been devoted to the analysis of such systems under different conditions on processor access rates, bus holding times, and arbitration protocols (c.f. [1, 21] Realistic consideration of certain stochastic systems, however, often requires the introduction of a random environment where system parameters are subject to randomly occuring fluctuations. This situation may be attribute to certain changes in the physical environment, or sudden personnel changes and work load alterations.

The queueing model

Consider a multiprocessor computer system in which N different processors with a common memory are connected by a single bus. A processor that generates a request to use the bus is said to be active, otherwise it is called inactive or idle. The bus arbitration protocol (selection rule) is assumed to be FCFS, that is, the arbiter selects the next processor to use the bus amongst the active ones in order of requests' arrivals. The time intervals from the completion of the previous bus usage to the generation of a new request as well as the holding times of the common bus are exponentially distributed random variables with parameter depending on the state of the corresponding random environment. Each processor is characterised by its own acces and service rate. The processors operate in a random environment governed by an ergodic Markov chain $(\xi_1(t), t \geq 0)$ with state space $(1, \ldots, r_1)$ and with transition rate matrix $a^{(1)}_{i_1 j_1}, i_1, j_1 = 1, \ldots, r_1, \quad a^{(1)}_{i_1 i_1} = \sum_{j \neq i_1} a^{(1)}_{i_1 j})$. Moreover, it is assumed that each processor can have at most one outstanding request at any time, i.e., each processor can generate a new request only after the bus usage of the previous request has been completed.

Whenever the environmental process is in state i_1 , let $\lambda_p(i_1,\varepsilon)$ be the access rate for processor $p, p=1,\ldots,N$, respectively. Similarly, the shared bus is supposed to operate in a random environment governed by an ergodic Markov chain $(\xi_2(t), t\geq 0)$ with state space $(1,\ldots,r_2)$ and with transition rate matrix $(a^{(2)}_{i_2j_2}, i_2,j_2=1,\ldots,r_2, \quad a^{(2)}_{i_2i_2}=\sum_{j\neq i_2} a^{(2)}_{i_2j})$. Whenever the environmental process is in state i_2, let $\mu_p(i_2)$ be the service rate for processor $p, p=1,\ldots,N$, respectively. To this end the probability that processor p generates a request in the time interval $(t,t+h)$ is $\lambda_p(i_1,\varepsilon)h+o(h)$, where $\varepsilon>0, \quad i_1=1,\ldots r_1$, and the probability that processor p completes the bus usage in time interval $(t,t+h)$ is $\mu_p(i_2)h+o(h), i_2=1,\ldots,r_2, p=1,\ldots,N$. All random variables and the random environment are assumed to be independent of each other. Let us consider the system under the heavy traffic assumption, i.e., $\lambda_p(i_1,\varepsilon)\to\infty$ as $\varepsilon\to 0$. For simplicity let $\lambda_p(i_1,\varepsilon=\lambda_p(i_1)/\varepsilon, p=1,\ldots,N, i_1=1,\ldots,r_1$.

Denote by $Y_\varepsilon(t)$ the number of inactive processors at time t, and let

$$\Omega_\varepsilon(m)=\inf\{t:\ t>0, Y_\varepsilon(t)=m+1/Y_\varepsilon(0)\leq m\},$$

i.e., the instant at which the number of inactive processors reaches the $(m+1)$-th level for the first time, provided that at the beginning their number is not greater than $m, m = 1, \ldots, N-1$. In particular, if $m = N-1$ then the bus becomes idle since there is no active processor and, hence $\Omega_\varepsilon(N-1)$ can be referred to as the busy period length of the bus. Denote by $\pi_o(i_1, i_2 : 0; k_1, \ldots, k_N)$ the steady-state probability that $\xi_1(t)$ is in state $i_1, \xi_2(t)$ is in state i_2, there is no idle processor and the order of requests' arrival to the bus is $(k_1, \ldots, k_N)$. Similarly, denote by $\pi_o(i_1, i_2 : 1; k_2, \ldots, k_N)$ the steady-state probability that the first random environment is in state i_1, the second one is in state i_2 , processor k_1 is inactive and the other processors sent their requests in order $(k_2, \ldots, k_N)$. Clearly $(k_s, \ldots, k_N) \in V_N^{N-s+1}, s = 1, 2$, where V_N^{N-s+1} denotes the set of all variations of order $N-s+1$ of integets $1, \ldots, N$. Now we have:

Theorem 2. *For the system in question under the above assumptions, independently of the initial state, the distribution of the normalized random variable $\varepsilon^m \Omega_\varepsilon(m)$ converges weakly to an exponentially distributed random* *variable with parameter*

$$\Lambda = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \sum_{(k_1,\ldots k_N)\in V_N^N} \pi_o(i_1, i_2 : 1; k_2, \ldots, k_N)$$

$$\times \frac{\mu_{k_2}(i_2)}{\lambda_{k_1}(i_1)} \frac{\mu_{k_3}(i_2)}{\lambda_{k_1}(i_1) + \lambda_{k_2}(i_1)} \times \ldots \times \frac{\mu_{k_{m+1}}(i_2)}{\lambda_{k_1}(i_1) + \ldots \lambda_{k_m}(i_1)} \frac{1}{D},$$

where

$$D = \sum_{\substack{i_1,j_1=1 \\ j_1 \neq i_1}}^{r_1} \sum_{\substack{i_2,j_2=1 \\ j_2 \neq i_2}}^{r_2} \sum_{(k_1,\ldots k_N)\in V_N^N} \pi_o(i_1, i_2 : 0; k_1, \ldots, k_N)$$

$$\times \frac{a_{i_1 j_1}^{(1)} + a_{i_2 j_2}^{(2)}}{(a_{i_1 i_1}^{(1)} + a_{i_2 i_2}^{(2)} + \mu_{k_1}(i_2))^2}$$

Proof. Let us introduce the following stochastic process

$$Z_\varepsilon(t) = (\xi_1(t), \xi_2(t) : Y_\varepsilon(t); \beta_1(t), \ldots, \beta_{N-Y_\varepsilon(t)}(t))$$

where $\beta_1(t), \ldots, \beta_{N-Y_\varepsilon(t)}(t))$ denotes the indices of the active processors in the order of their request arrival to the bus. It is easy to see that $(Z_\varepsilon(t), t \geq 0)$ is a multi-dimensional Markov chain with state space

$$E = ((i_1, i_2 : s; k_1, \ldots, k_{N-s}), \quad i_1 = 1, \ldots, r_1, \quad i_2 = 1, \ldots, r_2, \\ (k_1, \ldots, k_{N-s}) \in V_N^{N-s}, s = 0, \ldots, N)$$

where $k_o = \{0\}$ by definition. Furthermore, let

$$\langle \alpha_m \rangle = ((i_1, i_2 : s; k_1, \ldots, k_{N-s}), \quad i_1 = 1, \ldots, r_1, \quad i_2 = 1, \ldots, r_2, \\ (k_1, \ldots, k_{N-s}) \in V_N^{N-s}, s = 0, \ldots, m).$$

Hence our aim is to determine the distribution of the first exit time of $Z_\varepsilon(t)$ from $\langle \alpha_m \rangle$, provided that $Z_\varepsilon(t) \in \langle \alpha_m \rangle$. It can easily be verified that the transition probabilities for the embedded Markov chain are

$$p_\varepsilon[(i_1, i_2 : s; k_1, \ldots, k_{N-s}), (j_1, i_2 : s; k_1, \ldots, k_{N-s})]$$
$$= \frac{a^{(1)}_{i_1 j_1}}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p \neq k_1, \ldots, k_{N-s}} \lambda_p(i_1)/\varepsilon + \mu_{k_1}(i_2)}, \quad s = 0, \ldots, N-1,$$

$$p_\varepsilon[(i_1, i_2 : N; 0), (j_1, i_2 : N; 0)] = \frac{a^{(1)}_{i_1 j_1}}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p=1}^{N} \lambda_p(i_1)/\varepsilon}, \quad s = N,$$

$$p_\varepsilon[(i_1, i_2 : s; k_1, \ldots, k_{N-s}), (i_1, j_2 : s; k_1, \ldots, k_{N-s})]$$
$$= \frac{a^{(2)}_{i_2 j_2}}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p \neq k_1, \ldots, k_{N-s}} \lambda_p(i_1)/\varepsilon + \mu_{k_1}(i_2)}, \quad s = 0, \ldots, N-1,$$

$$p_\varepsilon[(i_1, i_2 : N; 0), (j_1, i_2 : N; 0)] = \frac{a^{(2)}_{i_2 j_2}}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p=1}^{N} \lambda_p(i_1)/\varepsilon}, \quad s = N,$$

$$p_\varepsilon[(i_1, i_2 : s; k_1, \ldots, k_{N-s}), (i_1, i_2 : s+1; k_2, \ldots, k_{N-s})]$$
$$= \frac{\mu_{k_1}(i_2)}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p \neq k_1, \ldots, k_{N-s}} \lambda_p(i_1)/\varepsilon + \mu_{k_1}(i_2)}, \quad s = 0, \ldots, N-1,$$

$$p_\varepsilon[(i_1, i_2 : s; k_1, \ldots, k_{N-s}), (i_1, i_2 : s-1; k_1, \ldots, k_{N-s+1})]$$
$$= \frac{\lambda_{k_{N-s+1}}(i_1)/\varepsilon}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p \neq k_1, \ldots, k_{N-s}} \lambda_p(i_1)/\varepsilon + \mu_{k_1}(i_2)}, \quad s = 1, \ldots, N-1,$$

$$p_\varepsilon[(i_1, i_2 : N; 0), (i_1, i_2 : N-1; k)] = \frac{\lambda_k(i_1)}{a^{(1)}_{i_1 i_1} + a^{(2)}_{i_2 i_2} + \sum_{p=1}^{N} \lambda_p(i_1)/\varepsilon}, \quad s = N$$

As $\varepsilon \longrightarrow 0$ this implies

$$p_\varepsilon[(i_1,i_2:0;k_1,\ldots,k_N),(j_1,i_2:0;k_1,\ldots,k_N)] = \frac{a^{(1)}_{i_1j_1}}{a^{(1)}_{i_1i_1}+a^{(2)}_{i_2i_2}\mu_{k_1}(i_2)}, \quad s=0$$

$$p_\varepsilon[(i_1,i_2:0;k_1,\ldots,k_N),(i_1,j_2:0;k_1,\ldots,k_N)] = \frac{a^{(2)}_{i_2j_2}}{a^{(1)}_{i_1i_1}+a^{(2)}_{i_2i_2}\mu_{k_1}(i_2)}, \quad s=0$$

$$p_\varepsilon[(i_1,i_2:s;k_1,\ldots,k_{N-s}),(j_1,i_2:s;k_1,\ldots,k_{N-s})] = o(1), \quad s=1,\ldots,N,$$

$$p_\varepsilon[(i_1,i_2:s;k_1,\ldots,k_{N-s}),(i_1,j_2:s;k_1,\ldots,k_{N-s})] = o(1), \quad s=1,\ldots,N,$$

$$p_\varepsilon[(i_1,i_2:0;k_1,\ldots,k_N),(i_1,i_2:1;k_2,\ldots,k_N)] = \frac{\mu_{k_1}(i_2)}{a^{(1)}_{i_1i_1}+a^{(2)}_{i_2i_2}+\mu_{k_1}(i_2)}, \quad s=0,$$

$$p_\varepsilon[(i_1,i_2:s;k_1,\ldots,k_{N-s}),(i_1,i_2:s+1;k_2,\ldots,k_{N-s})]$$
$$= \frac{\mu_{k_1}(i_2)\varepsilon}{\sum_{p\neq k_1,\ldots,k_{N-s}}\lambda_p(i_1)}(1+o(1)), \quad s=1,\ldots,N-1$$

This agrees with the conditions (1)-(4), but here the zero level is the set

$$((i_1, i_2 : 0; k_1, \ldots, k_N), (i_1, i_2 : 1; k_1, \ldots, k_{N-1} \quad i_1 = 1, \ldots, r_1, \quad i_2 = 1, \ldots, r_2,$$
$$(k_1, \ldots, k_{N-s} \in V_N^{N-s}, s = 0, 1),$$

while the q-th level is the set

$$((i_1, i_2 : q+1; k_1, \ldots, k_{N-q-1}), \quad i_1 = 1, \ldots, r_1, \quad i_2 = 1, \ldots, r_2,$$
$$(k_1, \ldots, k_{N-q-1} \in V_N^{N-q-1}).$$

Since the level 0 in the limit forms an essential class, the probabilities $\pi_o(i_1, i_2 : 0; k_1, \ldots, k_N), \pi_o(i_1, i_2 : 1; k_1, \ldots, k_{N-1}) i_1 = 1, \ldots, r_1, i_2 = 1, \ldots, r_2, (k_1, \ldots, k_{N-s}) \in V_N^{N-s}, s = 0, 1$, satisfy the following system of equations

$$\pi_o(j_1, j_2 : 0; k_1, \ldots, k_N) = \tag{63}$$
$$= \sum_{i_1 \neq j_1} \pi_o(i_1, j_2 : 0; k_1, \ldots, k_N) a^{(1)}_{i_1 j_1} / [a^{(1)}_{i_1 i_1} + a^{(2)}_{j_2 j_2} + \mu_{k_1}(j_2)] +$$
$$+ \sum_{i_2 \neq j_2} \pi_o(j_1, i_2 : 0; k_1, \ldots, k_N) a^{(2)}_{i_2 j_2} / [a^{(2)}_{j_1 j_1} + a^{(2)}_{i_2 i_2} + \mu_{k_1}(i_2)] +$$
$$+ \pi_o(j_1, j_2 : 1; k_1, \ldots, k_{N-1}),$$

$$\pi_o(j_1, j_2 : 1; k_1, \ldots, k_{N-1}) = \tag{64}$$
$$= \pi_o(j_1, j_2 : 0; k_N, k_1, \ldots, k_{N-1}) \mu_{k_N}(j_2) / [a^{(1)}_{j_1 j_1} + a^{(2)}_{j_2 j_2} + \mu_{k_N}(j_2)].$$

To apply the asymptotic expressions (62), it is necessary to solve system

(63),(64), subject to normalizing condition

$$\sum_{i_1=1}^{r_1}\sum_{i_2=1}^{r_2}\sum_{(k_1,\ldots,k_N)}\{\pi_o(i_1,i_2:0;k_1,\ldots,k_N)+\pi_o(i_1,i_2:1;k_1,\ldots,k_{N-1})\}=1.$$

Suppose this solution is known. Then by substituting it into (62) it follows that

$$g(\langle\alpha_m\rangle)=\varepsilon^m\sum_{i_1=1}^{r_1}\sum_{i_2=1}^{r_2}\sum_{(k_1,\ldots,k_N)\in V_N^N}\pi_o(i_1,i_2:1;k_2,\ldots,k_N) \tag{65}$$

$$\frac{\mu_{k_2}(i_2)}{\lambda_{k_1}(i_1)}\frac{\mu_{k_3}(i_2)}{\lambda_{k_1}(i_1)+\lambda_{k_2}(i_1)}\times\ldots\times\frac{\mu_{k_{m+1}}(i_2)}{\lambda_{k_1}(i_1)+\ldots\lambda_{k_m}(i_1)}(1+o(1)).$$

Taking into account the exponentiality of $\tau_\varepsilon(j_1,j_2:s;k_1,\ldots,k_{N-s})$ for fixed

Θ it is implied that

$$E\exp\{i\varepsilon^m\Theta\tau_\varepsilon(j_1,j_2:0;k_1,\ldots,k_B)\} = 1+\varepsilon^m\frac{i\Theta}{a^{(1)}_{j_1j_1}+a^{(2)}_{j_2j_2}+\mu_{k_1}(j_2)}(1+o(1)),$$

$$E\exp\{i\varepsilon^m\Theta\tau_\varepsilon(j_1,j_2:s;k_1,\ldots,k_{N-s})\} = 1+o(\varepsilon^m), \quad s>0.$$

Notice that $\beta_\varepsilon=\varepsilon^m$ and therefore from Corollary 1 our statement immediately follows.

However, if $\mu_p(i_2)=\mu(i_2), \quad p=1,\ldots,N, i_2=1,\ldots,r_2$, then by substituting (64) into (63) then we get

$$\pi_o(j_1, j_2 : 0; k_1, \ldots, k_N) = \tag{66}$$
$$= \sum_{i_1 \neq j_1} \pi_o(i_1, j_2 : 0; k_1, \ldots, k_N) a^{(1)}_{i_1 j_1} / [a^{(1)}_{i_1 i_1} + a^{(2)}_{j_2 j_2} + \mu(j_2)] +$$
$$+ \sum_{i_2 \neq j_2} \pi_o(j_1, i_2 : 0; k_1, \ldots, k_N) a^{(2)}_{i_2 j_2} / [a^{(2)}_{j_1 j_1} + a^{(2)}_{i_2 j_2} + \mu(i_2)]$$
$$+ \pi(j_1, j_2 : 0; k_N, k_1, \ldots, k_{N-1}) \mu(j_2) / [a^{(1)}_{j_1 j_1} + a^{(2)}_{j_2 j_2} + \mu(j_2)].$$

Since the steady-state distribution of the governing Markov chains satisfies

$$\pi^{(1)}_{j_1} a^{(1)}_{j_1 j_1} = \sum_{i_1 \neq j_1} \pi^{(1)}_{i_1} a^{(1)}_{i_1 j_1}, \quad \pi^{(2)}_{j_2} a^{(2)}_{j_2 j_2} = \sum_{i_2 \neq j_2} \pi^{(2)}_{i_2} a^{(2)}_{i_2 j_2}, \tag{67}$$

it can easily be verified, that the solution of (66) together with (67) is

$$\pi_o(i_1, i_2 : 0; k_1, \ldots, k_N) = B\pi_{i_1}^{(1)}\pi_{i_2}^{(2)}(a_{i_1 i_1}^{(1)} + a_{i_2 i_2}^{(2)} + \mu(i_2)),$$

$$\pi_o(i_1, i_2 : 1; k_1, \ldots, k_{N-1}) = B\pi_{i_i}^{(1)}\pi_{i_2}^{(2)} + \mu(i_2)),$$

where B is the normalizing constant, i.e.

$$1/B = N! \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \pi_{i_1}^{(1)}\pi_{i_2}^{(2)}((a_{i_1 i_1}^{(1)} + (a_{i_2 i_2}^{(2)} + 2\mu(i_2)).$$

Thus, from Theorem 2 follows that $\varepsilon^m \Omega_\varepsilon(m)$ converges weakly to an exponentially distributed random variable with parameter

$$\Lambda = \frac{1}{N!} \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \sum_{(k_1,\ldots,k_N)\in V_N^N} \pi_{i_1}^{(1)}\pi_{i_2}^{(2)}\mu(i_2)^{m+1} \frac{1}{\lambda_{k_1}(i_1)} \frac{1}{\lambda_{k_1}(i_1) + \lambda_{k_2}(i_1)}$$

$$\times \ldots \times \frac{1}{\lambda_{k_1}(i_1) + \ldots + \lambda_{k_m}(i_1)}.$$

Consequently, the distribution of the time while the number of idle processors reaches the $(m+1)$-th level for the first time is approximated by

$$P(\Omega_\varepsilon(m) > t) = P(\varepsilon^m \Omega_\varepsilon(m) > \varepsilon^m t) \approx \exp(-\varepsilon^m \wedge t).$$

In particular, when $m = N-1$, we get that the busy period length of the bus is asymptotically an exponentially distributed random variable with parameter

$$\varepsilon^{N-1}\wedge = \varepsilon^{N-1}\frac{1}{N!}\sum_{i_1=1}^{r_1}\sum_{i_2=1}^{r_2}\sum_{(k_1,\ldots,k_N)\in V_N^N} \pi_{i_1}^{(1)}\pi_{i_2}^{(2)}\mu(i_2)^N \times$$
$$\times \frac{1}{\lambda_{k_1}(i_1)}\frac{1}{\lambda_{k_1}(i_1)+\lambda_{k_2}(i_1)} \times \ldots \times \frac{1}{\lambda_{k_1}(i_1)+\ldots+\lambda_{k_N}(i_1)}. \quad (68)$$

In the case when there are no random environments, i.e., $\mu(i_2) = \mu$, and $\lambda_p(i_1) = \lambda_p, i_1 = 1,\ldots,r_1,\ i_2 = 1,\ldots,r_2, p = 1,\ldots,N$, from (68) it follows that

$$\varepsilon^{N-1}\wedge = \frac{\mu^N}{N!} \sum_{(k_1,\ldots,k_N)\in V_N^N} \frac{1}{\lambda_{k_1}/\varepsilon} \frac{1}{\lambda_{k_1}/\varepsilon + \lambda_{k_2}/\varepsilon} \times$$
$$\times \ldots \times \frac{1}{\lambda_{k_1}/\varepsilon + \ldots + \lambda_{k_{N-1}}/\varepsilon}. \tag{69}$$

Finally, for the special case of totally homogeneous processors (i.e., $\lambda_p = \lambda, \quad p = 1, \ldots, N$) expression (69) reduces to

$$\varepsilon^{N-1}\wedge = \frac{1}{(N-1)!} \frac{\mu^N}{(\lambda/\varepsilon)^{N-1}} \tag{70}$$

Performance measures

This section deals with the derivation of the main steady-state performance measures relating to the heterogeneous multiprocessor model treated in the previous section.

Utilizations

The utilization U of the bus is defined as the fraction of time during which it is busy. The idle period of the bus starts when each processor is idle at the end of a service completion, and terminates when a processor generates a request. It is clear that the mean idle period length is

$$\sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\sum_{p=1}^{N} \lambda_p(i_1)/\varepsilon}.$$

Hence for U the following expression is obtained

$$U = \frac{\frac{1}{\varepsilon^{N-1}\wedge}}{\frac{1}{\varepsilon^{N-1}\wedge} + \sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\sum_{p=1}^{N} \lambda_p(i_1)/\varepsilon}}. \tag{71}$$

The bus utilization U_p of processor p is defined as the fraction of time that processor p uses the bus. Since the processors have identically distributed holding times we get

$$U_p = U \sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \left(\lambda_p(i_1) / \sum_{k=1}^{N} \lambda_k(i_1) \right) \tag{72}$$

Throughput

The throughput γ_p of processor p is defined as the mean number of requests of processor p served per unit time. It is well-known that

$$U_p = \gamma_p b_p$$

where b_p is the mean bus usage(service) time of a request by processor p. In this case

$$U_p = \gamma_p \sum_{i_2=1}^{r_2} \pi_{i_2}^{(2)} \frac{1}{\mu(i_2)}$$

and thus

$$\gamma_p = U_p / \sum_{i_2=1}^{r_2} \pi_{i_2}^{(2)} \frac{1}{\mu(i_2)}.$$

Mean delay and waiting times

The mean delay T_p of processor p is the average time from the instant at which a request is generated at processor p to the instant at which the bus usage of that request has been completed. In other words, T_p is the mean duration of an active state at processor p. Since the state of processor p alternates between the active state of average duration T_p and the inactive state of mean duration

$$\sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\lambda(i_1)/\varepsilon}$$

the following relationship clearly holds

$$\gamma_p = \frac{1}{T_p + \sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\lambda(i_1)/\varepsilon}}.$$

Thus,

$$T_p = \frac{1}{\gamma_p} - \sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\lambda_p(i_1)/\varepsilon}.$$

Furthermore, for the mean waiting time W of processor p it follows that

$$W_p = T_p - \sum_{i_2=1}^{r_2} \pi_{i_2}^{(2)} \frac{1}{\mu(i_2)}.$$

Average number of requests served during a busy period

A pair of an idle period followed by an busy period is called a cycle, whose mean length is denoted by C. Clearly,

$$C = \frac{1}{\varepsilon^{N-1}\wedge} + \sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\sum_{p=1}^{N} \lambda_p(i_1)/\varepsilon}.$$

Denote by N_p the mean number of requests of processor p served during a cycle. The throughput γ_p of processor p is then given by $\gamma_p = N_p/C$, which yields that the total number of requests served during an busy period is

$$\sum_{p=1}^{N} N_p = \sum_{p=1}^{N} \gamma_p C.$$

Mean number of active processors

Let us denote by $Q^{(p)}$ the steady-state probability that processor p is idle. Clearly, we have

$$Q^{(p)} = \gamma_p \sum_{i_1=1}^{r_1} \pi_{i_1}^{(1)} \frac{1}{\lambda_p(i_1)/\varepsilon}.$$

Hence, the mean number of active processors is

$$\sum_{p=1}^{N} (1 - Q^{(p)}) = N - \sum_{p=1}^{N} Q^{(p)}.$$

Numerical results

This section presents a number of validation experiments (c.f., Table 7) examining the credibility of the proposed approximation against exact results for the performance measure of processor utilization at equilibrium. Note that an exact formula for the utilization is known only when the system is not effected by random environment and it is given (via Palm-formula) by

$$U_p^* = \frac{1}{N} \frac{\sum_{k=1}^{N} \binom{N}{k} k! \rho^k}{1 + \sum_{k=1}^{N} \binom{N}{k} k! \rho^k},$$

where $\rho = \frac{\lambda/\varepsilon}{\mu}$. In this case relations (70-72) reduce to the following approximation

$$U_p = \frac{1}{N} \frac{N!}{N! + (\frac{\mu}{\lambda/\varepsilon})^N}.$$

The following results are derived:

$N = 3$			$N = 4$		
ρ	U_p^*	U_p	ρ	U_p^*	U_p
1	0.3125	0.285714286	1	0.246153846	0.24
2	0.329113924	0.326530612	2	0.249605055	0.249350649
2^2	0.332657201	0.332467532	2^2	0.249968310	0.249959317
2^3	0.333237575	0.333224862	2^3	0.249997756	0.249997457
2^4	0.333320592	0.333319771	2^4	0.249999999	0.249999999
2^5	0.333333169	0.333331638	2^5	0.25	0.25
2^6	0.333333125	0.333333121			
2^7	0.333333307	0.333333307			
2^8	0.333333333	0.333333333			

$N=5$			$N=6$		
ρ	U_p^*	U_p	ρ	U_p^*	U_p
1	0.199386503	0.198347107	1	0.166581502	0.166435506
2	0.199968409	0.199947930	2	0.166664473	0.166666305
2^2	0.199998732	0.199998372	2^2	0.166666623	0.166666661
2^3	0.199999955	0.199999949	2^3	0.166666666	0.166666666
2^4	0.199999998	0.199999998			
2^5	0.2	0.2			

$N=7$			$N=8$		
ρ	U_p^*	U_p	ρ	U_p^*	U_p
1	0.142846715	0.142828804	1	0.124998860	0.1249969
2	0.142857009	0.142856921	2	0.124999993	0.124999988
2^2	0.142857142	0.142857141	2^2	0.125	0.125
2^3	0.142857143	0.142857143			

$N=9$			$N=10$		
ρ	U_p^*	U_p	ρ	U_p^*	U_p
1	0.111110998	0.111110805	1	0.099999999	0.099999999
2	0.111111111	0.111111111	2	0.1	0.1

Table 7: Exact and asymptotic results

It can be observed from Table 7 that the approximate values for $\{U_p\}$ are very much comparable in accuracy to those provided by the exact results for $\{U_p^*\}$. However, the computational complexity, due to the proposed approximation, has been considerably reduced. As λ/ε becomes greater than μ, the $\{U_p\}$ approximations, as expected, approach the exact values of $\{U_p^*\}$. Clearly, the greater the number of processors the less number of steps are needed to reach the exact results.

Other papers on systems with randomly changing parameters:
[4, 68, 69, 70, 71, 72, 73]

References

[1] Ajmone, M., Balbao, G., and Conte, G. *Performance Modeling of Multiprocessor Systems*. The MIT Press, Cambridge, 1986.

[2] Allen, A. O. *Probability, statistics, and queueing theory. With computer science applications. 2nd ed.* Academic Press, Inc., Boston, MA, 1990.

[3] Anisimov, V. Limit distributions of functionals of a semi-Markov process defined on a fixed state space until the first exit time,. *Soviet Math.Dokl. 11* (1970), 1002–1004.

[4] Anisimov, V., and Sztrik, J. Asymptotic analysis of some controlled finite-source queueing systems. *Acta Cybern. 9*, 1 (1989), 27–38.

[5] Anisimov, V., Zakusilo, O., and Donchenko, V. *Elements of queueing theory and asymptotic analysis of systems*. Visha Skola, KIev, 1987.

[6] Babitskij, A., Dudin, A., and Klimenok, V. On calculation of the characteristics of unreliable queueing systems with finite source. *Autom. Remote Control 57* (1996), 75–84.

[7] Baccelli, F., and Makowski, A. Stability and bounds for single server queues in random environment . *Communications in Statistics: Stochastic Models 2* (1986), 281–292.

[8] Begain, K., Bolch, G., and Herold, H. *Practical Perfromance Modeling, Application of the MOSEL Language*. Wiley & Sons, New York, 2001.

[9] Bolch, G., Greiner, S., de Meer, H., and Trivedi, K. *Queueing Networks and Markov Chains*. Wiley & Sons, New York, 1998.

[10] Boxma, O. A queueing model of finite and infinite source interaction. *Oper. Res. Lett. 5* (1986), 245–254.

[11] BUNDAY, B., AND R.E., S. The G/M/r machine interference model. *Eur. J. Oper. Res. 4*, 9 (1980), 399–402.

[12] CARMICHAEL, D. *Engineering Queues in Contruction and Mining*. Ellis Horwood, Chichester, 1987.

[13] CHANDRA, M. J. A study of multiple finite-source queueing models. *J. Oper. Res. Soc. 37* (1986), 275–283.

[14] CHEN, I.-R., AND WANG, D.-C. Repairman models for replicated data management: A case study. In *Fourth International Conference on Parallel and Distributed Information Systems (PDIS '96)* (Los Alamitos, Ca., USA, Dec. 1996), IEEE Computer Society Press, pp. 184–195.

[15] COOPER, R. B. *Introduction to queueing theory. 2nd ed.* Edward Arnold, London, 1981.

[16] DAIGLE, J. N. *Queueing theory for telecommunications.* Addison-Wesley, Reading, MA, 1992.

[17] Dshalalow, J. H. *Advances in queueing. Theory, methods, and open problems.* CRC Press., Boca Raton, 1995.

[18] Dshalalow, J. H. *Frontiers in Queueing : Models and Applications in Science and Engineering*. CRC Press., Boca Raton, 1997.

[19] Franken, P., Konig, D., and Arndt, U. Schmidt, V. *Queues and Point Processesy*. Academie Verlag, Berlin, 1981.

[20] Gaver, D., Jacobs, P., and Latouche, G. Finite birth-and-death models in randomly changing environments . *Advances in Applied Probability 16* (1984), 715–731.

[21] Gelenbe, E. *Multiprocessor Performance*. Wiley & Sons, Chichester, 1989.

[22] Gelenbe, E., and Mitrani, I. *Analysis and Synthesis of Computer Systems*. Academic Press, London, 1980.

[23] Gelenbe, E., and Pujolle, G. *Introduction to Queueing Networks*. Wiley & Sons, Chichester, 1987.

[24] German, R. *Performance analysis of communication systems. Modeling with non-Markovian stochastic Petri nets.* John Wiley & Sons, Chichester, 2000.

[25] Gertsbakh, I. Asymptotic methods in reliability theory; A review. *Advances in Applied Probability 16*, 1 (1984), 147–175.

[26] Gertsbakh, I. *Statistical reliability theory.* Marcel Dekker, Inc., New York, 1989.

[27] Gnedenko, B., Belyayev, Y., and Solovyev, A. *Mathematical methods of reliability theory*. Academic Press, New York, London, 1969.

[28] Gnedenko, B., and Kovalenko, I. *Introduction to queueing theory. Transl. from the Russian by Samuel Kotz. 2nd ed., rev. and suppl.* Birkhaeuser, Boston, MA, 1991.

[29] GROSS, D., AND HARRIS, C. M. *Fundamentals of queueing theory. 2nd ed.*. John Wiley & Sons., New York, 1985.

[30] GUPTA, S. Finite source Erlang based queueing systems: Complementarity, equivalence and their implications. *Comput. Math. Appl. 28*, 6 (1994), 57–74.

[31] GUPTA, U., AND RAO, T. V. A recursive method to compute the steady state probabilities of the machine interference model: (M/G/1)/K. *Comput. Oper. Res. 21*, 6 (1994), 597–605.

[32] HALL, R. W. *Queueing methods. For services and manufacturing.* Prentice Hall, Englewood Cliffs, NJ, 1991.

[33] HAVERKORT, B. *Performance of Computer Communication Systems: A Model-Based Approach*. Wiley & Sons, New York, 1998.

[34] Haverkort, B., Marie, R., Rubino, G., and Trivedi, K. *Performability Modelling: Techniques and Tools*. Wiley & Sons, New York, 2001.

[35] Ivcsenko, G., Kastanov, V., and Kovalanko, I. *Theory of Queueing Systems*. Nauka, Moscow, 1982.

[36] Jain, R. *The Art of Computer Systems Performance Analysis*. Wiley & Sons, New York, 1991.

[37] Jaiswal, N. *Priority Queues*. Academic Press, New York, 1969.

[38] Jeya Chandra, M., and Shantikumar, J. On the machine interference problem with several types of machines attended by a single repairman. *Int. J. Prod. Res. 7* (1983), 529–541.

[39] Keilson, J. *Markov Chain Models - Rarity and Exponentiality*. Springer-Verlag, New York, 1979.

[40] King, J. *Computer and Communication Systems Performance Modeling*. Prentice Hall, Englewood Cliffs, N.J., 1990.

[41] Kleinrock, L. *Queueing systems. Vol. I. Theory*. John Wiley & Sons, New York, 1975.

[42] Kleinrock, L. *Queueing systems. Vol. II: Computer applications.* John Wiley & Sons, New York, 1976.

[43] Kobayashi, H. *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology*. Addison-Wesley, Reading, MA, 1978.

[44] Korolyuk, V., and Korolyuk, V. *Stochastic Models of Systems*. Kluwer Academic Publishers, Dordrecht, London, 1999.

[45] Koulamas, C., and Iakovou, E. Machining economics with machine interference considerations. *Int. J. Prod. Res. 33*, 1 (1995), 71–78.

[46] KOVALENKO, I. Rare events in queueing systems, A survey, . *Queueing Systems 16* (1994), 1–49.

[47] KOVALENKO, I., PEGG, P., AND KUZNETZOV, N. *Mathematical Theory of Reliability of Time Dependent Systems with Practical Applications*. Wiley & Sons, New York, 1997.

[48] KULKARNI, V. *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, London, 1996.

[49] LINDEMANN, C. *Stochastic Modeling using DSPNexpress*. Oldenburg, Munich, 1994.

[50] MAHMOUD, M. S., AND BAHNASAWI, A. A. Sensitivity analysis of machine interference in manufacturing systems. *Int. J. Syst. Sci. 26*, 9 (1995), 1673–1686.

[51] MITRANI, I. *Modelling of Computer and Communication Systems*. No. 24 in Cambridge Computer Science Texts. Cambridge University Press, 1987.

[52] NELSON, R. *Probability, stochastic processes, and queueing theory. The mathematics of computer performance modeling.* Springer-Verlag, New York, 1995.

[53] PECK, L., AND HAZELWOOD, R. *Finite Queueing Tables*. Wiley & Sons, New York, 1958.

[54] SAHNER, R., TRIVEDI, K., AND PULIAFITO, A. *Performance and Reliability Analysis of Computer Systems – An Example-Based Approach Using the SHARPE Software Package*. Kluwer Academic Publisher, Boston, M.A., 1996.

[55] SAUER, C., AND CHANDY, K. *Computer Systems Performance Modelling*. Prentice Hall, Englewood Cliffs, N.J., 1981.

[56] Scherr, A. *An Analysis of Time-Shared Computer Systems*. MIT, Cambridge, MA, 1967.

[57] Schlott, R. Bandwidth allocation in an IBCN queueing system with quasirandom input. *ETT AEI 4, 6* (1993), 705–722.

[58] Stecke, K. E., and Aronson, J. E. Review of operator/machine interference models. *Int. J. Prod. Res. 23, 1* (1985), 129–151.

[59] Stewart, W. *MARCA: MARkov Chain Analyser, A Software Package for Makrov Modelling, Version 2.0*. North Carolina State University, Department of Computer Science, Raleigh, N.C., 1990.

[60] Stewart, W. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, 1995.

[61] Sztrik, J. On the machine interference. *Publ. Math. 30* (1983), 165–175.

[62] Sztrik, J. Probability model for non-homogeneous multiprogramming computer system. *Acta Cybern. 6* (1983), 93–101.

[63] Sztrik, J. A queueing model for multiprogrammed computer systems with different I/O times. *Acta Cybern. 7* (1985), 127–135.

[64] Sztrik, J. A finite-source queueing model for some manufacturing processes. *Probl. Control Inf. Theory 16*, 6 (1987), 449–457.

[65] Sztrik, J. Investigation of stationary characteristics of controlled systems $\vec{G}/M/r$ with finite source. *Serdica 14*, 2 (1988), 179–184.

[66] Sztrik, J. Some contributions to the machine interference problem with heterogeneous machines. *J. Inf. Process. Cybern. 24*, 3 (1988), 137–143.

[67] Sztrik, J. The $\vec{G}/M/1$ machine-interference model with state-dependent speeds. *J. Oper. Res. Soc. 39*, 2 (1988), 201–207.

[68] Sztrik, J. Modelling of heterogeneous multiprocessor systems with randomly changing parameters. *Acta Cybern. 10* (1991), 71–84.

[69] Sztrik, J. Modelling of a single bus multiprocessor system operating in Markovian environments. *Comput. Math. Appl. 23*, 11 (1992), 57–65.

[70] Sztrik, J. Asymptotic analysis of a heterogeneous finite-source communication system operating in random environments. *Publ. Math. 42*, 3-4 (1993), 225–238.

[71] Sztrik, J. Modelling of a multiprocessor system in a randomly changing environment. *Perform. Eval. 17*, 1 (1993), 1–11.

[72] Sztrik, J., and Lukashuk, L. Modelling of a communication system evolving in a random environment. *Acta Cybern. 10* (1991), 85–91.

[73] Sztrik, J., and Rigo, R. On a closed communication system with fast sources and operating in Markovian environments. *J. Inf. Process. Cybern. 29*, 4 (1993), 241–246.

[74] TAKACS, L. *Introduction to the Theory of Queues*. Oxford University Press, New York, 1962.

[75] TAKAGI, H. *Queueing analysis. A foundation of performance evaluation. Volume 1: Vacation and priority systems, part 1.* North-Holland, Amsterdam, 1991.

[76] TAKAGI, H. *Queueing analysis. A foundation of performance evaluation. Volume 2: Finite Systems*. North-Holland, Amsterdam, 1993.

[77] TIJMS, H. *Stochastic Modelling and Analysis: A Computational Approach*. Wiley & Sons, New York, 1986.

[78] TOSIRISUK, P., AND CHANDRA, J. Multiple finite source queueing model with dynamic priority scheduling. *Nav. Res. Logist. 37*, 3 (1990), 365–381.

[79] TRIVEDI, K. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. Prentice-Hall, Englewood Cliffs, N.J., 1982.

[80] USHAKOV, I. A., AND HARRISON, R. A. *Handbook of reliability engineering. Transl. from the Russian. Updated ed.* John Wiley & Sons, New York, NY, 1994.

[81] VAN HOORN, M. *Algorithms and approximations for queueing systems.* Centrum voor Wiskunde en Informatica, Amsterdam, 1984.

[82] WANG, K., AND SIVAZLIAN, B. Comparative analysis for the G/G/r machine repair problem. *Computers and Industrial Engineerings 18*, 4 (1990), 511–520.

[83] YOUNIS, M. A., AND HAMED, A. F. A stochastic approach to machine interference problems in robot-served manufacturing systems. *Int. J. Syst. Sci. 28*, 6 (1997), 563–570.