Transient Models for Queueing Networks

A THESIS
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
OF THE UNIVERSITY OF STELLENBOSCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Graham P. Phillips
August 1995

Supervised by: Prof A. E. Krzesinski

Declaration

I the undersigned hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree.

Signature: .... Date: ....

Abstract

Models for communication networks provide tools for network dimensioning and facilitate the design of control mechanisms, for example, access control, congestion control and routing mechanisms. The aim of the thesis is to investigate models for *transient* queueing systems. Exact models for transient systems are computationally expensive and therefore *approximate* models are compared to determine which models are best suited.

A literature survey is given and the following models are selected for further analysis: Filipiak's model for the $M/M/1/K$ queue, Moore's model for the $M^X/E_Y/1/K$ queue as well as Filipiak's and Tipper & Sundareshan's (TS) models for a network of $M/M/1/K$ queues. The AQ-model is derived in this thesis from Filipiak's model.

The models are implemented in the C programming language and compared over a wide range of conditions. The accuracy for an approximate model is defined in terms of the difference between the queue length yielded by the exact model and the queue length yielded by the approximate model. The following results are obtained:

- The models for the $M/M/1/K$ queue listed in order of decreasing accuracy are: the AQ-model, Moore's model and Filipiak's model while the models in order of increasing computation are: Moore's model, Filipiak's model and the AQ-model.
- The models for a *network* of $M/M/1/K$ queues listed in order of decreasing accuracy are: the TS-model, the AQ-model and Filipiak's model. Filipiak's model

and the AQ-model have similar computational requirements, while the TS-model has a significantly greater requirement. Therefore the AQ-model provides a good compromise between accuracy and computation for transient networks.

Opsomming

Modelle vir kommunikasienetwerke verskaf hulpbronne vir netwerkdimensionering en bevorder die ontwerp van beheermeganismes, byvoorbeeld, toegangsbeheer, opeenhopingsbeheer en roeteringsmeganismes. Hierdie tesis ondersoek modelle vir tydafhanklike-toustaanmodelle. Eksakte modelle vir dié stelsels vereis hoë-koste berekenings en word derhalwe met benaderde modelle vergelyk om sodoende die mees geskikte model vir toue en toustaannetwerke te bepaal.

'n Literatuuroorsig word gegee en die volgende modelle word breedvoerig ondersoek: Filipiak se model vir die $M/M/1/K$-tou, Moore se model vir die $M^X/E_Y/1/K$-toue asook Filipiak en Tipper & Sundareshan (TS) se modelle vir 'n netwerk van $M/M/1/K$-toue. Die AQ-model word in dié tesis uit Filipiak se model afgelei.

Die modelle word in die C programmeringstaal geïmplementeer en oor 'n wye reeks toestande in terme van hul noukeurigheid en berekeningskoste vergelyk. Die akkuraatheid van 'n model word gedefineer as die verskil tussen die toulengte van die eksakte model en die toulengte van die benaderde model. Resultate dui op die volgende:

- Die modelle vir die $M/M/1/K$-toue in afnemende volgorde van akkuraatheid is: die AQ-model, Moore se model en Filipiak se model terwyl die modelle in toenemende volgorde van berekeningskoste is: Moore se model, Filipiak se model en die AQ-model.

- Die modelle vir 'n *netwerk* van $M/M/1/K$-toue in afnemende volgorde van akkuraatheid is: die TS-model, die AQ-model en Filipiak se model. Filipiak se model en die AQ-model het soortgelyke berekeningskostes, terwyl die TS-model se berekeningskoste aansienlik hoër is. Gevolglik bied die AQ-model 'n goeie middeweg tussen die noukeurigheid en berekeningskoste vir tydafhanklikenetwerke.

Acknowledgements

I am indebted to the following people who have been particularly helpful while writing this thesis

- Prof. Tony Krzesinski for his guidance and support, without which I would have been unable to attempt this thesis.
- My parents and friends for their continued support and encouragement.

This thesis is funded in part by the Telkom Teletraffic Application Development Programme, the Foundation for Research and Development (FRD) and the Harry Crossley Scholarship.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The most widely used models for the performance evaluation of communication networks are based on queueing theory models which represent the communication system as a *queueing network*. The earliest problems studied were those of telephone traffic congestion pioneered by Danish mathematician A. K. Erlang in 1909 while additional influences include the works of two Russian mathematicians, A. N. Kolmogorov in the 1930's and A. A. Markov in the 1950's.

Tractable analysis of a queueing network is possible if the network has a product-form solution. Networks of infinite servers with non-homogeneous Poisson input are known to have a product-form transient solution [BT93, Fol82, HL81, Kin69]. There is a need to develop computationally efficient models for transient queueing networks composed of queues consisting of a finite number of servers and finite waiting lines.

1.1 Problem statement

The aim of the thesis is to investigate models for transient queueing systems. More specifically, we are interested in answering the following questions.

- *Which models found in the literature provide the best compromise between accuracy and computation for modelling communication networks?*
- *Can these models be improved in terms of accuracy or computation?*

1.2 The importance of transient network models

Due to the mathematical difficulty of solving transient queueing networks, researchers have previously emphasised steady-state models over transient ones. Morse [Mor58, p. 67] has even claimed that

> ...transient solutions are not usually of much practical importance.

However Morse qualifies this statement by stating that transient solutions are useful either when relaxation times are long or when one wishes to determine the time taken for a particular queue to grow to an unmanageable length (given that the arrival rate is increased to a value larger than the service rate).

Transients arise not only from changes in the statistics of the arrival processes but also from changes in the configuration of the network. Such behavioural changes are caused by modifications to routing tables, changes in network topology and failures of communication links. Consequently there has been an increased research effort devoted to transient queueing network models [TS90, TS89, AC89, Fil88, Har81, Ste79, Kot78, Cha77].

1.3 The importance of finite capacity models

An important characteristic of communication networks is the *blocking probability* which is the probability that a customer on arriving to a particular service centre cannot be accommodated and is rejected. The blocking probability is an important design criterion as it indicates the level of the network's availability to accept communication requests. Blocking is reduced at the expense of installing faster switching equipment or installing greater queue capacity.

Unlike finite capacity queueing models, infinite capacity models are unable to represent blocking. This deficiency limits the application of infinite capacity models to systems in which blocking plays a significant role in its performance. Consequently there is a need for finite capacity queueing models.

1.4 Scope of thesis

Models for both transient queues and transient queueing networks are investigated. Queues with finite capacity are emphasized over queues with infinite capacity and similarly, queues with single servers are emphasized over queues with multiple servers.

Several *approximate* models for the $M/M/1/K$ queue and a network of $M/M/1/K$ queues are compared in terms of their accuracy and computation. The accuracy of the approximate model is defined in terms of the difference between the queue length yielded by *Kolmogorov's system of differential equations* and the queue length yielded by the approximate model. The computational requirements for Kolmogorov's system severely restrict the size of the queueing network that is analysed.

1.5 Research contributions

We derive the AQ-model (approximate queue model) from Filipiak's *dynamic-flow model* developed in [Fil88] and show that the AQ-model is more accurate than Filipiak's model for a small packet-switched communication network under various transient conditions.

The computation for the AQ-model, like Filipiak's model, is dependent on the number of links in the network and is relatively small compared with other transient network models. Therefore we believe that the AQ-model is suitable for modelling larger communication networks.

1.6 Notation and terminology

The shorthand queueing notation due largely to Kendall is used [Ken53]. This notation is augmented with the subscript t that denotes a time dependence for the parameters of the particular process. For example, while $M/M/1$ describes a queue with exponential arrival and service processes with constant parameters, $M_t/M/1$ denotes a queue with a time dependent arrival process parameter.

The meanings of several words that are used throughout this thesis are given as follows:

exact/approximate model The system of equations describing a queueing system is referred to as a *model*. A model is *exact* if no approximations are made in the solution of its equations, otherwise the model is *approximate*.

transient/steady-state A *transient queueing system* is a system that has time dependent performance measures. Similarly a *transient solution* is a solution that is time dependent, while a solution that is independent of time is called a *steady-state solution*.

homogeneous/non-homogeneous A *homogeneous queueing system* is a system that has constant arrival and service process parameters. If either the arrival process or the service process has time dependent parameters then the system is *non-homogeneous*.

1.7 Overview

The remainder of this thesis is organized as follows.

Chapter 2 is a survey of transient queueing models found in the literature since 1953.

Chapter 3 compares two techniques for evaluating the queue length probabilities for the $M/M/1/K$ queue. These techniques provide an exact model for calculating the accuracy for the approximate models investigated in chapter 4.

Chapter 4 compares two approximate models for the $M/M/1/K$ queue: Filipiak's *dynamic-flow model* and a new model called the AQ-model. Both models consist of differential equations that are solved by numerical integration. The models provide approximations for mean queue length, mean rejected traffic and mean carried traffic for the $M/M/1/K$ queue.

Chapter 5 investigates Moore's approximate models for the $M^X/E_Y/1/K$ queue. The $M^X/E_Y/1/K$ queue is able to model a wider range of arrival and service processes than the $M/M/1/K$ queue and reduces to the $M/M/1/K$ queue as a special case. Moore's model is derived using greater mathematical rigour than Filipiak's model and, unlike Filipiak's model, approximates the queue length *distribution* from which several performance measures may be determined.

Chapter 6 compares three approximate models for the $M/M/1/K$ queue—Filipiak's model, the AQ-model and Moore's nonstationary arrival model—in terms of their accuracy and computation.

Chapter 7 compares three approximate queueing *network* models—Filipiak's model, Tipper & Sundareshan's model and the AQ-model—for a small packet-switched communication network. (Filipiak's model and the AQ-model consist of the same equations as the models in chapter 4, however additional equations are required to model the flow of traffic between the queues in the network.)

Finally, chapter 8 presents a summary of conclusions.

Chapter 2

Survey of transient queueing models

This chapter investigates several models of transient queues found in the literature. Two criteria are used to select the models. The first criterion is that the queue has a *finite* number of servers (infinite server queues are not considered). The second criterion is that the model yields an expression for the mean queue length, either explicitly or via the distribution of queue length.

The models are classified into two groups, those modelling queues with infinite capacity and those modelling queues with finite capacity. The chapter is concluded with a chronological summary.

2.1 Queues with infinite capacity

This section surveys models for queues that have infinite capacity. Three exact models are discussed (the $M/M/1$, the $M/E_k/1$ and the non-homogeneous $M_t/M_t/1$ queue) followed by several approximate models. The models are listed according to their queueing structure in table 1.

2.1.1 Exact models

The M/M/1 queue

A closed-form expression for the transient queue length distribution for the $M/M/1$ queue was first derived by Clarke in an unpublished paper in 1953 [Cla53], while the first *published* results were given by Ledermann & Reuter [LR54]. The claim that Clarke produced the first transient solution is substantiated by a quote from Champernowne [Cha56],

> The simple queue problem with constant parameters, dealt with first by W. Ledermann and G. E. H. Reuter (1954) and by A. B. Clarke (1953) was considered by these writers in relation to far more general problems and the solutions are accordingly less simple than is possible in the special case ...

Ledermann & Reuter used spectral analysis to derive their results for a general birth-death process with state-dependent parameters[1]. They include results for the special case of constant parameters (that is the $M/M/1$ queue) and state that in their opinion the standard method of generating functions cannot be used in this special case [LR54, section 14]. However Bailey shows that the generating function technique can be used to solve the $M/M/1$ queue [Bai54].

Although the queue length distribution for the $M/M/1$ queue is known, numerical procedures have not been given enough emphasis in the literature. Many textbooks present the queue length distribution in terms of modified Bessel functions without any discussion about the evaluation of this distribution (see equation (59)). However there are several formulations of the queue length distribution that require less computation to evaluate.

[1] A definition for a birth-death process is given in [Kle75].

- Takács derives an expression for queue length using results from the $M/M/1/K$ queue and taking the limit as the size of the waiting-line tends to infinity [Tak62].

- Jones et al. considers the problem of evaluating an infinite sum of Bessel functions in equation (59) [JCJ80]. The infinite sum is eliminated using Marcum's Q-function for which computational procedures are known [Mar60, DJ62, SBS66]. The introduction of this function leads to an expression for queue length in terms of Q-functions and a *finite* number of Bessel functions (equation (61)).

- Cantrell proposes a similar method where the infinite sum of Bessel functions is eliminated using the generalized Q-function [Can86]. The evaluation of the resulting expression is two to eight times faster than the evaluation of the expression given by Takács [AW89, p. 1103].

The M/E$_k$/1 queue

The Erlang random variable of type k is modelled by the sum of k independent and identically distributed exponential random variables. Therefore if the state of the queue is defined as the number of exponential phases waiting or being processed we have a Markov process (and therefore the queue can be described by Kolmogorov's system of equations).

The $M/E_k/1$ queue was solved for the idle probability $P_0(t)$ by Luchak [Luc56]. The phase probabilities $P_n(t)$ for $n > 0$ were found using $P_0(t)$ and Kolmogorov's system of equations (equations (65) and (67)). Note that the $M/E_k/1$ queue includes the $M/M/1$ queue as a special case when $k = 1$.

The non-homogeneous M$_t$/M$_t$/1 queue

The $M_t/M_t/1$ queue was solved by Clarke by converting Kolmogorov's equations into a *single* hyperbolic partial differential equation [Cla56]. This differential equation is solved using a standard method to obtain a solution in terms of a Volterra type integral

Queue	Model accuracy	Author(s)
$M/M/1$	Exact	Cantrell [Can86], Jones [JCJ80], Takács [Tak62], Luchak [Luc56], Champernowne [Cha56], Morse [Mor55], Bailey [Bai54], Ledermann et al. [LR54], Clarke [Cla56]
	Approx.	Kotiah [Kot78], Rider [Rid76], Rothkopf & Oren [RO79]
$M/E_k/1$	Exact	Luchak [Luc56]
	Approx.	
$BMAP/G/1$	Exact	
	Approx.	Lucantoni, Choudhury & Whitt [LCW93]
$M_t/M/1$	Exact	Lees & Boyd [LB66], Wragg [Wra63]
	Approx.	Filipiak [Fil88], Clark [Cla81], Rothkopf & Oren [RO79], Rider [Rid76]
$M_t/M_t/s$	Exact	Kolmogorov's system of forward DEs
	Approx.	Clark [Cla81], Rothkopf & Oren [RO79]

Table 1: Models for infinite capacity queues.

equation [Cla56, equations (17) and (18)]. The solution is numerically difficult to evaluate since the kernel of the solution tends to infinity at one end of the range of integration [Cla56, equation (18)]). Wragg proposed an alternative method [Wra63]. This method is considered the most favourable out of 8 numerical methods investigated by Lees & Boyd [LB66].

2.1.2 Approximate models

Kotiah's model for the M/M/1 queue

The *Laplace transform* for the queue length distribution is given in terms of a complex root inside the unit circle of an auxiliary quadratic equation (equation (63)). Inverting this transform yields a complicated expression for the queue length distribution (see section A.1). Kotiah approximated this transform by rational polynomials that are easy to invert [Kot78]. The model has the notable feature that upper and lower bounds for the approximation error are determined. Unfortunately the model is limited in its current form since results are given only for an initial queue length of zero.

Rider's model for the $\mathrm{M}_t/\mathrm{M}_t/1$ queue

The kernel of Rider's model for the $M_t/M_t/1$ queue is the equation

$$\frac{d}{dt}E[Q(t)] = \lambda(t) - \mu(t)\Big(1 - P_0(t)\Big)$$

where $E[Q(t)]$ is the mean queue length [Rid76]. (This equation is derived by multiplying Kolmogorov's system of forward DEs by n and summing over n.) Although this equation seems straightforward, an expression for the idle probability $P_0(t)$ is involved even for the simple case of constant $\lambda(t)$ and $\mu(t)$ (see equation (62)). Instead of evaluating an exact expression for $P_0(t)$, $P_0(t)$ is approximated in terms of known quantities: mean queue length, mean arrival rate and mean service rate. The model is solved by numerical integration.

Rothkopf & Oren's model for the $\mathrm{M}_t/\mathrm{M}_t/\mathrm{s}$ queue

Rothkopf & Oren's model for the $M_t/M_t/1$ queue [RO79] uses the same kernel as Rider's model. The idle probability $P_0(t)$ however is approximated by the negative binomial distribution. This distribution is chosen for two reasons. Firstly this distribution ensures that the model yields exact steady-state results [RO79, p. 525]. Secondly the parameters of the negative binomial distribution are uniquely specified by its mean and variance. The significance of this is that by augmenting the kernel with a single equation (equation (70)), the distribution's parameters (and therefore $P_0(t)$) can be determined.

Similarly for the $M/M/s$ queue Rothkopf & Oren use the negative binomial distribution to approximate the probabilities $P_0, P_1, \ldots, P_{s-1}$. A correcting term is required to ensure that the model yields the correct steady-state results.

Clark's model for the $\mathbf{M}_t/\mathbf{M}_t/\mathbf{s}$ queue

Clark's model for the $M_t/M_t/s$ queue uses a Polya-Eggenberger distribution [JK69] to reduce Kolmogorov's system of forward DEs to 5 differential equations [Cla81, equations (15) to (19)]. In addition to calculating the mean queue length, the queue length variance and mean waiting time are available from two supplementary equations. This model is more accurate than Rothkopf & Oren's model for $s > 1$.

Filipiak's model for the $\mathbf{M}_t/\mathbf{M}/\mathbf{1}$ queue

Filipiak's model for the $M_t/M/1$ queue has the form [Fil88]

$$\frac{d}{dt}E[Q(t)] \quad = \quad -\mu\frac{E[Q(t)]}{E[Q(t)]+1} + \lambda(t)$$

If $\lambda(t)$ is constant this equation is solved by separation of variables [Agn75]. However if $\lambda(t)$ is time dependent then numerical integration is used to solve the model. Filipiak's model approximates mean queue length and mean carried traffic. The model yields exact steady-state results if the queue is homogeneous (that is if $\lambda(t)$ is constant), however an estimate of the model's approximation error under transient conditions is not given. The model is computationally efficient and is suited for modelling transient queueing *networks*.

Lucantoni's model for the BMAP/G/1 queue

The $BMAP/G/1$ queue is modelled by Lucantoni et al. [LCW93]. This queue has a batch Markovian arrival process that is able to model a general class of Markovian arrival processes. (The arrival process includes the phase-type renewal process and the Markov modulated Poisson process as special cases and allows correlated inter-arrival times and batch sizes.) The service times are independent and identically distributed random variables.

Queue	Model accuracy	Author(s)
$M/M/1/K$	Exact	Abou-el-ata et al. [AAK93], Takács [Tak62], Morse [Mor58]
	Approx.	Stern [Ste79]
$M_t/M/1/K$	Exact	Kolmogorov's system of forward DEs
	Approx.	Filipiak [Fil88], Moore [Moo75]
$M_t^X/E_Y/1/K$	Exact	
	Approx.	Moore [Moo75]

Table 2: Models for finite capacity queues.

Numerical results for the queue length distribution are obtained by numerically inverting its Laplace transform.

2.2 Queues with finite capacity

This section surveys models for queues that have a finite capacity. Exact models are covered first followed by several approximate models. The models are listed in table 2 according to their queueing structure.

2.2.1 Exact models

The first solution for the queue length distribution for the $M/M/1/K$ queue was given by Morse [Mor58][2]. Takács derives an equivalent result using techniques from the theory of Linear Algebra [Tak62]. (See equations (77) and (78).)

The $M/M/1/K$ queue with balking is solved by Abou-el-ata et al. [AAK93]. The solution for the queue length distribution is intricate and includes integrals of Bessel functions. No consideration is given to the evaluation of the solution.

[2]This author has not found a solution for the $M/M/1/K$ queue prior to the Morse's paper in 1958.

2.2.2 Approximate models

Stern's model for the M/M/1/K queue

Stern's model for the $M/M/1/K$ queue takes advantage of the form of the queue length distribution for the exact model [Ste79]. The distribution is expressed as a sum of exponential terms (equations (77) and (78)) having the form

$$E[Q(t)] \quad = \quad L_0 + \sum_{i=1}^{K} L_i e^{-\alpha_i t}$$

where $\alpha_i > 0$. The computation required to evaluate $E[Q(t)]$ is reduced by removing insignificant terms from the summation. The terms considered insignificant are those terms $L_i e^{-\alpha_i t}$ with large values of α_i since these terms decay rapidly as t increases. The approximation error of the model is easily determined since the largest error occurs when t is zero.

Filipiak's model for the M$_t$/M/1/K queue

Filipiak's model for the $M_t/M/1/K$ queue is similar to Filipiak's model for the $M/M/1$ queue [Fil88]

$$\frac{d}{dt}E[Q(t)] \quad = \quad -\mu\, G(E[Q(t)]) + \lambda(t)$$

The functional form $G(\cdot)$ is determined such that the model yields a small error over a wide range of conditions. The model approximates mean queue length, mean rejected traffic and mean carried traffic. Filipiak's model is dealt with in greater detail in chapter 4.

Moore's model for the M$_t^X$/E$_Y$/1/K queue

Moore's model approximates the $M_t^X/E_Y/1/K$ queue [Moo75]. This queue has an arrival process that is modelled by a Poisson distribution with arrival groups of random size X, and service times that are modelled by a weighted-sum Erlang distribution. The model approximates the non-homogeneous arrival process by dividing the time axis into intervals such that the arrival process in each interval is homogeneous. The embedded Markov chain method is used to calculate the queue length distribution at the departure instants. This model is discussed in greater detail in chapter 5.

2.3 Chronological summary

The following table provides a chronological summary of the queueing models investigated in the previous sections. (The most recent paper is given first.)

The column labeled E shows whether the model is *exact* or approximate, the column labeled NR shows whether the paper includes *numerical results* for queue length while the column labeled QD shows whether the model calculates the *queue length distribution*. The last column shows some of the references made by the article.

Author(s)	Queue	E	NR	QD	Method	Importance of article	Refs made
Abou-el-ata et al. [AAK93]	$M/M/1/K$	✓	×	✓	Uses a generating function and Bessel functions.	Includes balking.	[Par87]
Lucantoni et al. [LCW93]	$BMAP/G/1$	×	✓	✓	Numerically inverts a Laplace transform.	Yields numerical results for a general Markov arrival process.	[AW92] [Ç67]
Filipiak [Fil88]	$M_t/M/1$ $M_t/M/1/K$ $M_t/M/c/c$	×	✓	×	Numerically integrates a single non-linear DE.	Yields a computationally efficient approximate solution.	[Rid76] [Moo75] [LB66]

continued on next page

Author(s)	Queue	E	NR	QD	Method	Importance of article	Refs made
Cantrell [Can86]	$M/M/1$	✓	✓	✓	The generalized Q-function replaces an infinite sum of Bessel functions.	Yields a computationally efficient solution.	[Ack82] [Ste79]
Jones et al. [JCJ80]	$M/M/1$	✓	✓	✓	Marcum's Q-function replaces an infinite sum of Bessel functions.	Avoids difficulties in the evaluation of an infinite sum of Bessel functions.	[Ste79]
Clark [Cla81]	$M_t/M_t/s$	×	✓	×	Approximates $P_s(t)$ and $P_{s+1}(t)$ using a Polya distribution. Integrates 5 DEs.	Improves accuracy of Rothkopf & Oren's model for $s > 1$.	[RO79]
Rothkopf & Oren [RO79]	$M_t/M_t/s$	×	✓	×	Approximates $P_0(t)$ using a negative binomial distribution. Integrates 2 DEs.	Improves accuracy of Rider's model.	[Rid76]
Stern [Ste79]	$M/M/1/K$	×	✓	×	Approximates the solution by the slowest transient component.	Yields a computationally efficient approximate solution.	[CS61]
Kotiah [Kot78]	$M/M/1$	×	✓	✓	Approximates a complex root by a rational polynomial.	Gives an error-bounds analysis for $P_n(t)$ and $E[Q(t)]$.	[Gav66]
Rider [Rid76]	$M_t/M/1$	×	✓	×	Approximates $P_0(t)$. Numerically integrates a single non-linear DE.	Yields a computationally efficient approximate solution.	[Cla56] [Koo72]
Moore [Moo75]	$M_t^X/E_Y/1/K$	×	✓	✓	Uses the embedded Markov chain method.	Yields numerical results for a non-homogeneous bulk arrival system.	[Koo72] [GW58]

continued on next page

Author(s)	Queue	E	NR	QD	Method	Importance of article	Refs made
Lees & Boyd [LB66]	$M_t/M_t/1$	✓	✓	×	Uses Wragg's method of solution.	Compares 8 numerical methods. Yields numerical results for Wragg's method.	[Wra63]
Wragg [Wra63]	$M_t/M/1$	✓	×	✓	Gives solution in terms of a Volterra integral equation.	Avoids numerical difficulties found in [Cla56].	[Cla56]
Takács [Tak62]	$M/M/1$ $M/M/1/K$	✓	×	✓	Solves Kolmogorov's system of DEs using results from Linear Algebra.	Derives the $M/M/1$ solution from the $M/M/1/K$ solution.	[Mor58]
Morse [Mor58]	$M/M/1/K$	✓	×	✓	Solves Kolmogorov's system of DEs directly.	Yields queue length probabilities for the $M/M/1/K$ queue.	
Luchak [Luc56]	$M/E_k/1$	✓	✓	✓	System state represents the number of service phases to be completed.	Generalizes the $M/M/1$ queue to include the Erlang service distribution.	[Bai54] [LR54] [Mor55]
Clarke [Cla56]	$M_t/M_t/1$	✓	×	✓	Uses a generating function to reduce system of DEs to a partial DE.	Yields results for the non-homogeneous $M/M/1$ queue.	[Bai54] [LR54]
Champernowne [Cha56]	$M/M/1$	✓	×	✓	Uses advanced combinatorial methods.	Gives an alternate method of solution for the $M/M/1$ queue.	[LR54] [Cla53]
Morse [Mor55]	$M/M/1$	✓	×	✓	Solves Kolmogorov's system of DEs directly.	Yields a solution in terms of integrals of trigonometrical terms.	

continued on next page

Author(s)	Queue	E	NR	QD	Method	Importance of article	Refs made
Bailey [Bai54]	$M/M/1$	✓	×	✓	Uses the Laplace transform-generating function technique.	Yields the results of [LR54] using a generating function.	[LR54] [Ken53]
Ledermann & Reuter [LR54]	State dependent birth-death process.	✓	×	✓	Uses spectral analysis.	Yields results for constant parameters λ and μ as a special case.	[Ken51]
Clarke [Cla53]	$M/M/1$	✓	×	✓		First expression for queue length probabilities for the $M/M/1$ queue.	

Table 3: A chronological summary of models.

Chapter 3

Exact models of the M/M/1/K queue

An exact model for the $M/M/1/K$ queue is required for two reasons, first to calculate the accuracy for approximate models for the $M/M/1/K$ queue (investigated in chapter 4) and second to implement Tipper & Sundareshan's approximate network model (investigated in chapter 7). The purpose of this chapter is to compare two exact evaluation methods in terms of their ease of implementation and their computational requirements.

The chapter is divided into three sections. The first section investigates the known *closed-form solution* for queue length probabilities while the second section investigates the numerical integration of *Kolmogorov's system of differential equations.* The last section compares the two methods and concludes that integrating Kolmogorov's equations is the preferred method because it is easy to implement and furthermore it can be used to solve a range of queueing systems.

3.1 Evaluation of the closed-form solution

The first exact model we investigate is the closed-form solution for the $M/M/1/K$ queue found by Morse [Mor58]. Takács [Tak62] provides an alternative solution given by

$$P_{i,n}(t) = P_n + p_{i,n}(t) \tag{1}$$

where

$$P_n = \frac{(1-\rho)}{1-\rho^{K+1}}\,\rho^n \tag{2}$$

and

$$\begin{aligned} p_{i,n}(t) = {} & \frac{2}{K+1}\rho^{(n-i)/2} \sum_{m=1}^{K} \frac{e^{-\left(\lambda+\mu-2\sqrt{\lambda\mu}\,\cos(m\pi/(K+1))\right)t}}{1-2\sqrt{\rho}\,\cos(m\pi/(K+1))+\rho} \\ & \cdot \left(\sin\frac{im\pi}{K+1} - \sqrt{\rho}\,\sin\frac{(i+1)m\pi}{K+1}\right) \\ & \cdot \left(\sin\frac{nm\pi}{K+1} - \sqrt{\rho}\,\sin\frac{(n+1)m\pi}{K+1}\right) \end{aligned} \tag{3}$$

where $\rho = \lambda/\mu$.

This solution gives the probability $P_{i,n}(t)$ that the queue length is n at time t given that the queue length is i at time $t = 0$. Note that $P_{i,n}(t)$ is the sum of a steady-state component P_n and a transient component $p_{i,n}(t)$.

It would seem at first glance that if $n < i$ then the factor $\rho^{(n-i)/2}$ causes the transient component $p_{i,n}(t)$ to become unbounded as ρ approaches zero. However this is not the case. The transient component $p_{i,n}(t)$ is bounded from equation (1) since both $P_{i,n}(t)$ and P_n are bounded (we assume that the probabilities $P_{i,n}(t)$ and P_n lie between 0 and 1). Therefore, given that $n < i$, as ρ decreases the summand in equation (3) decreases in such a way that $p_{i,n}(t)$ remains bounded. Although we have shown that $p_{i,n}(t)$ is bounded, numerical errors from the evaluation of the factor $\rho^{(n-i)/2}$ are expected when ρ is small. However, as we shall see, numerical errors are obtained for larger values of ρ than expected.

The evaluation of $P_{i,n}(t)$ yields numerical inaccuracies under various conditions and these inaccuracies are discussed in sections 3.1.1, 3.1.2 and 3.1.3.

3.1.1 Evaluating the transient component

One cause of numerical inaccuracy arises from the evaluation of the transient component $p_{i,n}(t)$ under the condition $\rho \sim 0$ and $n < i$ (see equation (3)). More specifically, the transient component $p_{i,n}(t)$ is evaluated using double precision floating point arithmetic (15 significant digits) and not a single significant digit is obtained for the values: $\lambda = 0.1, \mu = 1, t = 0.0, K = 8, i = 8$ and $n = 0$. The cause of this inaccuracy is not the factor $\rho^{(n-i)/2}$ (since $\rho^{(n-i)/2} = 10000$ and overflow does not occur) but from the summand. The reason for the inaccurate evaluation of the summand is that precision is lost when summing floating-point values that vary greatly in magnitude.

An attempt to overcome the numerical errors

We attempt to overcome the numerical errors found when evaluating $p_{i,n}(t)$ under the condition $\rho \sim 0$ and $n < i$. Greater accuracy for $p_{i,n}(t)$ is required when t is small than when t is large, because $p_{i,n}(t)$ decreases exponentially as t increases and therefore the error introduced by $p_{i,n}(t)$ also decreases as t increases. We therefore attempt to find an alternative expression for $p_{i,n}(t)$ under the condition $\rho \sim 0, i < n$ and $t \sim 0$. This attempt was not successful and therefore only partial results are given.

The transient term is rewritten as

$$p_{i,n}(t) = \frac{2}{K+1} \frac{N_{i,n}\left(\sqrt{\rho}, \mu t\right)}{D_{i,n}\left(\sqrt{\rho}\right)} \tag{4}$$

where

$$\begin{aligned} N_{i,n}(x,\nu) &= \sum_{m=1}^{K} e^{-\nu\left(1-2x\cos m\theta + x^2\right)} M_{i,n}(x,m) \\ M_{i,n}(x,m) &= \Big(\sin im\theta - x\sin(i+1)m\theta\Big) \cdot \Big(\sin nm\theta - x\sin(n+1)m\theta\Big) \end{aligned}$$

$$
\begin{aligned}
& \cdot \prod_{\ell=1,\ell\neq m}^{K} \left(1 - 2x\cos\ell\theta + x^2\right) \\
D_{i,n}(x) &= x^{i-n} \prod_{m=1}^{K} \left(1 - 2x\cos m\theta + x^2\right) \\
\theta &= \pi/(K+1) \\
\rho &= \lambda/\mu
\end{aligned}
$$

Now

$$
\begin{aligned}
N_{i,n}(x,\nu) &= \sum_{m=1}^{K} M_{i,n}(x,m) + \sum_{m=1}^{K} -\nu\left(1 - 2x\cos m\theta + x^2\right) M_{i,n}(x,m) \\
&\quad + \ldots + \sum_{m=1}^{K} \frac{(-\nu)^z}{z!}\left(1 - 2x\cos m\theta + x^2\right)^z M_{i,n}(x,m) + \ldots \qquad (5)
\end{aligned}
$$

We wish to expand $N_{i,n}(x,\nu)$ in terms of its power series in x and obtain the coefficients of this series. The coefficients are dependent on ν and their second order approximation in ν requires the expansion of the first two summands in equation (5). Therefore we wish to expand the first two summands as two power series in x and collect the coefficients of like powers of x.

As has been mentioned above, complete results were not obtained and only the first two coefficients of the first summand are given by the following result. (The results are proved in appendix B. Note that these coefficients are independent of K.)

Result 3.1.1 *If $M_{i,n}(x,m)$ is given by equation (4) and the sum $\sum_{m=1}^{K} M_{i,n}(x,m)$ is expanded in terms of its power series in x, that is*

$$
\sum_{m=1}^{K} M_{i,n}(x,m) = c_0 + c_1 x + c_2 x + \ldots + c_{2K} x^{2K} \qquad (6)
$$

then

$$
c_0 = \begin{cases} 1, & \text{if } i = n \text{ and } i \neq 0, \\ 0, & \text{otherwise.} \end{cases}
$$

$$
c_1 = \begin{cases} -1, & \text{if } i + n = 1, \\ 0, & \text{if } i + n \neq 1. \end{cases}
$$

An alternative evaluation method

We propose an alternative evaluation method to avoid the numerical errors found when evaluating the transient component $p_{i,n}(t)$. The evaluation method consists of using *computer algebra* to mechanically expand $N_{i,n}(x,\nu)$ (equation (5)) in terms of a power series. An algorithm for evaluating $p_{i,n}(t)$ using this method is implemented in the Maple programming language [Hec93]. (The program appears in section B.3.) We found that this method is possible provided the variables i, n, ν, K are known prior to the expansion process.

Although this method can be used to evaluate $p_{i,n}(t)$ to any given accuracy, the computation required is significant. For example, the Maple program in section B.3 running on an IBM PC under a WINOS2 environment takes 942 seconds to evaluate a single value for $p_{i,n}(t)$. (The PC consists of a 33MHz 386DX CPU, a 8087 coprocessor and 16Mbytes of main memory.)

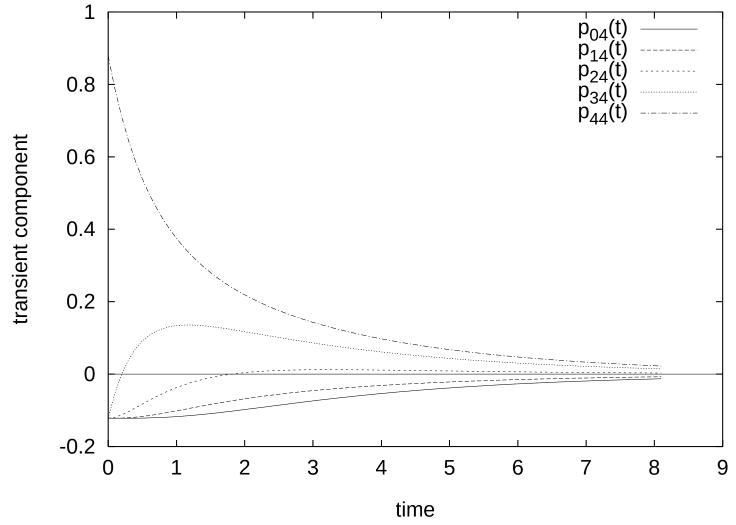
3.1.2 Evaluating the steady-state component

A second cause for numerical inaccuracy arises from evaluating the steady-state component P_n under the condition $\rho \sim 1$ (see equation (2)). The reason for this inaccuracy is that when $\rho = 1$, equation (2) yields an indeterminate form. However the indeterminate form is avoided by evaluating P_n using the expression

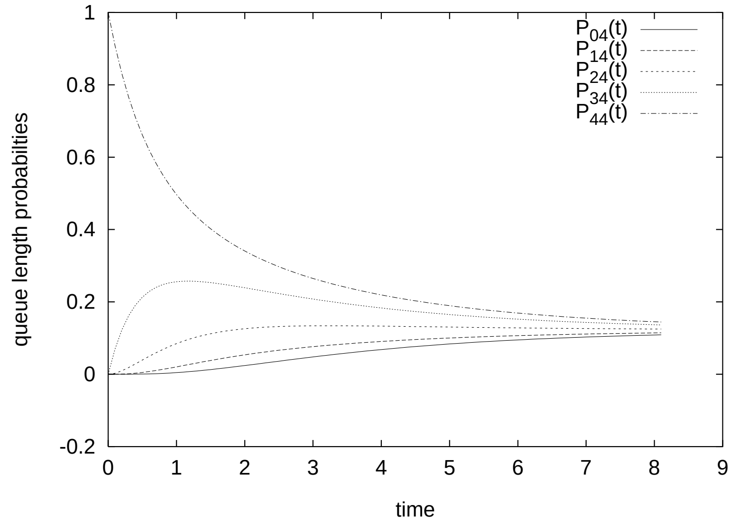
$$P_n \quad = \quad \left(\sum_{r=0}^{K} \binom{K+1}{r+1} (\rho - 1)^r \right)^{-1} \rho^n$$

3.1.3 Summing the transient and steady-state components

Another cause of numerical inaccuracy arises from summing the transient and steady-state components under the condition $t \sim 0$ and $i \neq n$ (see equation (1)) . The reason for this inaccuracy is that under this condition $p_{i,n}(t) \approx -P_n$ and therefore cancellation errors occur when summing these values. For example, section B.3



(a) transient components



(b) probabilities (transient plus steady-state components)

Figure 1: Evolution of blocking probabilities for the $M/M/1/4$ queue.

shows that 6 significant digits are lost when evaluating $P_{i,n}(t) = P_n + p_{i,n}(t)$ for $\lambda = 0.2, \mu = 1, t = 0.5, K = 8, i = 8$ and $n = 0$, since $P_n = 0.8000004096$ and $p_{i,n}(t) = -0.8000001470$.

Graphs of the queue length probability for the $M/M/1/4$ queue are shown in figure 1. Figure 1(a) shows the transient component of queue length probability $p_{i,n}(t)$ for different initial queue lengths i. (Note that the transient component may be negative and, unlike the steady-state component, does not represent a probability.) Figure 1(b) shows the sum of the transient and steady-state components.

3.2 Evaluation by numerical integration of Kolmogorov's equations

The second exact model we investigate is Kolmogorov's system of differential equations for the $M/M/1/K$ queue (equation (76)). These equations form a system of linear first order differential equations. It is easily shown that

$$\sum_{n=0}^{K} \frac{d}{dt} P_{i,n}(t) = 0$$

and this equation ensures that sum of probabilities is 1 for all t, given that the sum of probabilities is 1 for $t = 0$.

The model is solved by numerically integrating the system of equations. The numerical method is easily implemented due to the availability of numerical integration routines (for example, the Runga-Kutta integration routine `rkqc` found in [PFTV88]). This method has the advantage that it can be used to solve many queueing systems (in general, all queueing systems that are Markov processes). In particular this method is used to solve the non-homogeneous $M_t/M/1/K$ queue (in chapter 4) and also used to solve an exact model for a queueing network (in chapter 7).

3.3 Comparison of evaluation methods

The advantages and disadvantages of both methods are summarized as follows.

- Comparing the methods in terms of computation is difficult because the models are solved differently. For example, the closed-form solution yields a single value for queue length probability $P_{i,n}(t)$ while the numerical integration method may yield many intermediate values for queue length probability as the integration routine solves the system of equations from 0 to t.
- Comparing the methods in terms of ease of implementation, the numerical integration method is easier to implement due to the availability of integration routines. On the other hand special techniques are required to avoid the numerical difficulties when evaluating the closed-form solution under certain conditions.
- Comparing the methods in terms of their ability to solve a range of queueing systems, the integration method can be applied to all queueing systems that are Markov processes. On the other hand the closed-form solution for the $M/M/1/K$ queue is limited to solving this queue.

Therefore we conclude that the numerical integration method is the preferred method for providing an exact solution for the $M/M/1/K$ queue because it is easy to implement and furthermore it can be used to solve a range of queueing systems.

Chapter 4

Approximate models of the M/M/1/K queue

Two approximate models for the $M_t/M/1/K$ queue—Filipiak's *dynamic-flow model* and a new model, called the AQ-model—are investigated. Both Filipiak's model and the AQ-model consist of differential equations that contain certain *characteristic functions*. The functional form of these functions are determined such that the models yield approximate solutions under a wide range of transient conditions. Once the characteristic functions are determined, the models are solved numerically to yield the mean queue length for the $M_t/M/1/K$ queue.

The chapter is organized as follows. The derivation for Filipiak's model is summarized and the functional forms for the characteristic functions are determined. The AQ-model model is developed and the characteristic functions for this model are also determined. The evaluation of the characteristic functions is discussed. Finally, several evaluation methods for approximating the characteristic functions are compared in terms of their computation.

Filipiak's model and the AQ-model are compared together with another approximate model in terms of their accuracy and computation in chapter 6. Both models are developed further in chapter 7 to yield computationally efficient models for queueing networks.

4.1 Filipiak's approximate dynamic-flow model

The dynamic-flow model for the $M_t/M/1/K$ queue is developed by Filipiak in [Fil88]. The model is called a fluid flow approximation because the kernel of the model consists of an equation describing the rate of flow of customers into and out of the queue. Filipiak's dynamic-flow model is similar to the models of Rider [Rid76] and Agnew [Agn75] for the $M/M/1$ queue and Le Gall's model for the $M/M/1/K$ queue [LGB83].

4.1.1 Derivation

The mean queue length $E[Q(t)]$ for the $M_t/M/1/K$ queue is given by the solution for the equation

$$\frac{d}{dt}E[Q(t)] \quad = \quad -\mu\left(1 - P_0(t)\right) + \lambda(t)\left(1 - P_K(t)\right) \tag{7}$$

where

$$E[Q(t)] \quad = \quad \sum_{n=1}^{K} n\, P_n(t)$$

Equation (7) is obtained by multiplying Kolmogorov's system of differential equations for the $M_t/M/1/K$ queue (equation (79)) by n and summing over all values of n.

Substituting $x = E[Q(t)]$ into equation (7) and rewriting yields

$$\frac{d}{dt}x(t) \quad = \quad -\mu\left(1 - P_0(t) + \rho(t)P_K(t)\right) + \lambda(t)$$

where $\rho(t) = \lambda(t)/\mu$. Filipiak proposes that the term $1 - P_0(t) + \rho(t)P_K(t)$ is approximated by $G(x)$ which leads to the equation

$$\frac{d}{dt}x \quad = \quad -\mu\, G(x) + \lambda(t) \tag{8}$$

Filipiak further proposes that the rejected traffic defined by $\lambda^r(t) = \lambda(t)P_K(t)$ and carried traffic defined by $\lambda^c(t) = \mu(1 - P_0(t))$ are approximated by the equations

$$\lambda^r(t) \approx \mu\, G(x(t))\, H(x(t)) \tag{9}$$

$$\lambda^c(t) \approx \mu\, G(x(t)) \Big(1 - H(x(t))\Big) \tag{10}$$

4.1.2 Finding functional forms for the characteristic functions

We call $G(x)$ and $H(x)$ *characteristic functions* and these functions are determined such that the model yields approximate solutions under a wide range of conditions. Agnew proposes that the functions should be determined statistically if experimental data from existing systems can be obtained [Agn75]. However if such data are not available then the characteristic functions must be determined by other methods. Agnew proposes an alternative method which results in a model that is exact under steady-state conditions. It is this method that we follow.

Equation (8) reduces to

$$G(x) = \lambda/\mu = \rho \tag{11}$$

under steady-state conditions where λ is defined by

$$\lambda = \lim_{T\to\infty} \frac{1}{T} \int_0^T \lambda(t)dt$$

The characteristic function $G(x)$ is defined by equation (11). However before $G(x)$ is evaluated, equation (11) requires that ρ is defined in terms of x, that is

$$G(x) = \rho(x)$$

A rational function for ρ in terms of x cannot be found, however Filipiak observes that a rational function for its inverse, namely

$$x = L(\rho) \tag{12}$$

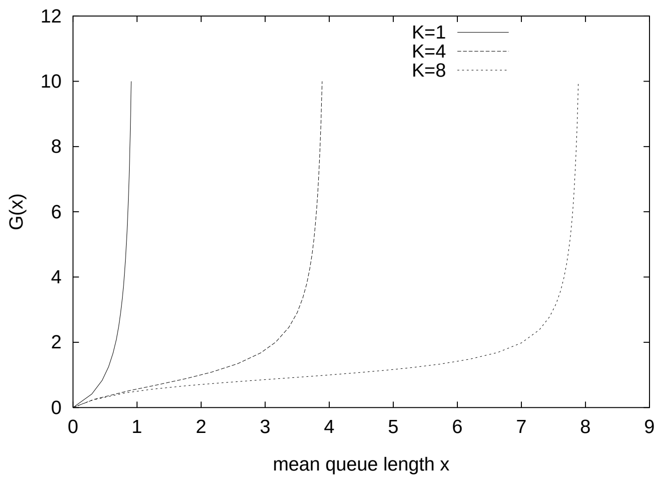


Figure 2: Characteristic function $G(x) = L^{-1}(x)$ for the $M_t/M/1/K$ queue.

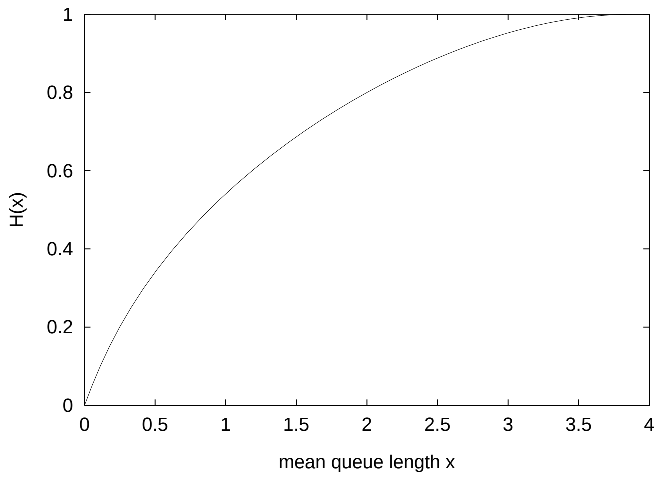


Figure 3: Characteristic function $H(x)$ for the $M_t/M/1/4$ queue.

is given by the steady-state queue length for the $M/M/1/K$ queue,

$$L(\rho) = \frac{\rho\left(1-(K+1)\rho^K + K\rho^{K+1}\right)}{(1-\rho)(1-\rho^{K+1})} \tag{13}$$

Therefore the characteristic function $G(x)$ is defined by

$$G(x) = L^{-1}(x) \tag{14}$$

and this functional form for $G(x)$ ensures that equation (8) yields exact results for queue length under steady-state conditions. A plot for $G(x)$ for various values of K is shown in figure 2.

The characteristic function $H(x)$ is determined such that the model yields accurate values for the rejected traffic under steady-state conditions. The rejected traffic under steady-state conditions is

$$\lambda^r = \lambda\, P_K \tag{15}$$

while the rejected traffic given by the dynamic-flow model (equation (9)) is

$$\lambda^r = \mu G(x) H(x) = \lambda H(x) \tag{16}$$

(The last equality is true since $G(x) = \rho$ and $\rho = \lambda/\mu$.) Given steady-state conditions, P_K is expressed as a function of ρ using the standard equation

$$P_K(\rho) = \frac{1-\rho}{1-\rho^{K+1}}\,\rho^K$$

Comparing equations (15) and (16) and using equation (12) yields

$$H(x) = P_K(\rho) = P_K\left(L^{-1}(x)\right) \tag{17}$$

and this functional form for $H(x)$ ensures that the model yields exact results for the rejected traffic under steady-state conditions. A plot for $H(x)$ for the $M_t/M/1/4$ queue is shown in figure 3.

4.2 The approximate AQ-model

We develop a new model called the AQ-model (approximate queue model). Like Filipiak's model, the AQ-model is derived from equation (7) for the mean queue length for the $M_t/M/1/K$ queue.

We approximate the terms $1 - P_0(t)$ and $P_K(t)$ by the characteristic functions $\widehat{G}(\cdot)$ and $\widehat{H}(\cdot)$ such that

$$\widehat{G}(x) \approx 1 - P_0(t) \tag{18}$$

$$\widehat{H}(x) \approx P_K(t) \tag{19}$$

under a wide range of conditions. Substituting equations (18) and (19) into equation (7) yields

$$\frac{d}{dt}x = -\mu\,\widehat{G}(x) + \lambda(t)\left(1 - \widehat{H}(x)\right) \tag{20}$$

where x is the mean queue length. The rejected traffic $\lambda^r(t) = \lambda(t)P_K(t)$ and the carried traffic $\lambda^c(t) = \mu(1 - P_0(t))$ are given by

$$\begin{aligned} \lambda^r(t) &= \lambda(t)\,\widehat{H}(x(t)) \\ \lambda^c(t) &= \mu\,\widehat{G}(x(t)) \end{aligned}$$

The characteristic functions $\widehat{G}(\cdot)$ and $\widehat{H}(\cdot)$ are determined in the same way as the characteristic functions for Filipiak's model, that is, the functions are determined such that the model yields accurate results under steady-state conditions. Equations (18) and (19) reduce to

$$\begin{aligned} \widehat{G}(x) &= 1 - P_0 \\ \widehat{H}(x) &= P_K \end{aligned}$$

under steady-state conditions. Using the result that P_0 and P_K are functions of ρ given by the standard equations

$$\begin{aligned} P_K &= \rho^K P_0 \\ P_0 &= \frac{1-\rho}{1-\rho^{K+1}} \end{aligned}$$

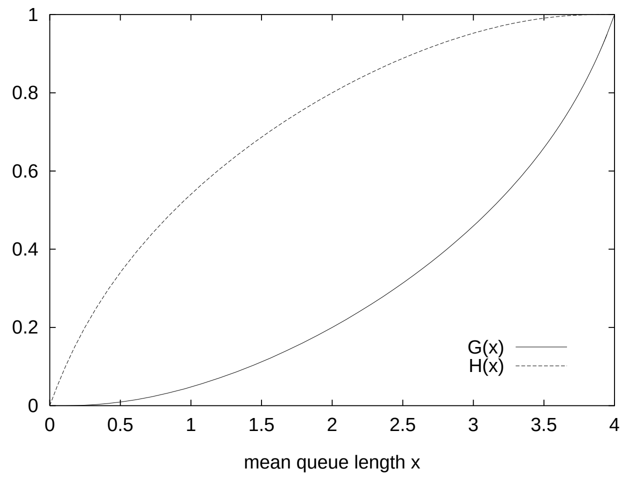


Figure 4: Characteristic functions $\widehat{G}(x)$ and $\widehat{H}(x)$ for the $M_t/M/1/4$ queue.

together with the result that ρ and x are dependent through equation (12), yields the functional forms for the characteristic functions

$$\widehat{G}(x) = 1 - P_0(L^{-1}(x)) \tag{21}$$

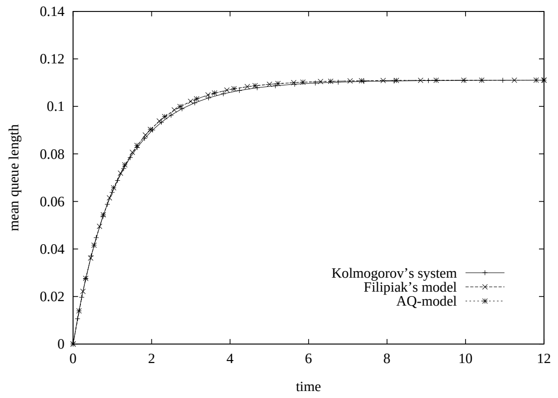
$$\widehat{H}(x) = P_K(L^{-1}(x)) \tag{22}$$

Figure 4 shows a plot of $\widehat{G}(x)$ and $\widehat{H}(x)$ for the $M_t/M/1/4$ queue. Note that both characteristic functions $\widehat{G}(x)$ and $\widehat{H}(x)$ are undefined for $x \geq 4$ because $L^{-1}(x)$ is asymptotic at $x = 4$ (see figure 2 for a plot of $L^{-1}(x)$). Also note that $\widehat{H}(x) = H(x)$.

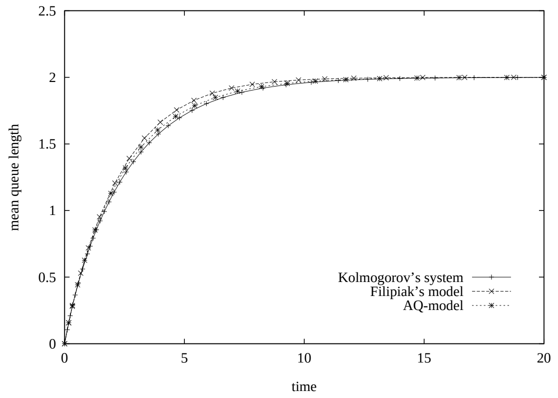
Evolutions of queue length for Filipiak's dynamic-flow model and the AQ-model for the $M/M/1/4$ are shown in figures 5(c) and 6(c).

4.3 Evaluating the characteristic functions

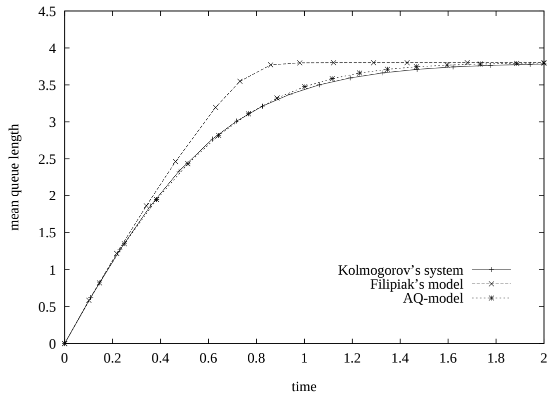
Filipiak's model and the AQ-model require the evaluation of the characteristic functions $G(x), H(x), \widehat{G}(x)$ and $\widehat{H}(x)$ and their dependence on $L^{-1}(x)$ consequently requires the evaluation of $L^{-1}(x)$. However $L^{-1}(x)$ cannot be expressed as a rational polynomial



(a) $\rho = 0.1$

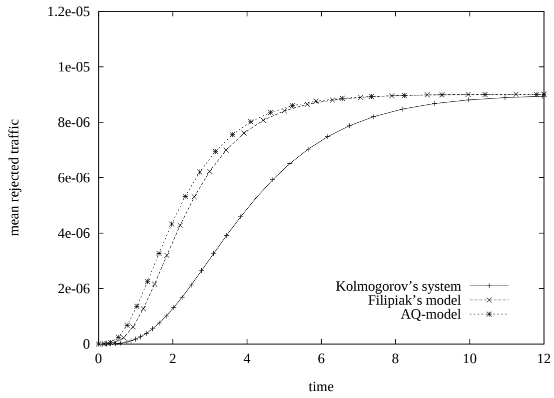


(b) $\rho = 1.0$

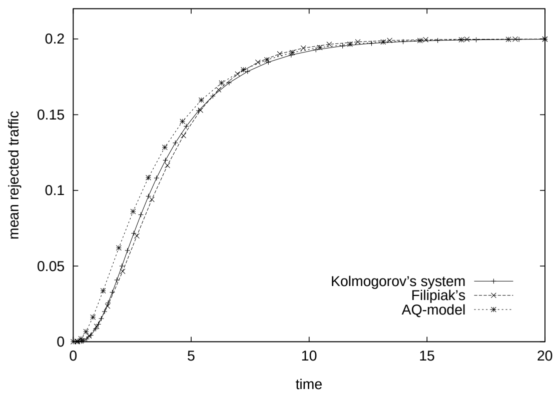


(c) $\rho = 6.0$

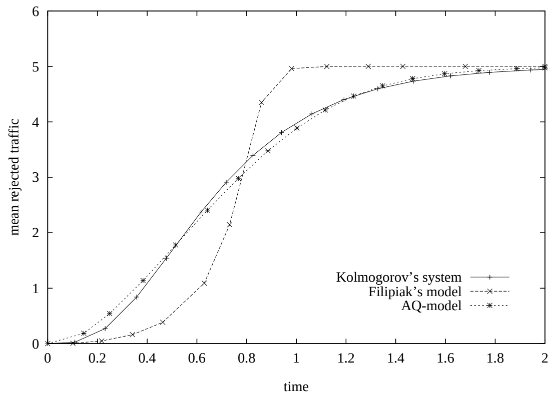
Figure 5: Evolution of queue length, $x(t)$, for the $M/M/1/4$ queue with initial condition $x(0) = 0$.



(a) $\rho = 0.1$



(b) $\rho = 1.0$



(c) $\rho = 6.0$

Figure 6: Evolution of mean rejected traffic, $\lambda^r(t)$, for the $M/M/1/4$ queue with initial condition $x(0) = 0$.

in x and therefore $L^{-1}(x)$ is determined by numerically inverting $L(\rho)$.

4.3.1 Evaluating $\mathbf{L}^{-1}(\cdot)$

The problem of inverting a function for an arbitrary value is reformulated to the problem of finding the root of an equation. Given the function $L(\rho)$ and an arbitrary value x, the value of $\rho = L^{-1}(x)$ is calculated by finding the root ρ that satisfies the equation

$$L(\rho) - x = 0 \tag{23}$$

Several numerical techniques are available for finding roots of non-linear equations. The choice is largely determined by whether or not the derivatives of the function are easily calculated. If the function's first derivative cannot be calculated, then either the *bisection method*, the *secant method*, the *false position method* or the *Wijngaarden-Dekker-Brent method* is used. The Wijngaarden-Dekker-Brent method (also called Brent's method) is often the method of choice due to its speed of convergence [PFTV88, p. 267].

4.3.2 Evaluating $\mathbf{L}(\cdot)$

Three numerical difficulties arise when evaluating $L(\cdot)$ using equation (13). First, numerical overflow occurs when evaluating ρ^K and ρ^{K+1} for large values of ρ and K. Second, numerical underflow occurs when evaluating ρ^K and ρ^{K+1} for small ρ and large K. Finally, the equation reduces to an indeterminate form when $\rho = 1$. The remainder of this section shows how these three problems are avoided.

Traffic intensity less than unity

Expanding equation (13) yields

$$L(\rho) = \frac{\rho - (K+1)\rho^{K+1} + K\rho^{K+2}}{1 - \rho - \rho^{K+1} + \rho^{K+2}} \tag{24}$$

If $\rho < 1$ then $L(\rho)$ is evaluated using the equation

$$L(\rho) = \frac{\rho(1 + \rho^K(\rho - (K+1)))}{1 - \rho(1 - \rho^K(\rho - 1))} \tag{25}$$

Accurate results are obtained even if numerical underflow occurs when evaluating ρ^K. Note that should numerical underflow occur then equation (25) reduces to the equation for the queue length for the $M/M/1$ queue, namely

$$L(\rho) = \frac{\rho}{1 - \rho}$$

(Unlike numerical overflows, numerical underflows caused by floating-point operations do not invoke exceptions in the C run-time environment and therefore the programmer need not code special routines to handle this type of exception.)

Traffic intensity greater than unity

If $\rho > 1$ then $L(\rho)$ is evaluated using the equation

$$L(\beta) = \frac{K + \beta(\beta^K - (K+1))}{1 - \beta(1 - \beta^K(\beta - 1))} \tag{26}$$

where $\beta = 1/\rho$. Equation (26) is obtained by setting $\beta = 1/\rho$ and rewriting equation (24). Equation (26) yields accurate results even if numerical underflow occurs when evaluating β^K.

Traffic intensity near unity

If $\rho = 1$ then the indeterminate form yielded by equation (24) is reduced to

$$L(\rho)\Big|_{\rho=1} = \frac{K}{2}$$

by applying L'hôpital's rule. However when ρ is near unity, equation (24) does not provide accurate results when evaluated using floating-point arithmetic. Accurate results are achieved by substituting $\rho = 1 + \alpha$ into equation (24) and expanding the factors $(1+\alpha)^{K+1}$ and $(1+\alpha)^{K+2}$ in terms of their binomial expansions to yield

$$L(\alpha) = \frac{K}{3}\frac{2\alpha K + \alpha + 3}{\alpha K + 2} \tag{27}$$

where $\alpha = \rho - 1$. This equation is valid for $|\alpha| \ll 1$.

The first four terms of the binomial expansions for the factors $(1+\alpha)^{K+1}$ and $(1+\alpha)^{K+2}$ are required because fewer terms do not remove the indeterminate form. The domain of $L(\alpha)$ for which equation (27) is sufficiently accurate, is determined by comparing this approximation to one of higher order (that is, an approximation containing more terms in the binomial expansion) for various values for ρ and K.

4.4 Approximating the characteristic functions

The disadvantage of evaluating the characteristic function using the root finding method found in section 4.3.1 is that substantial computation is required since equation (23) is computed many times until convergence for the root is attained.

This section investigates several methods for approximating arbitrary functions and compares these methods in terms of their computation for evaluating the characteristic function $G(x)$.

4.4.1 Interpolation

Given an arbitrary function $f(x)$ and a set of coordinates

$$\mathcal{S} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\}$$

such that $y_i = f(x_i)$ where $x_1 < x_2 < \ldots < x_i < \ldots < x_n$, the method of interpolation is used to estimate the value of the function between the coordinates.

Interpolation routines use the set of coordinates together with an *interpolating function.* Various functional forms are used for the interpolating function including polynomial functions, rational functions, trigonometric functions and cubic-spline functions. Both the accuracy and the computation for the interpolation method is dependent on the number of stored coordinates, the greater the number of coordinates, the greater the accuracy and the greater the computation.

The remainder of this section discusses how the polynomial interpolation method[1] is used to approximate $G(x)$. The coordinates (x_i, y_i) for the set $\mathcal{S}$ (where $y_i = G(x_i)$) are calculated parametrically using the equation

$$(x_i, y_i) \quad = \quad (L(\rho_i), \rho_i) \tag{28}$$

A small set of coordinates is calculated such that the interpolation routine yields a given accuracy[2]. The advantage of calculating the coordinates parametrically is that the computationally expensive root finding method is avoided.

4.4.2 Least-squares polynomial approximation

The least-squares polynomial approximation method can be applied to both discrete and continuous data. An approximation function for discrete data is determined such that the *sum* of the square of the errors is minimized while the approximation function for continuous data is determined such that the *integral* of the square of the error is minimized.

We use the method for discrete data because accurate coordinates for the characteristic function $G(x)$ are easily generated (see equation 28). The minimization function for coordinates $\mathcal{S}$ (where $y_i = G(x_i)$) is given by

$$\mathcal{M} \quad = \quad \sum_{i=1}^{N}(y_i - a_0 - a_1 x_i - \cdots - a_m x_i^m)^2$$

[1]The accuracy of polynomial interpolation does not necessarily increase with the order of the polynomial function and therefore the order of the function should not exceed 5 unless rigorous monitoring of errors is undertaken [PFTV88].

[2]The author has developed an algorithm that finds a small set of N coordinates such that the interpolation routine yields a given maximum error. Although the algorithm does not find the smallest set, the algorithm shows that the required maximum error is achieved using no more than N coordinates.

where m is the order of the polynomial ($m < N$). Standard techniques of calculus lead to a system of linear equations which uniquely determine the coefficients a_i [Sch68].

The polynomial approximation method has two disadvantages.

- Except for very low degree polynomials, the system of linear equations is *ill-conditioned* [Sch68]. This means that, although the minimizing function defines the coefficients a_i uniquely, in practice it may be impossible to determine these coefficients. Standard methods for solving linear systems either produce no solution at all or else magnify numerical errors. This disadvantage is avoided by introducing orthogonal polynomials, for example, Chebyshev polynomials (see section 4.4.3).
- Polynomials are not suited to approximating functions that contain poles. This particular problem arises when approximating the characteristic function $G(x)$, which contains a pole at $x = K$ (see figure 2). Although not investigated in this thesis, functions containing poles are approximated more accurately by rational functions.

The remainder of this section proposes a modification for the polynomial approximation method and shows how this proposed method is used to approximate $G(x)$. Inspection of $G(x)$ in figure 2 leads us to propose that a tan function could be used as a first order approximation for $G(x)$. Therefore we propose that the approximation function has the form

$$\widetilde{G}(x) \quad = \quad b\tan(ax) + \sum_{k=0}^{n} c_k x^k$$

which we call the *tan approximation.* The coefficient a is chosen such that the asymptote of $G(x)$ coincides with $x = K$. This implies that $a = \pi/(2K)$. Coefficient b is calculated by fitting $b\tan(ax)$ to the coordinates $\mathcal{S}$ (where $y_i = G(x_i)$), while the coefficients c_k are determined by fitting a polynomial to the coordinates $(x_i, y_i - b\tan(ax_i))$ for $1 \leq i \leq N$.

The reason for coefficient b being determined prior to the coefficients c_k is that if the coefficients are calculated simultaneously using the fitting program Fudgit [Lac93], a singular matrix is returned.

4.4.3 Chebyshev polynomial approximation

The Chebyshev polynomial approximation method is a special case of the least-squares method. The method requires finding M Chebyshev coefficients c_k for an arbitrary function $f(x)$ such that

$$f(x) \approx \left(\sum_{k=0}^{M-1} c_k T_k(x)\right) - \frac{c_0}{2}$$

where $T_k(x)$ is the Chebyshev polynomial of degree k given by

$$T_k(x) = \cos(k \arccos x)$$

The disadvantage of this method is that the computation required to determine the coefficients is dependent on the computation for evaluating $f(x)$. That is, if $f(x)$ is computationally expensive then the computation for the Chebyshev coefficients is also expensive.

The advantage of this method is that once these coefficients are known, the Chebyshev approximated function is evaluated reasonably quickly (see table 4). If the coefficients are decreasing rapidly (which is the typical case), the computation is reduced by using only the first m coefficients where $m \leq M - 1$. The computation may be further reduced by transforming the Chebyshev coefficients to polynomial coefficients thereby yielding a polynomial approximation.

4.4.4 Computation comparisons

The four evaluation methods—interpolation, least-squares polynomial approximation, tan approximation and Chebyshev approximation—are compared in terms of the time

		polynomial approximation	tan approximation	Chebyshev approximation	polynomial interpolation
$0 \leq G(x) \leq 1$	relative speed	7	13	14	64
	maximum error	1.42×10^{-3} #1	1.08×10^{-3}	0.239×10^{-3}	0.513×10^{-3}
	coefficients	5	6 #2	7 #3	7 #4
$0 \leq G(x) \leq 2$	relative speed	8	13	14	66
	maximum error	0.657×10^{-3}	0.333×10^{-3}	0.701×10^{-3}	0.838×10^{-3}
	coefficients	7	8	8	13
$0 \leq G(x) \leq 10$	relative speed	-	-	34	68
	maximum error	-	-	1.04×10^{-3}	1.22×10^{-3}
	coefficients	-	-	29	37

Table 4: Comparison of evaluation methods to approximate the characteristic function $G(x)$.

required to evaluate $G(x)$ for the $M_t/M/1/4$ queue. The evaluation routines[3] are implemented in C and compiled using the Borland C++ (version 1.00) for OS/2 (version 2.1). The hardware platform consists of an IBM compatible PC with a 80386DX CPU and a 8087 coprocessor. The results for the comparison are shown in table 4.

Table 4 is constructed as follows. Filipiak's approximation for $G(x)$ for the $M_t/M/1/4$ queue[4], valid for the range $0 \leq \rho \leq 1$ yields a maximum error of $\epsilon = 1.42 \times 10^{-3}$. See #1 in table 4.

The remaining approximations are determined by finding the the least number of coefficients for each method that yield a maximum error no greater than ϵ. For example, to achieve a maximum error less than ϵ for the range $0 \leq G(x) \leq 1$, the tan approximation requires 6 polynomial coefficients, the Chebyshev approximation method requires 7 coefficients and the interpolation method requires 7 coordinates (see #2, #3 and #4 in the table).

The computation required for these approximations is denoted by *relative speed*. The

[3]The coefficients for the polynomial approximation and the tan approximation are calculated using the fitting program `Fudgit` [Lac93] and the Chebyshev coefficients are calculated using the routine `chebft` [PFTV88]. The Chebyshev approximation and the interpolation method are evaluated using the routines `chebev` and `polint` [PFTV88].

[4]Filipiak's approximation is given by the 5th order polynomial $\widetilde{G}(x) = 0.00086 + 0.96457x - 0.79800x^2 + 0.58778x^3 - 0.22297x^4 + 0.03523x^5$. See equation (3.33) in [Fil88].

following conclusions are made regarding the computation for the evaluation methods.

- The least-squares polynomial approximation method requires the least computation for both ranges $0 \leq G(x) \leq 1$ and $0 \leq G(x) \leq 2$. However, the fitting program Fudgit is unable to calculate the coefficients for the range $0 \leq G(x) \leq 10$. A possible reason for this limitation of Fudgit is that the asymptotic behaviour of $G(x)$ yields a system of ill-conditioned equations represented by the internal data structures of the program.

- The computation for each method increases as the required range of $G(x)$ increases from $0 \leq G(x) \leq 1$ to $0 \leq G(x) \leq 10$. The reason for this increase in computation is that more coefficients are required to ensure accuracy as the range of the function increases and particularly as the function is approximated closer to the pole at $x = K$.

Chapter 5

Approximate models for the $\mathbf{M}^X/\mathbf{E}_Y/\mathbf{1}/\mathbf{K}$ queue

Two approximate models for the $M^X/E_Y/1/K$ queue, developed by Moore [Moo75], are investigated in this chapter. The first model, the *stationary arrival model* assumes that the arrival and service processes are homogeneous, while the second model, the *nonstationary arrival model*, approximates a non-homogeneous arrival process by dividing the time axis into intervals, such that the arrival process in each interval is considered homogeneous.

The $M^X/E_Y/1/K$ queue reduces to the $M/M/1/K$ queue as a special case and therefore Moore's model may be compared with Filipiak's model and the AQ-model found in the previous chapter. These three models are compared in terms of their accuracy and computation in chapter 6.

5.1 Description for the $\mathbf{M}^X/\mathbf{E}_Y/\mathbf{1}/\mathbf{K}$ queue

The arrivals of groups of customers form a Poisson process with inter-arrival parameter λ. The number of customers in each group is a random variable X, which may take on any positive integral value with probability $g_j = P(X = j)$. This process is called a *compound Poisson process* and the distribution for inter-arrival times for this process is given by the compound Poisson distribution function

$$f_{A|T}(j|t) \;=\; \sum_{k=0}^{j} e^{-\lambda t}\frac{(\lambda t)^k}{k!}g_j^{(k*)} \tag{29}$$

for $j \geq 0$ [GH74]. The factor $g_j^{(k*)}$ is the k-fold convolution of distribution g_j with itself and denotes the probability $P(\sum_{i=1}^{k} X_i = j)$. The distribution $f_{A|T}(j|t)$ denotes the probability that there are j arrivals in a time interval, given that the time interval has length t.

The distribution for service times is modelled by the weighted-sum Erlang distribution function

$$h_T(t) \;=\; \sum_{k=1}^{\infty} c_k e^{-\mu t}\frac{\mu^k t^{k-1}}{(k-1)!} \tag{30}$$

for $t \geq 0$. The constant c_k is a weight for the Erlang distribution $e^{-\mu t}\mu^k t^{k-1}/(k-1)!$ that is subject to the condition $\sum_{k=1}^{\infty} c_k = 1$. The weighted-sum Erlang distribution is a general form for the Erlang distribution. The Erlang distribution models a range of service time distributions, containing the exponential distribution as a special case and the deterministic distribution as a limiting case [Coo81, p. 66]. The flexibility of the weighted-sum Erlang distribution is further emphasized by Gaver [Gav54] and Luchak [Luc56].

5.2 Review of the imbedded Markov chain method

Both Moore's stationary and nonstationary models use the imbedded Markov chain method to calculate the distribution for queue length. The imbedded Markov chain method was introduced by Kendall to study the $M/G/1$ and $GI/M/s$ queues [Ken51, Ken53]. Kendall's method for the $M/G/1$ queue is reviewed as follows.

Consider a departure instant for the $M/G/1$ queue. The length of time until the next *arrival* is independent of the departure instant because the arrival process is Markovian. Similarly, the length of time until the next *departure* is independent of the departure instant because the service time is an independent random variable. Therefore the length of time until the next event, whether it be an arrival or a departure, is independent of the departure instant. Therefore the future evolution of the queueing process, from the departure instant, is independent of time.

A *renewal point* for a stochastic process with time parameter t, is an instant at which the future evolution of the process depends only on the state of the process at time t. Therefore the departure instants of the $M/G/1$ queue are renewal points. Consequently the theory of Markov chains may be used to analyse the state of the $M/G/1$ queue at the departure instants.

Let $T_n, n \geq 0$ be the instant of the nth departure and let the state of the process be defined by the queue length Q_n immediately after T_n. Further, let A_n denote the number of customers arriving between departure instants T_{n-1} and T_n. The queue length Q_n is calculated from Q_{n-1} using the equation

$$Q_n = \begin{cases} Q_{n-1} - 1 + A_n, & \text{if } Q_{n-1} \geq 1, \\ A_n, & \text{if } Q_{n-1} = 0, \end{cases} \tag{31}$$

which expresses the transition of the queue length from one departure instant to the next.

The queue length distribution at the departure instants is given by the Chapman-Kolmogorov equation

$$P(Q_n = j) = \sum_{i=0}^{\infty} P(Q_n = j \mid Q_{n-1} = i)\, P(Q_{n-1} = i) \tag{32}$$

where equation (31) yields

$$P(Q_n = j \mid Q_n = 0) = P(A_{n-1} = j)$$

$$P(Q_n = j \mid Q_n = i) \;\; = \;\; P(A_{n-1} = j - i + 1) \tag{33}$$

for $i \geq 1$.

Let $L_j(n) = P(Q_n = j)$ be the jth component of $\vec{L}(n)$ and let $a_j(n) = P(A_n = j)$. Equation (32) together with equation (33) may be written in matrix notation as

$$\vec{L}(n) \;\; = \;\; \vec{L}(n-1) \cdot \mathbf{A}(n) \tag{34}$$

where matrix $\mathbf{A}(n)$ is given by

$$\mathbf{A}(n) \;\; = \;\; \begin{bmatrix} a_0(n) & a_1(n) & a_2(n) & \cdots \\ a_0(n) & a_1(n) & a_2(n) & \cdots \\ 0 & a_0(n) & a_1(n) & \cdots \\ 0 & 0 & a_0(n) & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

If the queue has finite capacity K, then equation (31) becomes

$$Q_n \;\; = \;\; \begin{cases} \min(Q_{n-1} - 1 + A_n, K-1), & \text{if } Q_{n-1} \geq 1, \\ \min(A_n, K-1), & \text{if } Q_{n-1} = 0, \end{cases} \tag{35}$$

and matrix $\mathbf{A}(n)$ becomes

$$\mathbf{A}(n) \;\; = \;\; \begin{bmatrix} a_0(n) & a_1(n) & a_2(n) & \cdots & a_{K-2}(n) & d_{K-2}(n) \\ a_0(n) & a_1(n) & a_2(n) & \cdots & a_{K-2}(n) & d_{K-2}(n) \\ 0 & a_0(n) & a_1(n) & \cdots & a_{K-3}(n) & d_{K-3}(n) \\ 0 & 0 & a_0(n) & \cdots & a_{K-4}(n) & d_{K-4}(n) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_0(n) & d_0(n) \end{bmatrix}$$

where $d_j(n) \; = \; 1 - \sum_{m=0}^{j} a_m(n)$. Note that the queue length immediately after a departure is at most $K-1$.

Calculating the matrix A

This section shows how the components of matrix $\mathbf{A}(n)$, namely $a_j(n)$, are calculated for both the $M^X/E_Y/1$ queue and the $M/M/1$ queue. Note that the components $a_j(n)$ for the finite capacity $M^X/E_Y/1/K$ queues are identical to the components $a_j(n)$ for the infinite capacity $M^X/E_Y/1$ queue.

The distribution for the number of arrivals A_n between departure instants T_{n-1} and T_n is obtained using the Stieltjes integral[1]

$$P(A_n = j) \quad = \quad \int_0^\infty f_{A|T}(j|t)\, dH_T(t) \tag{36}$$

Substituting equations (29) and (30) into equation (36) yields, after some manipulation,

$$P(A_n = j) \quad = \quad \sum_{r=0}^{j} g_j^{(r*)} \left(\frac{\lambda}{\lambda+\mu}\right)^r \sum_{k=1}^{\infty} \binom{r+k-1}{r} \left(\frac{\mu}{\lambda+\mu}\right)^k c_k \tag{37}$$

for $j \geq 0$.

The $M^X/E_Y/1$ queue reduces to the $M/M/1$ queue in the special case that the arrival groups have size $X = 1$ and the service time distribution reduces to the exponential distribution. (The weighted-sum Erlang distribution reduces to the exponential distribution if $c_1 = 1$ and $c_k = 0$ for $k \neq 1$.) Then

$$f_{A|T}(j|t) \quad = \quad e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

and

$$dH_T(t) \quad = \quad \mu e^{-\mu t}\, dt$$

and therefore equation (37) reduces to the geometric distribution

$$P(A_n = j) \quad = \quad \left(\frac{\lambda}{\lambda+\mu}\right)^j \left(\frac{\mu}{\lambda+\mu}\right) \tag{38}$$

for $j \geq 0$.

We now have expressions for the components $a_j(n) = P(A_n = j)$ of matrix $\mathbf{A}(n)$ for the $M^X/E_Y/1$ (equation (37)) and the $M/M/1$ queue (equation (38)).

[1]The Stieltjes integral is used with generality for discrete, continuous and mixed distributions. In the case where the density function exists, $dH_T(t)$ is replaced by $h_T(t)dt$ yielding the Riemann integral.

5.3 Moore's stationary arrival model

Moore's stationary arrival model for the $M^X/E_Y/1/K$ queue assumes that the arrival and service processes are homogeneous [Moo75]. The imbedded Markov chain method is used to calculate the queue length distribution $L_j(n)$ for $0 \leq j \leq K-1$ at each departure instant T_n for $n > 0$.

The imbedded Markov chain method has the disadvantage that the association between the time axis and the queue length distribution is lost. In other words, the queue length distribution at each departure instant T_n is known, but the time t for which the distribution is valid is not known because T_n is a random variable.

A second disadvantage for the Markov chain method is that the queue length distribution is calculated at *departure* instants, rather than arrival instants from which more meaningful performance measures are calculated. Note that the queue length distribution, as observed at arriving instants, can be calculated from the queue length distribution, as observed at departure instants, under the restrictive condition that the queueing system is a birth-death process and is in equilibrium [Coo81, p. 185].

5.3.1 Derivation

Let Z_n be the server idle time from T_{n-1} until the *next arrival* into the queue. The Z_n's comprise a sequence of identically distributed random variables having a common distribution

$$P(Z_n \leq t) = 1 - e^{-\lambda t} \tag{39}$$

for $t \geq 0$ with mean

$$E[Z_n] = \frac{1}{\lambda} \tag{40}$$

Let X_n be the server idle time in the interval (T_{n-1}, T_n). Therefore we have

$$X_n = \begin{cases} Z_n, & \text{if } Q_{n-1} = 0, \\ 0, & \text{if } Q_{n-1} > 0, \end{cases} \tag{41}$$

with mean

$$E[X_n] \quad = \quad E[Z_n]\, L_0(n-1) \tag{42}$$

Let S_n be the nth customer's service time. Now the nth departure instant is given by the recursive equation[2]

$$T_n \quad = \quad T_{n-1} + X_n + S_n \tag{43}$$

for $n \geq 1$. The expectation of T_n is given by

$$E[T_n] \quad = \quad E[T_{n-1}] + E[X_n] + E[S_n] \tag{44}$$

The queue length distribution $L_j(n) = P(Q_n = j)$ is obtained by calculating the distributions for X_n, Q_n and T_n sequentially for $n = 1, 2, 3, \ldots$, each time conditioning on the appropriate preceding distributions. However, this method is computationally expensive and therefore Moore makes the assumption that the queue length distribution *is* $L_j(n)$ when $t = E[T_n]$.

Moore's stationary arrival model is given by equations (34), (42) and (44), that is

$$\begin{aligned} E[T_n] &= E[T_{n-1}] + E[X_n] + E[S_n] \\ \vec{L}(n) &= \vec{L}(n-1) \cdot \mathbf{A}(n) \end{aligned}$$

where

$$E[X_n] \quad = \quad E[Z_n]\, L_0(n-1) \tag{45}$$

and where $E[T_0]$ and $\vec{L}(0)$ are given initial conditions.

5.3.2 Performance measures

The mean and variance for queue length (as observed at departure instants) is obtained directly from the queue length distribution $\vec{L}(n)$. The system utilization (immediately

[2]This relationship was used by Takács to study the departure process for the $M/G/1$ queue [Tak61].

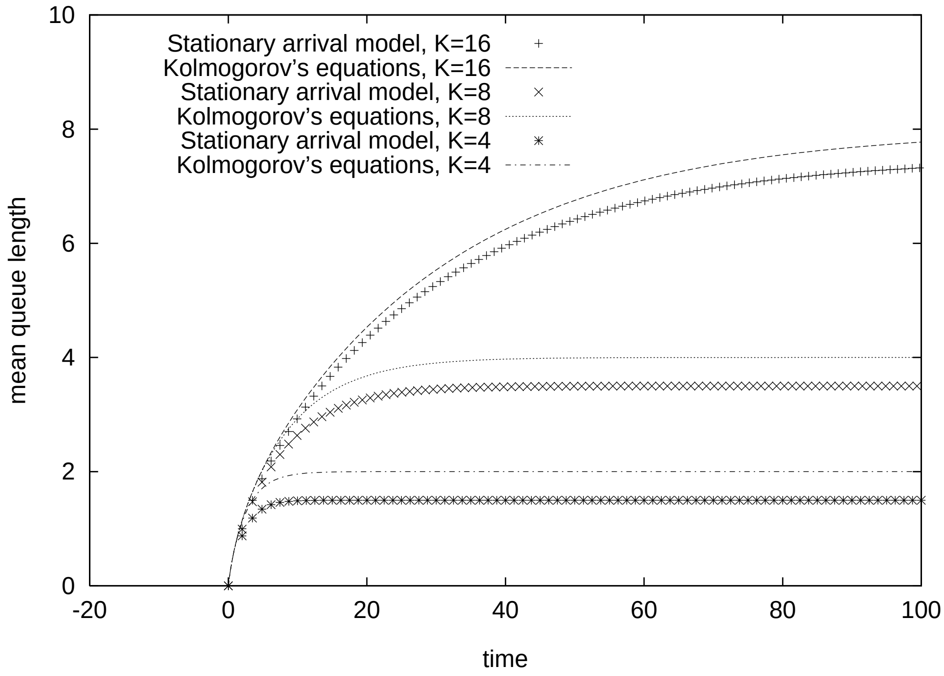


Figure 7: Comparison of mean queue length between Moore's stationary arrival model and Kolmogorov's system of equations for the $M/M/1/K$ queue.

after the nth departure) is defined as

$$U_n = \begin{cases} 1, & \text{if } Q_n > 0, \\ 0, & \text{if } Q_n = 0. \end{cases}$$

Therefore

$$E[U_n] = 1 - L_0(n)$$

and

$$\text{var}[U_n] = \Big(1 - L_0(n)\Big) L_0(n)$$

The reader is referred to [Moo75] for the derivation of other performance measures.

Figure 7 compares the evolution of queue length for both Moore's stationary arrival model and the solution for Kolmogorov's system of differential equations for the

$M/M/1/K$ queue. The graphs of the queue length for Moore's model are predictably lower than the solutions for Kolmogorov's system because the maximum number of customers as seen by a departing customer is at most $K-1$. However as K increases the relative error between the models decreases. (The time axis is given in units of mean service time $1/\mu$.)

5.3.3 Implementing the model

The following algorithm is used to solve Moore's model.

STATIONARY_ARRIVALS($E[T_0], \vec{L}(0), t_{\max}$)
BEGIN
    $t := E[T_0]$;
    n := 1;
    # Calculate components $a_j(n)$ of matrix $\mathbf{A}$
    FOR j FROM 0 TO $K-2$
        $a_j := P(\mathcal{A}=j)$;
    WHILE ($t \leq t_{\max}$)
    BEGIN
        $t := t + L_0(n)E[Z_n] + E[S_n]$;
        $\vec{L}(n) := \vec{L}(n-1)\cdot\mathbf{A}$;
        $n := n+1$;
        Calculate_Performance_Measures($\vec{L}(n), t$);
    END;
END.

The matrix notation $\vec{L}(n) := \vec{L}(n-1)\mathbf{A}$ is used for clarity, however the regular structure of matrix $\mathbf{A}(n)$ allows the computational operation

$$L_0(n) \quad := \quad \Big(L_0(n-1) + L_1(n-1)\Big)\, a_0$$

$$L_j(n) \;:=\; \Big(L_0(n-1)+L_1(n-1)\Big)\,a_j + \sum_{i=2}^{j+1} L_i(n-1)a_{j-i+1}$$

for $1 \le j \le K-2$ and

$$L_{K-1}(n) \;:=\; \Big(L_0(n-1)+L_1(n-1)\Big)\left(1-\sum_{m=0}^{K-2} a_m\right) + \sum_{i=2}^{K-1} L_i(n-1)\left(1-\sum_{m=0}^{K-i-1} a_m\right)$$

5.3.4 Modification of Moore's model for the $M/M/1/K$ queue

We modify Moore's stationary arrival model to calculate the approximate queue length, as observed at arrival instants, from the queue length distribution $\vec{L}(n)$. This modification ensures that Moore's model yields the same steady-state results for queue length as Filipiak's model and the AQ-model and allows these models to be compared in terms of accuracy.

Let A be the queue length mean for the $M/M/1/K$ queue under steady-state conditions. Let B be the queue length mean for the $M/M/1/K$ given by Moore's model. The modified queue length mean χ_n for the nth departure is given by

$$\chi_n \;=\; \frac{A}{B}\,E[X_n]$$

where $E[X_n]$ is given by equation (45).

The steady-state value A is calculated using equation (13) while B is given by the mean for the queue length distribution $\vec{L}$ where $\vec{L}$ is the solution for the equation

$$\vec{L} \;=\; \vec{L}\cdot\mathbf{A} \tag{46}$$

Note that equation (46) is the steady-state equivalent of equation (34). Equation (46) is solved by numerically iterating the equation until convergence is attained.

5.4 Moore's nonstationary arrival model

Moore's *nonstationary arrival model* for the $M_t^X/E_Y/1/K$ queue approximates a nonhomogeneous arrival process by dividing the time axis into intervals $I_i(1 \le i \le N+1)$

where the arrival process is considered to be homogeneous in each interval [Moo75]. The arrival process in each interval is modelled by the compound Poisson distribution with inter-arrival parameter λ_i and group size probability mass function $g_{ij} = P(X_i = j)$. The service process is homogeneous and identical to the service process for the stationary arrival model.

5.4.1 Derivation

The random variables—T_n, Q_n, Z_n, X_n and A_n—are defined in the same way as the stationary model. The random variables Q_n, X_n and T_n satisfy the same equations as the stationary model, that is, equations (35), (41) and (43). However the random variables Z_n and A_n have new dependences. Z_n is dependent on the departure instant T_{n-1} because Z_n is dependent on λ_i through the equation $P(Z_n \leq t) = 1 - e^{-\lambda_i t}$ (see equation (39)) and λ_i is dependent on the interval in which T_{n-1} occurs. Similarly A_n, the number of arrivals during the nth customer's service, is dependent not only on the length of service but also on the time at which the service commences, given by $T'_n = T_n + X_n$.

Let $\mathcal{Z}(t)$ be the length of time from an *arbitrary* instant t to the first arrival. Further let $\mathcal{A}(t)$ be the number of customers arriving in the interval $[t, S(t)]$ where $S(t)$ is the time at which a customer completes its service given that it arrives at time t. The dependence of Z_n on T_{n-1} is expressed as

$$Z_n = \mathcal{Z}(T_{n-1}) \tag{47}$$

and the dependence of A_n on T'_n is expressed as

$$A_n = \mathcal{A}(T'_n) \tag{48}$$

where

$$T'_n = T_{n-1} + X_n \tag{49}$$

The evaluation of the queue length distribution $L_j(n) = P(Q_n = j)$ requires the evaluation of the distributions for Z_n, X_n, T'_n, A_n, Q_n and T_n sequentially for $n = 1, 2, 3, \ldots$, each time conditioning on the appropriate preceding distributions. This method to calculate $\vec{L}(n)$ is computationally expensive and the computation is reduced by the following approximations,

$$\begin{aligned}
\mathcal{Z}(T_{n-1}) &\approx \mathcal{Z}(E[T_{n-1}]) \\
\mathcal{A}(T'_n) &\approx \mathcal{A}(E[T'_n])
\end{aligned}$$

Moore's nonstationary arrival model is obtained by substituting these approximations into equations (47) and (48) and using equations (34), (42) and (44) to yield

$$\begin{aligned}
E[T_n] &= E[T_{n-1}] + E[X_n] + E[S_n] \\
E[X_n] &= E[Z_n]L_0(n-1) \\
\vec{L}(n) &= \vec{L}(n-1)\mathbf{A} \\
E[Z_n] &= E[\mathcal{Z}(E[T_{n-1}])] \\
E[T'_n] &= E[T_{n-1}] + E[X_n]
\end{aligned}$$

where the components of matrix $\mathbf{A}$ are $a_j(n) = P(\mathcal{A}(E[T'_n]) = j)$.

Expressions for $P(\mathcal{A}(t) = j)$ and $E[\mathcal{Z}(t)]$ are derived in appendix C. In particular, $P(\mathcal{A}(t) = j)$ is given by equations (88) and (98) and $E[\mathcal{Z}(t)]$ is given by equations (109) and (110).

5.4.2 Implementing the model

Moore gives the following algorithm for solving the nonstationary arrival model.

NONSTATIONARY_ARRIVALS($E[T_0], \vec{L}(0), t_{\max}$)
BEGIN
    $t := E[T_0]$;
    n := 1;
    WHILE ($t \leq t_{\max}$)
    BEGIN

```
        t := t + L_0(n)E[Z(t)];
        # Calculate components a_j(n) of matrix A(n)
        FOR j FROM 0 TO K − 2
            a_j(n) := P(A(t) = j);
        L⃗(n) := L⃗(n − 1) · A(n);
        t := t + E[S_n];
        n := n + 1;
        Calculate_Performance_Measures(L⃗(n), t);
    END;
END.
```

5.5 Discussion

Moore's model facilitates the modelling of a wider range of queueing processes than the $M/M/1/K$ queue. Furthermore Moore's models, unlike Filipiak's model, calculates the queue length *distribution*, from which several performance measures may be calculated.

However, the disadvantage of Moore's models, which originates from the embedded Markov chain method, is that the analysis of the queue length distribution takes place at *departure* instants. The queue length distribution, as observed at arrival instants, cannot in general be calculated from the distribution (as observed at departure instants), with the exception that the queueing process is a birth-death process under steady-state conditions. A further disadvantage is that Moore's model requires greater programming effort that the models found in the previous chapter.

Chapter 6

Comparison of models for the M/M/1/K queue

This chapter compares the approximate models investigated in chapters 4 and 5. In particular, Filipiak's model, the AQ-model and Moore's modified model are compared in terms of their accuracy and computation for various values for K and ρ. Note that Moore's model for the $M^X/E_Y/1/K$ queue reduces to a model for the $M/M/1/K$ queue as a special case.

6.1 Definition of modelling error

The *modelling error* for an approximate model is defined as follows. Let $Q(t)$ be the queue length yielded by an exact model and let $\widetilde{Q}(t)$ be the queue length for the approximate model. The *modelling error* ϵ_Q is given by

$$\epsilon_Q = \max_t \left(\left| Q(t) - \widetilde{Q}(t) \right| \right)$$

which expresses the maximum absolute difference between exact and approximate values of the queue length over a period of time[1]. This definition is used to compare the accuracy for the approximate models for the $M/M/1/K$ queue and is also used in chapter 7 to compare several approximate models for a network of $M/M/1/K$ queues.

6.2 Calculation of modelling error

The modelling error for the approximate models is calculated for each pair (K, ρ) on the grid $g = \{(K, \rho) : K = 2, 3, \ldots, 11; \ \rho = 0.2, 0.6, \ldots, 3.8\}$ given that the queue is empty at time $t = 0$ and the service rate $\mu = 1$.

Kolmogorov's system of differential equations is used to calculate the exact value of the queue length. Filipiak's model and the AQ-model are compared with Kolmogorov's system by calculating the queue length for each model at equally spaced intervals along the time axis. The differential equations are integrated using the Runga-Kutta routine `rkqc` where the accuracy parameter `eps` $= 10^{-4}$. The characteristic functions used in Filipiak's model and the AQ-model (namely $G(\cdot), H(\cdot), \widehat{G}(\cdot)$ and $(\widehat{H}(\cdot))$ are evaluated using the Chebyshev evaluation routine `chebev` with 20 coefficients. (The `rkqc` and `chebev` routines are found in [PFTV88].)

However Moore's model is compared with Kolmogorov's system by calculating the queue length at the *departure instants* yielded by Moore's model.

6.3 Calculation of computation

The time required to compute the modelling error is measured for each pair (K, ρ) on the grid g. This time includes the initialization of data structures as well as the calculation of Chebyshev coefficients for the characteristic functions but does not include any input/output activity (all i/o routines are removed).

[1]For an alternative modelling error, refer to [AC89, equation (14)].

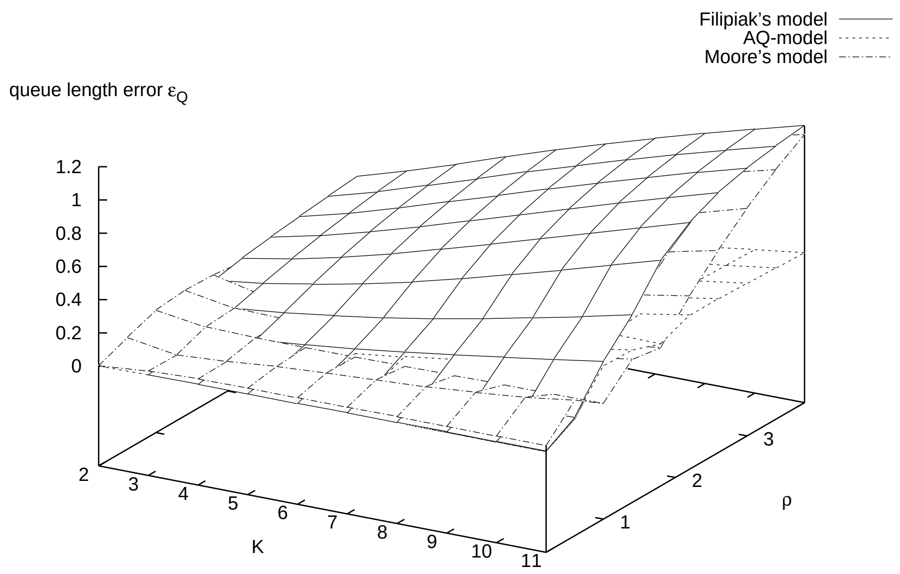


Figure 8: Comparison of queue length error for 3 models for the $M/M/1/K$ queue.

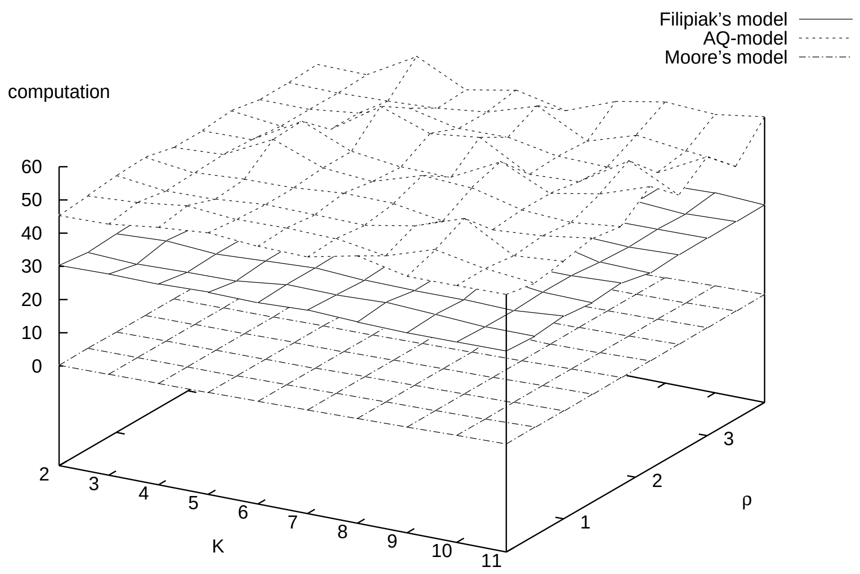


Figure 9: Comparison of computation for 3 models for the $M/M/1/K$ queue.

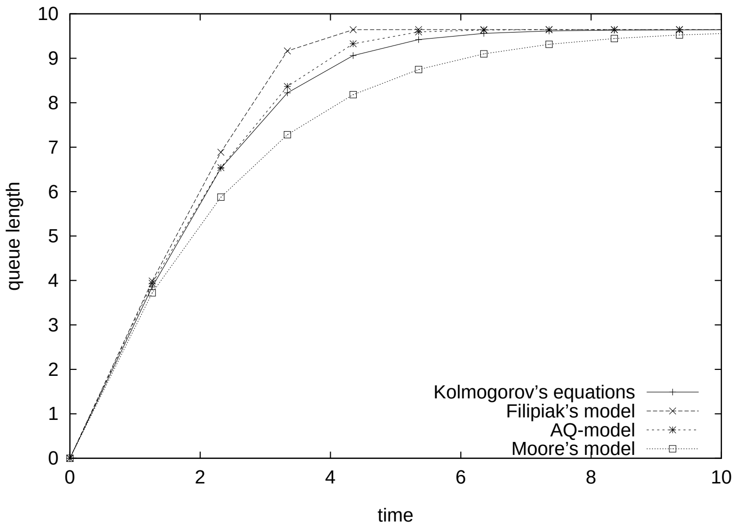


Figure 10: Comparison of evolution of queue length for 3 models for the $M/M/1/10$ queue ($\rho = 3.8$).

The programs are compiled using the GNU C compiler (v2.6) for the Unix (System V release 4.0) operating system running on a Sparcserver 1000 consisting of 4 50MHz CPUs and 256 Mbytes of main memory. The Unix `time` command is used to measure the elapsed time for the user process.

6.4 Results

Figure 8 compares the modelling error for the three approximate models. This figure shows that the AQ-model yields accurate results as ρ increases while Moore's model and Filipiak's model become increasingly inaccurate as ρ increases.

The computation for the approximate models is given in figure 9 and shows that Moore's model requires far less computation than the other models. The computation surfaces are plotted *relative* to each other, that is, the units of time are unspecified.

Figure 10 compares the evolution of queue length for a large value for ρ ($\rho = 3.8$). Note that the maximum error between the graphs for an approximate model and Kolmogorov's system corresponds to a single point on a surface in figure 8. The units of time are the same as the units for the mean service time $1/\mu$.

Finally, figure 11 shows a scatter plot of computation versus queue length error. This graph shows that although Moore's model requires less computation than the other models, it is not as accurate as the AQ-model. Note that the computation and error data are identical to the data used in figures 8 and 9.

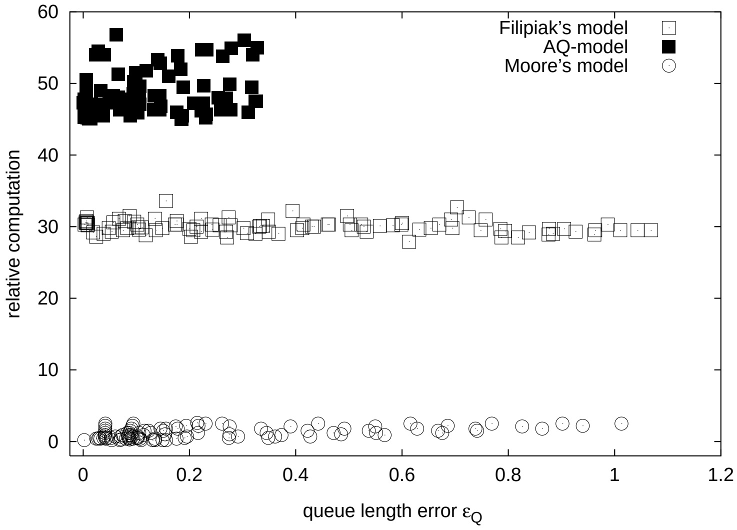


Figure 11: Computation versus queue length error for 3 models for the $M/M/1/K$ queue.

Chapter 7

Approximate models for transient networks

Three approximate models of a packet-switched communication network are compared in terms of accuracy and computation. This chapter is organized as follows. First, the importance of transient queueing network models is highlighted. Second, the exact model together with the three approximate models—the TS-model, Filipiak's model and the AQ-model—are presented. Finally the approximate models are compared in terms of their accuracy and computation for a small communication network under various conditions.

We show that under the conditions examined, the AQ-model affords greater accuracy than Filipiak's model while having a similar computational requirement. The TS-model is the most accurate model but has considerably greater computational requirements that the other models.

7.1 The queueing network model for a packet-switched communication network

Consider a packet-switched communication network consisting of M transmission links. Each transmission link is modelled by a single-server queue with index i where $1 \leq i \leq M$. Queue i has finite capacity K_i and exponentially distributed service times with mean $1/\mu_i$. Let n_i denote the number of customers either being served or waiting for service at queue i and let the network state descriptor be given by $\mathbf{n} = (n_1, n_2, \ldots, n_M)$.

External customers entering the network via queue i form a nonstationary Poisson process with parameter $\nu_i(t)$. Let $\alpha_{i,j}(t)$ denote the probability that a customer, upon completing service at queue i, proceeds directly to queue j. The probability that a customer leaves the network from queue i is given by $1 - \sum_{j=1}^{M} \alpha_{i,j}(t)$. Transit times between service centres are zero. Finally, any customer that is blocked at a queue is rejected from the network.

7.2 Discussion

The property that blocked customers are immediately rejected from the network is not suitable for modelling *circuit-switched* communication networks because circuit-switched networks typically route blocked traffic to alternative queues for further processing.

Note that the queueing discipline at each queue is not specified. The reason for this omission is because the approximate models (discussed in this thesis) calculate *queue length* and this particular performance measure is independent of the queueing discipline.

7.3 Four transient network models

This section presents four models for the transient queueing network described above[1]. Kolmogorov's system of equations is used as the exact model, while the remaining three models—the TS-model, Filipiak's model and the AQ-model—are approximate. The TS-model is attributed to Tipper & Sundareshan in [TS90], Filipiak's model is found in [Fil88] and the AQ-model is developed in section 4.2.

7.3.1 The exact model

Kolmogorov's system of differential equations

Let $P_{\mathbf{n}}(t)$ denote the probability that the queueing network is in state $\mathbf{n}$ at time t. Let $\mathbf{1}_j$ denote a unit vector in the j direction. Kolmogorov's system of forward differential equations for the queueing network is given by

$$\begin{aligned}
\frac{d}{dt}P_{\mathbf{n}}(t) \quad = \quad & -\left(\sum_{i=1}^{M} \mu_i \, \epsilon(n_i > 0) + \sum_{i=1}^{M} \nu_i(t) \, \epsilon(n_i < K_i)\right) \, P_{\mathbf{n}}(t) \\
& + \sum_{i=1}^{M} \nu_i(t) \, \epsilon(n_i > 0) \, P_{\mathbf{n}-\mathbf{1}_i}(t) \\
& + \sum_{i=1}^{M} \mu_i \left(1 - \sum_{j=1}^{M} \alpha_{i,j}(t)\right) \, \epsilon(n_i < K_i) \, P_{\mathbf{n}+\mathbf{1}_i}(t) \\
& + \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} \mu_i \, \alpha_{i,j}(t) \, \epsilon(n_i < K_i) \, \epsilon(0 < n_j < K_j) \, P_{\mathbf{n}+\mathbf{1}_i-\mathbf{1}_j}(t) \\
& + \sum_{i=1}^{M} \sum_{j=1, j\neq i}^{M} \mu_i \, \alpha_{i,j}(t) \, \epsilon(n_i < K_i) \, \epsilon(n_j = K_j) \, P_{\mathbf{n}+\mathbf{1}_i}(t) \\
& + \sum_{i=1}^{M} \mu_i \, \alpha_{i,i}(t) \, \epsilon(n_i > 0) \, P_{\mathbf{n}}(t) \qquad (50)
\end{aligned}$$

[1] Other transient queueing network models, although not investigated in this chapter, are found in [AC89, Har81, Cha77].

where

$$\epsilon(\mathcal{S}) = \begin{cases} 1, & \text{if } \mathcal{S} \text{ is true,} \\ 0, & \text{if } \mathcal{S} \text{ is false.} \end{cases}$$

The indicator function $\epsilon(\mathcal{S})$ ensures that a customer cannot leave a queue if there are no customers in the queue and also ensures that a customer cannot join a queue if the queue is full.

The interpretation of the right hand side of equation (50) is as follows. The first term represents the probability of a transition from state $\mathbf{n}$ to another state. The second term is the probability of an external customer arriving to the network. The third term is the probability of a customer departing from the network. The fourth term is the probability of a customer completing service at queue i and being routed to queue j. The fifth term is the probability of a customer competing service at queue i and being rejected at queue j while the sixth term is the probability of a customer completing service at queue i and being routed back to queue i.

Kolmogorov's system comprises $\prod_{i=1}^{M}(K_i+1)$ differential equations, one for each system state. The practicality of solving this system is limited owing to the dimensionality of the system state space.

Performance measures

Let $Q_i(t)$ denote the expected number of customers in queue i given by

$$Q_i(t) = \sum_{n=1}^{K_i} n \sum_{\{\mathbf{n}:\, n_i=n\}} P_{\mathbf{n}}(t)$$

The expected number of customers in the network is given by

$$Q_N(t) = \sum_{i=1}^{M} Q_i(t)$$

7.3.2 The approximate TS-model

In order to reduce the computation of the exact model, Tipper & Sundareshan [TS90] we assume that customers departing from a service centre form a Poisson process. This assumption has been shown to be reasonably accurate by Taafe in a study of the output process of nonstationary finite capacity Markov systems [Taa82]. Furthermore this assumption is weaker than the commonly made assumption, that the queues are independent. The latter assumption is not made by this model (as can be seen by the dependence between queues in the TS model – see equation (52)).

The introduction of the Poisson assumption allows each queue to be modelled as an $M/M/1/K_i$ queue and therefore the TS-model consists of fewer equations than Kolmogorov's system. In particular, the TS-model consists of $\sum_{i=1}^{M}(K_i+1)$ differential equations compared with Kolmogorov's system of $\prod_{i=1}^{M}(K_i+1)$ differential equations. Note that the Poisson approximation becomes increasingly accurate as K increases and is exact in the limit (if no feedback occurs) so that the output process of the infinite capacity $M/M/1$ queue *is* a Poisson process [Kan92, p. 162].

System of differential equations

Let $P_{n,i}(t)$ denote the probability that there are n customers in queue i at time t. The system of differential equations for the queue length probability for queue i is given by

$$\begin{aligned}
\frac{d}{dt}P_{0,i}(t) &= -\lambda_i(t)P_{0,i}(t) + \mu_i P_{1,i}(t) \\
\frac{d}{dt}P_{K_i,i}(t) &= \lambda_i(t)P_{K_i-1,i}(t) - \mu_i P_{K_i,i}(t) \\
\frac{d}{dt}P_{n,i}(t) &= -\Big(\lambda_i(t)+\mu_i\Big)P_{n,i}(t) + \lambda_i(t)P_{n-1,i}(t) + \mu_i P_{n+1,i}(t)
\end{aligned} \tag{51}$$

for $1 < n < K_i$. The arrival rate at queue i is given by

$$\lambda_i(t) = \nu_i(t) + \sum_{\ell=1}^{M}\mu_\ell\Big(1-P_{0,\ell}(t)\Big)\alpha_{\ell,i}(t) \tag{52}$$

The first term on the right of equation (52) represents the rate at which external customers arrive to the network via queue i. The second term represents the arrival rate of traffic from other queues and is given by the service rate multiplied by the probability that the server is busy. The reader is referred to [TS90] for further discussion.

Performance measures

The expected number of customers at queue i is given by

$$Q_i(t) = \sum_{n=1}^{K_i} n\, P_{n,i}(t)$$

The performance measures, carried traffic and rejected traffic, are given by

$$\begin{aligned} \lambda_i^c(t) &= \mu_i \left(1 - P_{0,i}(t)\right) \\ \lambda_i^r(t) &= \lambda_i(t)\, P_{K_i,i}(t) \end{aligned}$$

7.3.3 Filipiak's approximate dynamic-flow model

Filipiak extends the dynamic-flow model for a single queue to a queueing network model (see chapter 4). Filipiak's network model accommodates two routing mechanisms, namely *adaptive routing* and *alternate routing.* Adaptive routing refers to routing where the rate at which customers are routed is measured in terms of time dependent *probabilities.* Alternate routing on the other hand refers to *deterministic* routing of blocked customers to alternative service centres. While both forms of routing may be included in Filipiak's model, the model presented in this section considers only adaptive routing.

System of differential equations

Filipiak's model is given by M differential equations, one for each communication link

$$\frac{d}{dt}x_i(t) = -\mu_i\, G_i\left(x_i(t)\right) + \lambda_i(t)$$

for $1 \leq i \leq M$. The total arrival rate $\lambda_i(t)$ at each queue is

$$\lambda_i(t) = \nu_i(t) + \sum_{\ell=1}^{M} \mu_\ell \, G_\ell\left(x_\ell(t)\right) \Big(1 - H_\ell\left(x_\ell(t)\right)\Big) \, \alpha_{\ell,i}(t) \tag{53}$$

The first term on the right represents the arrival rate of customers arriving to the network while the second term represents the arrival rate of customers from the other queues. Note the similarity between equation (53) and equation (52) of the TS-model. The reader is referred to equations (8) and (10) in section 4.1.1 and equation (4.3a) in [Fil88].

Performance measures

The performance measures—queue length, carried traffic and rejected traffic—at each queue are given by

$$\begin{aligned} Q_i(t) &= x_i(t) \\ \lambda_i^c &= \mu_i \, G_i(x_i(t)) \Big(1 - H_i(x_i(t))\Big) \\ \lambda_i^r &= \mu_i \, G_i(x_i(t)) \, H_i(x_i(t)) \end{aligned} \tag{54}$$

7.3.4 The approximate AQ-model

We generalize the AQ-model developed in chapter 4 to model a queueing network.

System of differential equations

The AQ-model for a queueing network is given by M differential equations

$$\frac{d}{dt} x_i(t) = -\mu_i \, \widehat{G}_i\left(x_i(t)\right) + \lambda_i(t) \Big(1 - \widehat{H}_i\left(x_i(t)\right)\Big) \tag{55}$$

for $1 \leq i \leq M$ where

$$\lambda_i(t) = \nu_i(t) + \sum_{\ell=1}^{M} \mu_\ell \, \widehat{G}_\ell\left(x_\ell(t)\right) \alpha_{\ell,i}(t) \tag{56}$$

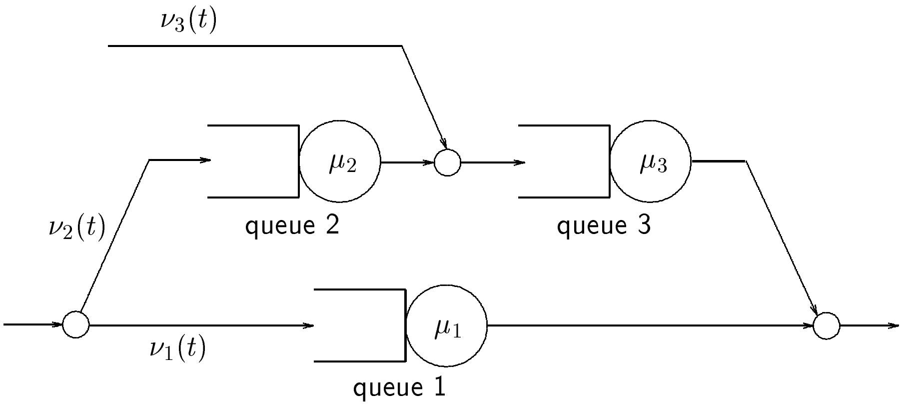


Figure 12: Queueing network for a packet-switched communication network.

Equation (56) is identical in form to equation (53), consisting of the arrival rate of customers entering the network and the arrival rate of customers from other queues. The form of equation (55) is described in chapter 4.

Performance measures

The performance measures for a particular queue are identical to those found in chapter 4 and are given by

$$\begin{aligned} Q_i(t) &= x_i(t) \\ \lambda_i^c &= \mu_i \widehat{G}_i(x_i(t)) \\ \lambda_i^r &= \mu_i \widehat{H}_i(x_i(t)) \end{aligned} \tag{57}$$

7.4 Comparison of approximate models

Due to the dimensionality of the state space of Kolmogorov's system, the comparison is restricted to a small communication network (see figure 12). This network models a primary communication link (queue 1) and an alternative path consisting of two links (queue 2 & 3) via a secondary node. The traffic stream $\nu_3(t)$ represents traffic disruptive to stream $\nu_2(t)$.

Let ϵ_{Q_i} denote the modelling error for queue i (see section 6.1) and let $\epsilon_Q = \max_i(\epsilon_{Q_i})$. (The following parameter values for figure 12 are used: $\mu_1 = \mu_2 = \mu_3 = 1$, $\nu_1 = 0.6$ and $K_1 = K_2 = K_3$.)

Filipiak's model

The modelling error for Filipiak's model is shown in figures 13 to 16. Figures 13 to 15 show the error ϵ_{Q_i} for each *individual* queue while figure 16 shows the combined error ϵ_Q for all queues in the network. The independence of ϵ_{Q_1} on the arrival rates ν_2 and ν_3 is clearly shown and has a value of $\epsilon_{Q_1} = 0.0383$ (figure 13). Figure 14 shows the dependence of ϵ_{Q_2} on ν_2 and figure 15 shows the dependence of ϵ_{Q_3} on ν_2 and ν_3.

Figure 17 shows the evolution of the queue length at queue 3 for the pair $(\nu_2, \nu_3) = (1, 1)$ and $(\nu_2, \nu_3) = (4, 4)$ associated with figure 15. Note that the maximum value attained by the error function is plotted on the vertical axis of figure 15.

The AQ-model

Graphs of the modelling error for the AQ-model are shown in figures 18 to 21. The modelling error shown in figure 18 is $\epsilon_{Q_1} = 0.0332$. Figure 20 shows that ϵ_{Q_3} becomes smaller as either ν_2 or ν_3 approaches $\mu = 1$ and figure 21 shows the combined error ϵ_Q for all queues in the network.

The TS-model

The queue length modelling error for the TS-model is shown in figure 22. The modelling error for queues 1 and 2 is zero since the TS-model and Kolmogorov's system yield identical results for independent queues. Consequently the combined error ϵ_Q for the

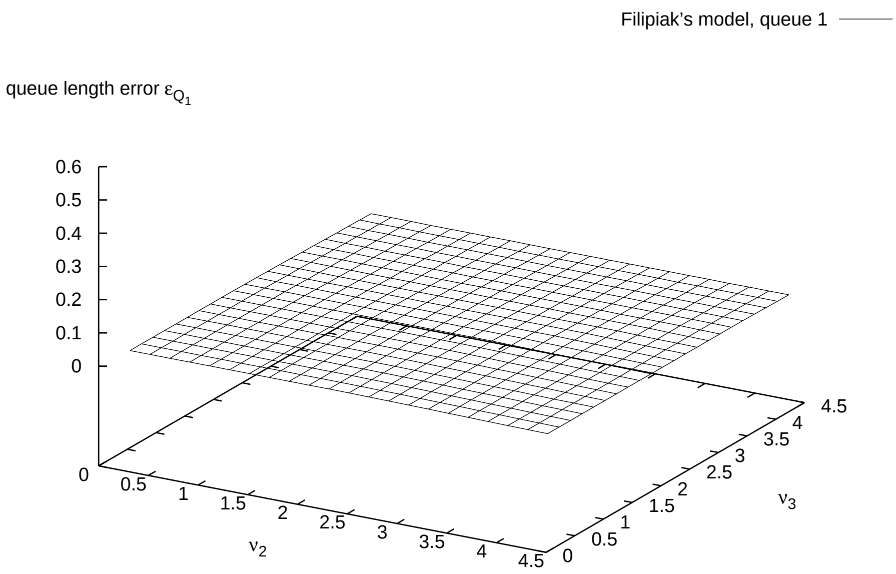


Figure 13: Queue length error for queue 1 (Filipiak's model).

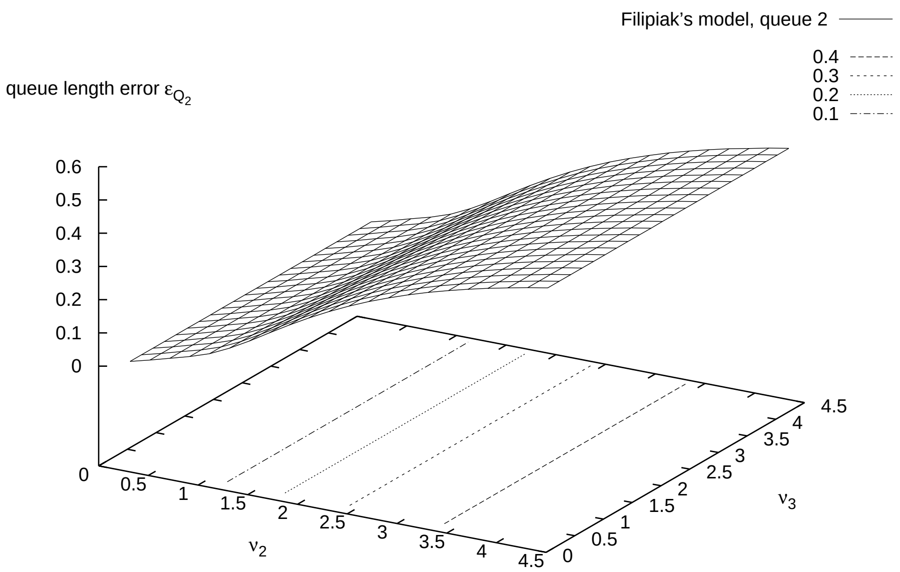


Figure 14: Queue length error for queue 2 (Filipiak's model).

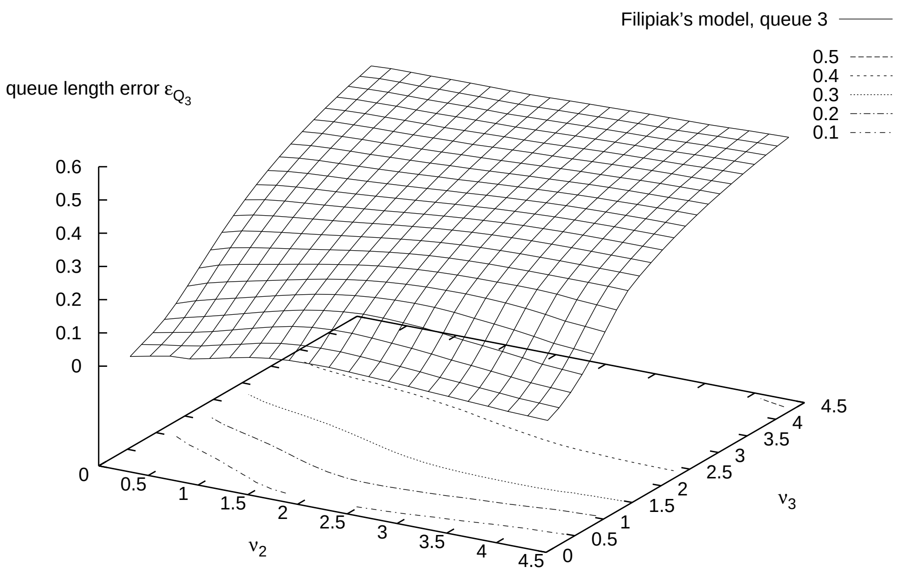


Figure 15: Queue length error for queue 3 (Filipiak's model).

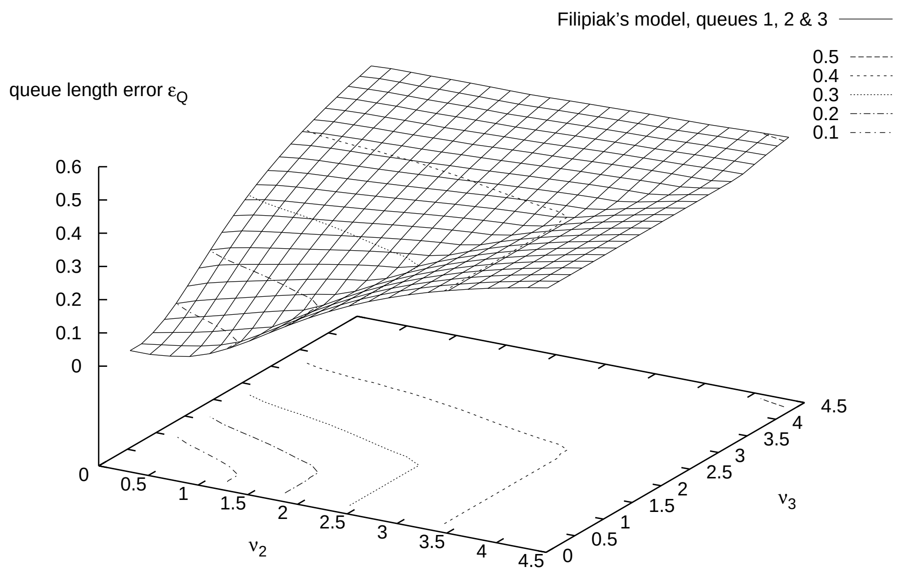
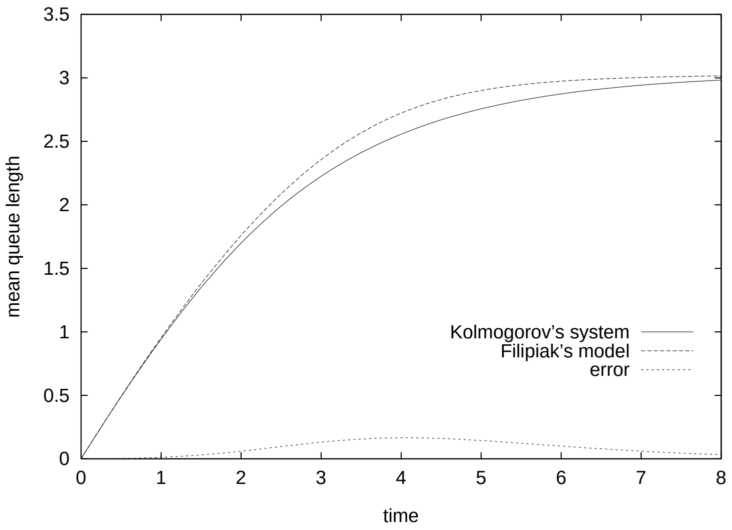
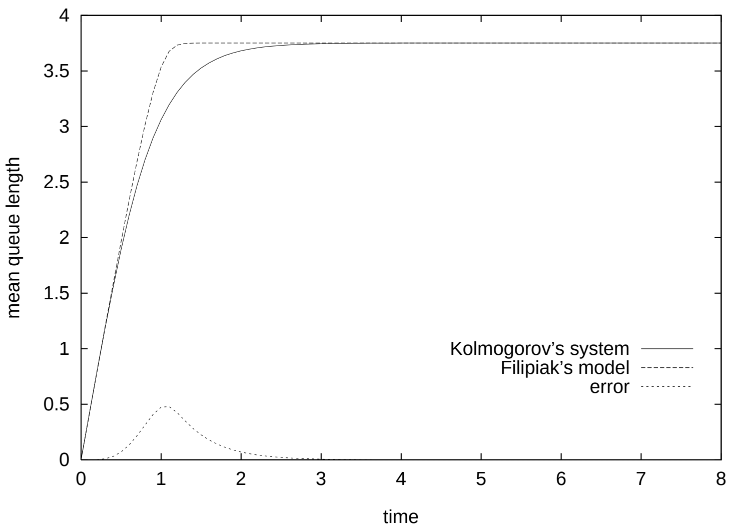


Figure 16: Queue length error for all queues (Filipiak's model).



(a) $\nu_2 = 1, \nu_3 = 1$



(b) $\nu_2 = 4, \nu_3 = 4$

Figure 17: Evolution of average queue length for queue 3 (Filipiak's model).

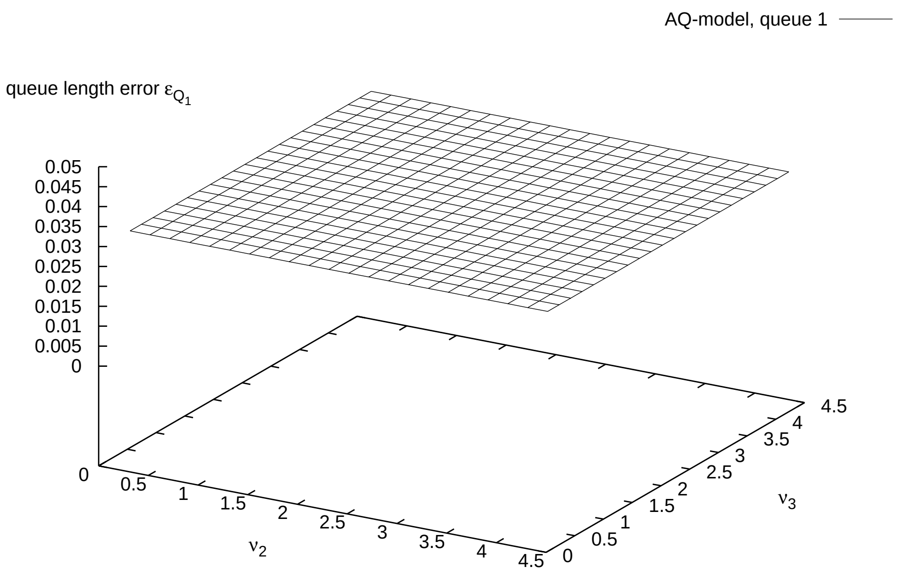


Figure 18: Queue length error for queue 1 (AQ-model).

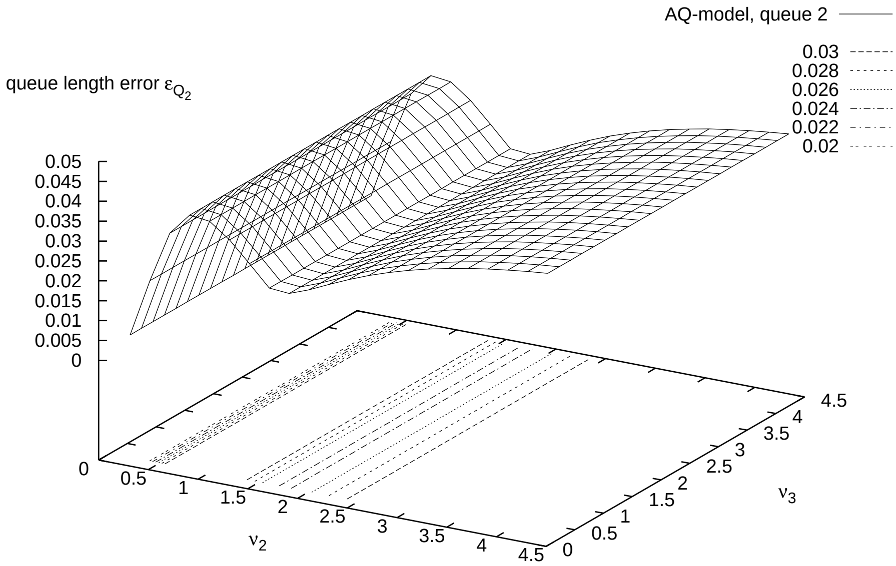


Figure 19: Queue length error for queue 2 (AQ-model).

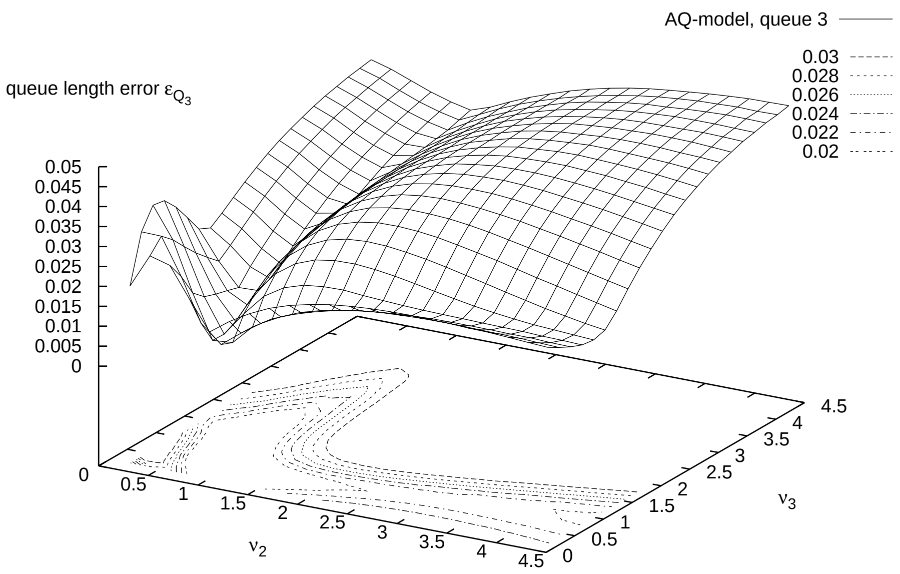


Figure 20: Queue length error for queue 3 (AQ-model).

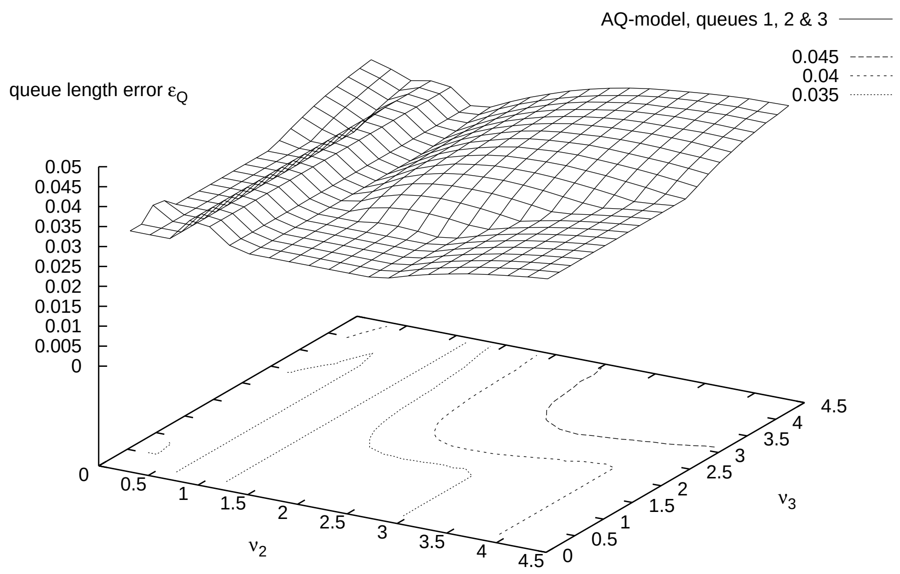


Figure 21: Queue length error for all queues (AQ-model).

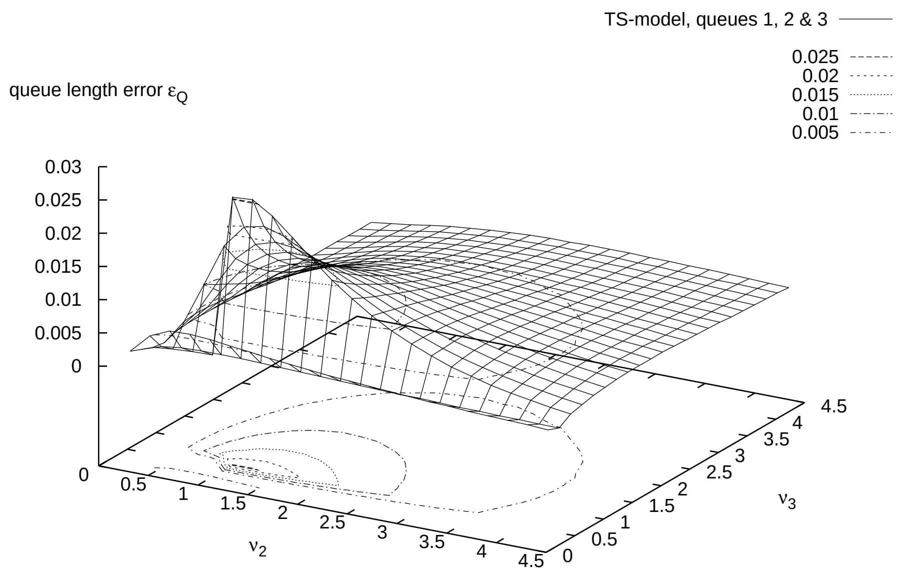


Figure 22: Queue length error for all queues (TS-model).

network is identical to ϵ_{Q_3}. The peak of the graph is centred over the pair $(\nu_2, \nu_3) = (1.0, 0.6)$. Whether the value $\nu_3 = 0.6$ is related to the value $\nu_1 = 0.6$ is not known and further analysis is required.

Comparisons

A comparison of the modelling error for the three approximate models is given in figure 23 which brings together the graphs of figures 16,21 and 22. The error surfaces for figures 23 listed in order of increasing accuracy are: Filipiak's model, the AQ-model and the TS-model[2].

A second comparison of these models for $K_1 = K_2 = K_3 = 8$ (figure 24) shows that the order of the models in terms of accuracy is maintained.

Figure 25 compares the computation required for the exact and approximate models for various values of $K_1 = K_2 = K_3$. The graph shows the elapsed time for the calculation of the queue length for 800 values of time (equally spaced with length 1). The method described in section 6.3 is used to measure the elapsed time.

These figures show that the TS-model is the most accurate *approximate model*. This is not surprising, since this model makes the least assumptions about the stochastic process. However accuracy is not the only consideration to be made when comparing these models. The computation of the models, which is dependent on the number of differential equations, must be taken into account. Both Filipiak's model and the AQ-model require M differential equations while the TS-model requires $\sum_{i=1}^{M}(K_i + 1)$ differential equations. We therefore conclude that the AQ-model provides a good compromise between accuracy and computation for modelling communication networks.

[2]The plotting program Gnuplot does not yield the correct key for figures 23 and 24.

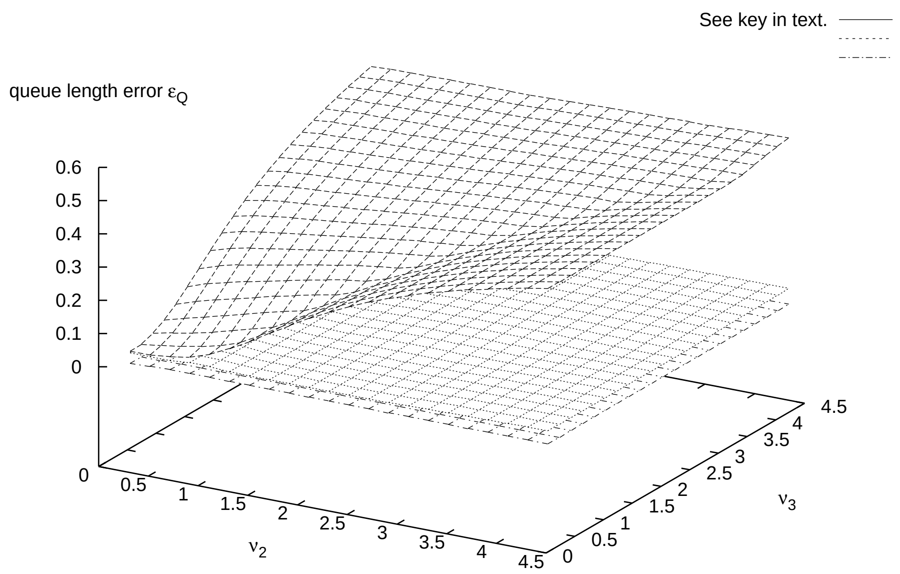


Figure 23: Comparison of queue length error for $K_1 = K_2 = K_3 = 4$.

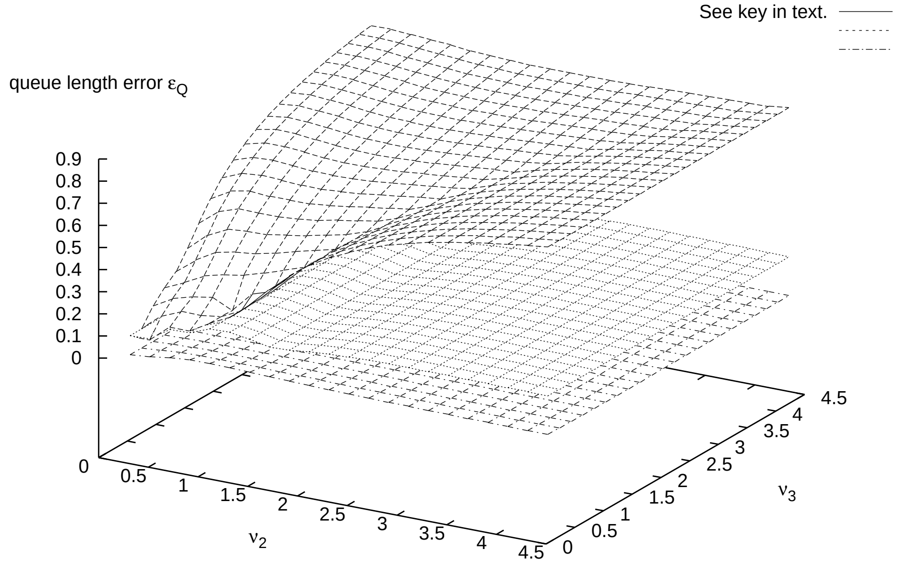


Figure 24: Comparison of queue length error for $K_1 = K_2 = K_3 = 8$.

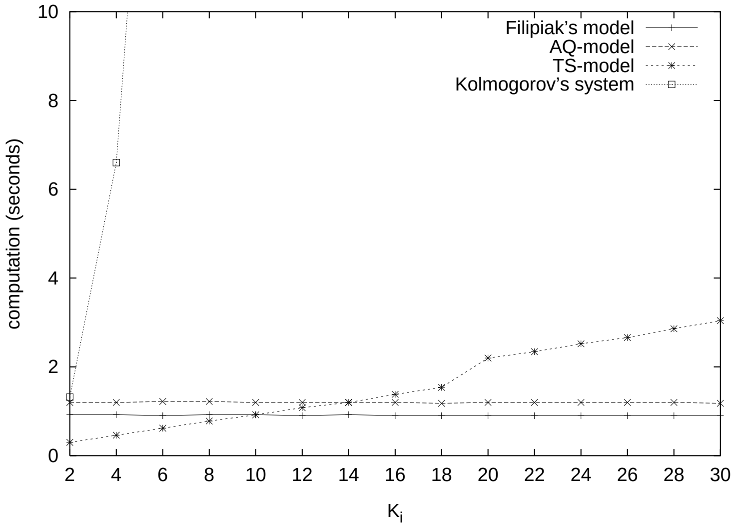


Figure 25: Comparison of computation for various $K_1 = K_2 = K_3$.

Chapter 8

Conclusions

We have investigated several approximate models for transient queueing systems to determine which models are best suited for modelling communication networks. We give a summary of our investigation.

A literature survey of various transient queueing models is given. A class of models consisting of a system of differential equations are solved by numerical integration. This class includes Rider's model for the $M_t/M_t/1$ queue, Rothkopf & Oren's model for the $M_t/M_t/s$ queue, Clark's model for the $M_t/M_t/s$ queue and Filipiak's *dynamic-flow* models. The advantage of this class is that non-homogeneous arrival processes may be modelled.

An exact model for the transient $M/M/1/K$ queue is required to calculate the accuracy for the approximate models. We examine the closed-form expression for the queue length distribution for the transient $M/M/1/K$ queue given by Takács and list several numerical difficulties that are encountered when evaluating this expression. An alternative method, requiring the numerical integration of Kolmogorov's system of differential equations, is easier to implement than the closed-form expression because of the availability of integration routines.

We investigate Filipiak's *dynamic-flow model* for the $M_t/M/1/K$ queue. This model consists of a differential equation that is integrated to yield approximate results for mean queue length, mean carried traffic and mean rejected traffic. A new model, the AQ-model (approximate queue model), is developed from Filipiak's model.

We investigate Moore's approximate model for the $M^X/E_Y/1/K$ queue. Moore's model facilitates the modelling of a wider range of queueing processes than the $M/M/1/K$ queue. Furthermore Moore's model, unlike Filipiak's model, calculates the queue length *distribution*, from which several performance measures may be determined. The disadvantage of Moore's model however, is that the analysis of the queue length distribution takes place at *departure* instants while the (more useful) queue length distribution, as observed at at arrival instants, is not determined. We modify Moore's *stationary arrival* model to approximate the mean queue length (as observed at arrival instants) from the queue length distribution yielded by Moore's model.

We compare Filipiak's model, the AQ-model and Moore's model in terms of accuracy and computation for the $M/M/1/K$ queue and show that the AQ-model is the most accurate while Moore's model is the most efficient in terms of computation.

Finally we compare Filipiak's model, the AQ-model and the TS-model (Tipper & Sundareshan's model) for a small packet-switched communication network under various conditions. The comparison shows that the models listed in order of decreasing accuracy are: the TS-model, the AQ-model and Filipiak's model. On the other hand, in terms of computation, Filipiak's model and the AQ-model have similar requirements while the TS-model has a significantly greater requirement. We conclude that the AQ-model provides a good compromise between accuracy and computation for communication networks.

Appendix A

Results for transient queueing systems

A.1 The M/M/1 queue

Kolmogorov's system of forward DEs

$$
\begin{aligned}
\frac{d}{dt}P_0(t) &= -\lambda P_0(t) + \mu P_1(t) \\
\frac{d}{dt}P_n(t) &= \lambda P_{n-1}(t) - (\lambda+\mu)\, P_n(t) + \mu P_{n+1}(t)
\end{aligned}
\tag{58}
$$

for $n \geq 1$.

Queue length distribution

The probability that the queue length is n at time t given that the initial queue length is j is given by [Tak62, Gif78, Kle75, GH74, Saa61]

$$
p_{jn}(t) = e^{-(\lambda+\mu)t} \left(\left(\frac{\lambda}{\mu}\right)^{(n-j)/2} I_{n-j}\left(2\sqrt{\lambda\mu}\, t\right) + \left(\frac{\lambda}{\mu}\right)^{(n-j+1)/2} I_{n+j+1}\left(2\sqrt{\lambda\mu}\, t\right) \right.
$$

$$+\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n}\sum_{m=n+j+2}^{\infty}\left(\frac{\lambda}{\mu}\right)^{-m/2} I_m\left(2\sqrt{\lambda\mu}\,t\right)\Bigg) \qquad (59)$$

where I_m is the modified Bessel function of the first kind of order m.

An equivalent expression is terms of trigonometrical terms is given by [Tak62, p. 23]

$$\begin{aligned} p_{jn}(t) &= P_n + 2e^{-(\lambda+\mu)t}\left(\frac{\lambda}{\mu}\right)^{(n-j)/2}\int_0^{\pi}\frac{e^{2\sqrt{\lambda\mu}\,t\,cos\theta}}{1-2\sqrt{\lambda/\mu}\,cos\theta+\lambda/\mu} \\ &\cdot\left(\sin j\theta-\sqrt{\lambda/\mu}\,\sin(j+1)\theta\right)\left(\sin n\theta-\sqrt{\lambda/\mu}\,\sin(n+1)\theta\right)\,d\theta \qquad (60) \end{aligned}$$

where

$$P_n = \begin{cases} (1-\lambda/\mu)(\lambda/\mu)^n & \text{for } \lambda<\mu \\ 0 & \text{for } \lambda\geq\mu \end{cases}$$

A further equivalent expression in terms of Marcum's Q-function is given by [JCJ80, Rob94]

$$\begin{aligned} p_{jn}(t) &= \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n} Q\left(\sqrt{2\mu t},\sqrt{2\lambda t}\right) \\ &+e^{-(\lambda+\mu)t}\Bigg(\left(\frac{\lambda}{\mu}\right)^{(n-j)/2} I_{n-j}\left(2\sqrt{\lambda\mu}\,t\right) \\ &+\left(\frac{\lambda}{\mu}\right)^{(n-j-1)/2} I_{n+j+1}\left(2\sqrt{\lambda\mu}\,t\right) \\ &-\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n}\sum_{m=0}^{n+j+1}\left(\frac{\lambda}{\mu}\right)^{-m/2} I_m\left(2\sqrt{\lambda\mu}\,t\right)\Bigg) \qquad (61) \end{aligned}$$

where

$$Q(a,b) = \exp\left(-\frac{a^2+b^2}{2}\right)\sum_{m=0}^{\infty}\left(\frac{a}{b}\right)^{m} I_m(a,b)$$

Queue length mean

The mean queue length at time t given that the initial queue size is j is given by [Tak62, p. 27]

$$\begin{aligned} E[Q(t) \mid Q(0) = j] &= L + 2e^{-(\lambda+\mu)t} \left(\frac{\lambda}{\mu}\right)^{(1-j)/2} \int_0^{\pi} \frac{e^{2\sqrt{\lambda\mu}\, t\, cos\theta} \cdot \sin\theta}{\left(1 - 2\sqrt{\lambda/\mu}\, cos\theta + \lambda/\mu\right)^2} \\ &\quad \cdot \left(\sin j\theta - \sqrt{\lambda/\mu}\, \sin(j+1)\theta\right) d\theta \end{aligned}$$

where

$$L = \begin{cases} \lambda/(\mu - \lambda) & \text{for } \lambda < \mu \\ 0 & \text{for } \lambda \geq \mu \end{cases}$$

Other integral representations of mean queue length are found in [AW88].

Idle probability

The server idle probability is given by [Rid76]

$$P_0(t) = (1-\rho) + e^{-(\lambda+\mu)t} \left(\rho e^{\lambda t} - \sigma e^{\mu\sigma t} + \sqrt{\rho}\mu t R(t)\right) \tag{62}$$

where

$$R(t) = \int_0^1 I_1\left(2\sqrt{\rho}\mu t\sqrt{1-s}\right) \frac{s}{\sqrt{1-s}} (\rho e^{s\lambda t} - \sigma e^{s\sigma\mu t}) ds$$

Laplace transforms

The Laplace transforms of queue length distribution $P_n^*(s) = \mathcal{L}(p_n(t))$ and mean queue length $\mathcal{L}(E[Q(t)])$ for an initial queue length of zero are given by

$$P_{0n}^*(s) = \left(\frac{\lambda}{\mu}\xi\right)^n \frac{1}{\lambda + s - \lambda\xi} \tag{63}$$

and

$$\mathcal{L}(E[Q(t) \mid Q(0) = 0]) = \frac{\lambda(1-\xi)}{s^2} \tag{64}$$

where ξ is the root inside the unit circle of the quadratic equation

$$\lambda z^2 - (\lambda + s + \mu)z + \mu = 0$$

These equations are derived from equation (11.50) of Bailey [Bai64]. Kotiah gives these results for $\lambda = 1$ (equations (1) and (2) of [Kot78]).

A.2 The M/E$_k$/1 queue

Kolmogorov's system of forward DEs

$$\begin{aligned}
\frac{d}{dt}P_0(t) &= -\lambda P_0(t) + \mu P_1(t) && (65)\\
\frac{d}{dt}P_n(t) &= -(\lambda+\mu)P_n(t) + \mu P_{n+1}(t) + \lambda \sum_{m=1}^{n} c_m P_{n-m}(t) && (66)
\end{aligned}$$

where $P_n(t)$ is the probability that at time t the queue consists of elements with a total of n phases either waiting or being processed [Luc56].

Idle probability

The probability that the queue is zero at time t when the initial queue length is j is given by [Luc56]

$$P_0(t,j) = 2\sum_{m=1}^{\infty}(j+m)\rho^{-(j+m-1)/(k+1)}\frac{1}{t}I^k_{j+m}(t)\exp\left(-\frac{(1+\rho)t}{2\rho^{1/(k+1)}}\right) \qquad (67)$$

where

$$\begin{aligned}
I^k_n(z) &= \sum_{r=0}^{\infty}\frac{(z/2)^{n+(k+1)r}}{r!(rk+n)!}\\
z &= 2\rho^{1/(k+1)}\mu t\\
\rho &= \lambda/\mu
\end{aligned}$$

A.3 The M$_t$/M$_t$/1 queue

Kolmogorov's system of forward DEs

$$\begin{aligned}
\frac{dP_0(t)}{dt} &= -\lambda(t)P_0(t) + \mu(t)P_1(t) \\
\frac{dP_n(t)}{dt} &= \lambda(t)P_{n-1}(t) - \Big(\lambda(t) + \mu(t)\Big)\, P_n(t) + \mu(t)P_{n+1}(t) \qquad (68)
\end{aligned}$$

for $n > 1$.

Queue length mean

$$\frac{d}{dt}E[Q(t)] = \lambda(t) - \mu(t)\Big(1 - P_0(t)\Big) \qquad (69)$$

Queue length variance

$$\frac{d}{dt}\mathrm{var}[Q(t)] = \lambda(t) + \mu(t) - \mu(t)\Big(2E[Q(t)] + 1\Big)\, P_0(t) \qquad (70)$$

A.4 The M$_t$/M$_t$/s queue

Kolmogorov's system of forward DEs

$$\begin{aligned}
\frac{d}{dt}P_0(t) &= -\lambda(t)P_0(t) + \mu(t)P_1(t) \qquad (71) \\
\frac{d}{dt}P_n(t) &= \lambda(t)P_{n-1}(t) - \Big(\lambda(t) + n\mu(t)\Big)\, P_n(t) + (n+1)\mu(t)P_{n+1}(t) \qquad (72)
\end{aligned}$$

for $n = 1, 2, \ldots, s-1$

$$\frac{d}{dt}P_n(t) = \lambda(t)P_{n-1}(t) - \Big(\lambda(t) + s\mu(t)\Big)\, P_n(t) + s\mu(t)P_{n+1}(t) \qquad (73)$$

for $n = s, s+1, \ldots$

Queue length mean

$$\frac{d}{dt}E[Q(t)] = \lambda(t) - s\mu(t) + \mu(t)\sum_{n=0}^{s-1}(s-n)P_n(t) \quad (74)$$

Queue length variance

$$\frac{d}{dt}\mathrm{var}[Q(t)] = \lambda(t) + s\mu(t) - \mu(t)\sum_{n=0}^{s-1}\Big(2E[Q(t)] + 1 - 2n\Big)(s-n)P_0(t) \quad (75)$$

A.5 The M/M/1/K queue

Kolmogorov's system of forward DEs

$$\begin{aligned}
\frac{d}{dt}P_0(t) &= -\lambda P_0(t) + \mu P_1(t) \\
\frac{d}{dt}P_K(t) &= \lambda P_{K-1}(t) - \mu P_K(t) \\
\frac{d}{dt}P_n(t) &= \lambda P_{n-1}(t) - (\lambda+\mu)\,P_n(t) + \mu P_{n+1}(t) \quad (76)
\end{aligned}$$

for $1 \leq n < K$.

Queue length probabilities

The probability that the queue length is n at time t given that the initial queue length is j is given by [Tak62, Gif78, Mor58].

If $\lambda \neq \mu$ then

$$\begin{aligned}
p_{jn}(t) &= \frac{(1-\lambda/\mu)(\lambda/\mu)^n}{1-(\lambda/\mu)^{N+1}} \\
&\quad + \frac{2}{N+1}\sum_{m=1}^{N}\frac{(\lambda/\mu)^{(n-j)/2}\, e^{-\left(\lambda+\mu+2\sqrt{\lambda\mu}\,\cos(m\pi/(N+1))t\right)}}{1-2\sqrt{\lambda/\mu}\,\cos(m\pi/(N+1))+\lambda/\mu}
\end{aligned}$$

$$\cdot\left(\sin\frac{jm\pi}{N+1}-\sqrt{\lambda/\mu}\,\sin\frac{(j+1)m\pi}{N+1}\right)$$
$$\cdot\left(\sin\frac{nm\pi}{N+1}-\sqrt{\lambda/\mu}\,\sin\frac{(n+1)m\pi}{N+1}\right) \tag{77}$$

and if $\lambda=\mu$ then

$$\begin{aligned} p_{jn}(t) &= \frac{1}{N+1}+\frac{1}{N+1}\sum_{m=1}^{N}\frac{e^{-2\lambda(1+\cos(m\pi/(N+1)))\,t}}{1-\cos(m\pi/(N+1))} \\ &\cdot\left(\sin\frac{jm\pi}{N+1}-\sin\frac{(j+1)m\pi}{N+1}\right)\left(\sin\frac{nm\pi}{N+1}-\sin\frac{(n+1)m\pi}{N+1}\right) \end{aligned} \tag{78}$$

A.6 The M$_t$/M/1/K queue

Kolmogorov's system of forward DEs

$$\begin{aligned} \frac{d}{dt}P_0(t) &= -\lambda(t)P_0(t)+\mu P_1(t) \\ \frac{d}{dt}P_K(t) &= \lambda(t)P_{K-1}(t)-\mu P_K(t) \\ \frac{d}{dt}P_n(t) &= \lambda(t)P_{n-1}(t)-(\lambda(t)+\mu)\,P_n(t)+\mu P_{n+1}(t) \end{aligned} \tag{79}$$

for $1\le n<K$.

Appendix B

More results for the M/M/1/K queue

The results of section B.1 are required for section B.2 while the results of section B.2 are used in section 3.1.1.

B.1 Selection of trigonometric results

Result B.1.1

$$\sum_{n=0}^{N} \cos(n\theta + \phi) = \frac{\sin(\theta(N + 1/2) + \phi) - \sin(-\theta/2 + \phi)}{2\sin(\theta/2)}$$

Proof: The difference operator Δ of the finite difference calculus [Ham73] is defined by

$$\Delta f(n) \equiv f(n+h) - f(n)$$

Applying this operator to $\sin(an + b)$ yields

$$\Delta \sin(an + b) = 2\sin(ah/2)\ \cos(an + ah/2 + b) \tag{80}$$

The difference and summation calculi are related by

$$f(N+1) - f(0) \quad = \quad \sum_{n=0}^{N} \Delta f(n) \tag{81}$$

Summing equation (80) over n with $h = 1$ and using equation (81) yields

$$\sin(aN + a + b) - \sin b \quad = \quad 2\sin(a/2) \sum_{n=0}^{N} \cos(an + a/2 + b)$$

Changing variables $\theta = a$ and $\phi = a/2 + b$ and arranging terms gives the result. □

Result B.1.2 *If $k \in \mathbb{Z}$ then*

$$\sum_{n=0}^{N-1} \cos\frac{kn\pi}{N} \quad = \quad \begin{cases} N, & \text{if } |k| = 0, 2N, 4N, \ldots, \\ 1, & \text{if } |k| \neq 0, 2N, 4N, \ldots \text{ and } k \text{ is odd,} \\ 0, & \text{if } |k| \neq 0, 2N, 4N, \ldots \text{ and } k \text{ is even.} \end{cases}$$

Proof: If $|k| = 0, 2N, 4N, \ldots$ then the sum is N since $\cos(kn\pi/N) = 1$.

Using result B.1.1 yields

$$\sum_{n=0}^{N-1} \cos\frac{kn\pi}{N} \quad = \quad \frac{\sin(k\pi - \frac{k\pi}{2N}) + \sin\frac{k\pi}{2N}}{2\sin\frac{k\pi}{2N}} \tag{82}$$

on condition that $|k| \neq 0, 2N, 4N, \ldots$

If k is even then

$$\sin\left(k\pi - \tfrac{k\pi}{2N}\right) \quad = \quad -\sin\tfrac{k\pi}{2N}$$

and therefore the left hand side of equation (82) is 0.

If k is odd then

$$\sin\left(k\pi - \tfrac{k\pi}{2N}\right) \quad = \quad \sin\tfrac{k\pi}{2N}$$

and therefore the left hand side of equation (82) is 1. □

Result B.1.3 *If $k \in \mathbb{Z}$ and $m \in \mathbb{Z}$ then*

$$\sum_{n=0}^{N-1} \sin \frac{kn\pi}{N} \sin \frac{mn\pi}{N} = \begin{cases} \frac{N}{2}, & \text{if } (|k-m| = 0, 2N, 4N, \ldots \text{ and } |k+m| \neq 0, 2N, 4N, \ldots), \\ -\frac{N}{2}, & \text{if } (|k-m| \neq 0, 2N, 4N, \ldots \text{ and } |k+m| = 0, 2N, 4N, \ldots), \\ 0, & \text{otherwise.} \end{cases}$$

Proof: This result is obtained by using the product formula

$$\sin u \, \sin v = \frac{1}{2} \Big(\cos(u-v) - \cos(u+v) \Big)$$

together with result B.1.2. □

Result B.1.4 *If $k \in \mathbb{Z}$ and $m \in \mathbb{Z}$ then*

$$\sum_{n=0}^{N-1} \cos \frac{kn\pi}{N} \cos \frac{mn\pi}{N} = \begin{cases} \frac{N}{2}, & \text{if } (|k-m| = 0, 2N, 4N, \ldots \text{ and } |k+m| \neq 0, 2N, 4N, \ldots) \\ & \text{or } (|k-m| \neq 0, 2N, 4N, \ldots \text{ and } |k+m| = 0, 2N, 4N, \ldots), \\ 1, & \text{if } |k-m| \neq 0, 2N, 4N, \ldots \text{ and } |k+m| \neq 0, 2N, 4N, \ldots \\ & \text{and } k+m \text{ is odd,} \\ 0, & \text{if } (|k-m| = 0, 2N, 4N, \ldots \text{ and } |k+m| = 0, 2N, 4N, \ldots) \\ & \text{or } (|k-m| \neq 0, 2N, 4N, \ldots \text{ and } |k+m| \neq 0, 2N, 4N, \ldots \\ & \text{and } k+m \text{ is even).} \end{cases}$$

Proof: This result is obtained by using the product formula

$$\cos u \, \cos v = \frac{1}{2} \Big(\cos(u-v) + \cos(u+v) \Big)$$

together with result B.1.2. □

B.2 Expansions for the transient component of queue length

Let

$$\begin{aligned} M_{i,n}(x,m) &= \Big(\sin im\theta - x \sin(i+1)m\theta \Big) \cdot \Big(\sin nm\theta - x \sin(n+1)m\theta \Big) \\ &\quad \cdot \prod_{\ell=1, \ell \neq m}^{K} \Big(1 - 2x \cos \ell\theta + x^2 \Big) \end{aligned} \tag{83}$$

where

$$\theta = \pi/(K+1)$$

Then

$$\sum_{m=1}^{K} M_{i,n}(x,m) = c_0 + c_1 x + c_2 x^2 + \ldots + c_{2K} x^{2K}$$

where $0 \le i \le K$ and $0 \le n \le K$.

Result B.2.1 *If c_0 is the coefficient of the zero-th power of x defined by equation (83) then*

$$c_0 = \begin{cases} 1, & \textit{if } i = n \textit{ and } i \neq 0, \\ 0, & \textit{otherwise.} \end{cases} \tag{84}$$

Proof: Equation (83) yields

$$c_0 = \frac{2}{K+1} \sum_{m=1}^{K} \sin \frac{im\pi}{K+1} \sin \frac{nm\pi}{K+1}$$

Using result B.1.3 and noting that $0 \le |i+n| \le 2K$ and $0 \le |i-n| \le 2K$ yields

$$c_0 = \frac{2}{K+1} \begin{cases} \frac{K+1}{2}, & \text{if } |i-n| = 0 \text{ and } |i+n| \neq 0, \\ -\frac{K+1}{2}, & \text{if } |i-n| \neq 0 \text{ and } |i+n| = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Now $|i-n| = 0$ implies $i = n$. Since $0 \le i \le K$ and $0 \le n \le K$, $|i+n| = 0$ implies $i = 0$ and $n = 0$. Furthermore the condition ($|i-n| \neq 0$ and $|i+n| = 0$) is false. □

Result B.2.2 *If c_1 is the coefficient of x defined by equation (83) then*

$$c_1 = \begin{cases} -1, & \textit{if } i + n = 1, \\ 0, & \textit{if } i + n \neq 1. \end{cases} \tag{85}$$

Proof: Equation (83) yields

$$\begin{aligned}
c_1 &= -\frac{2}{K+1}\sum_{m=1}^{K}\Bigg(\sin im\theta\ \sin(n+1)m\theta + \sin(i+1)m\theta\ \sin nm\theta \\
&\quad +2\sin im\theta\ \sin nm\theta \sum_{\ell=1,\ell\neq m}^{K}\cos\ell\theta\Bigg) \\
&= -\frac{2}{K+1}\Bigg(\sum_{m=1}^{K}\sin im\theta\ \sin(n+1)m\theta + \sum_{m=1}^{K}\sin(i+1)m\theta\ \sin nm\theta \\
&\quad +2\left(\sum_{\ell=1}^{K}\cos\ell\theta\right)\sum_{m=1}^{K}\sin im\theta\ \sin nm\theta - \sum_{m=1}^{K}\cos(i-n)m\theta\ \cos m\theta \\
&\quad +\sum_{m=1}^{K}\cos(i+n)m\theta\ \cos m\theta\Bigg) \qquad (86)
\end{aligned}$$

where $\theta = \pi/(K+1)$.

Using result B.1.2 with $k = 1$ yields

$$\sum_{\ell=1}^{K}\cos\ell\theta = 0$$

Using results B.1.3 and B.1.4 yields the following 4 equations

$$\sum_{m=1}^{K}\sin im\theta\ \sin(n+1)m\theta = \begin{cases} \frac{K+1}{2}, & \text{if } i = n+1, \\ 0, & \text{if } i \neq n+1. \end{cases}$$

$$\sum_{m=1}^{K}\sin(i+1)m\theta\ \sin nm\theta = \begin{cases} \frac{K+1}{2}, & \text{if } n = i+1, \\ 0, & \text{if } n \neq i+1. \end{cases}$$

$$\sum_{m=1}^{K}\cos(i-n)m\theta\ \cos m\theta = \begin{cases} \frac{K-1}{2}, & \text{if } i = n+1 \text{ or } i = n-1, \\ -1, & \text{if } i \neq n+1 \text{ and } i \neq n-1 \text{ and } i+n \text{ is odd}, \\ 0, & \text{if } i \neq n+1 \text{ and } i \neq n-1 \text{ and } i+n \text{ is even}. \end{cases}$$

$$\sum_{m=1}^{K}\cos(i+n)m\theta\ \cos m\theta = \begin{cases} \frac{K-1}{2}, & \text{if } i+n = 1, \\ -1, & \text{if } i+n \neq 1 \text{ and } i+n \text{ is odd}, \\ 0, & \text{if } i+n \neq 1 \text{ and } i+n \text{ is even}. \end{cases}$$

Substituting these equations into equation (86) yields the result. □

B.3 Maple program for the evaluation of queue length probabilities

This Maple program is described in section 3.1.1.

```
> # Compute the transient component of the queue length probability.
> # Valid for lambda ~ 0 and i > n.
> restart;
```

```
> POWER_TERMS := 20;
```

$$POWER\_TERMS := 20$$

```
> M_in:=(i,n,x,m,K)->(sin(i*m*Pi/(K+1)) - x * sin((i+1)*m*Pi/(K+1)))
>     *(sin(n*m*Pi/(K+1)) - x * sin((n+1)*m*Pi/(K+1)));
```

$$M\_in := (i, n, x, m, K) \to \left(\sin\left(\frac{i\,m\,\pi}{K+1}\right) - x\sin\left(\frac{(i+1)\,m\,\pi}{K+1}\right)\right)\left(\sin\left(\frac{n\,m\,\pi}{K+1}\right) - x\sin\left(\frac{(n+1)\,m\,\pi}{K+1}\right)\right)$$

```
> # Expand exponential factor in terms of a power series in x.
> # Valid for small mu*t*(1-2*sqrt(lambda/mu)*cos( ) + lambda/mu).
> alpha:=(x,nu,m,K)->subs(y=x,convert(series(exp(-nu*(1-2*'y'*cos(m*Pi/(K+1))+'y'^2)),
>     'y',POWER_TERMS),polynom));
```

$$\alpha := (x, \nu, m, K) \to \mathrm{subs}\Big(y = x, \mathrm{convert}\Big(\mathrm{series}\left(\mathrm{e}^{\left(-\nu\left(1-2\,'y'\cos\left(\frac{m\,\pi}{K+1}\right)+'y'^2\right)\right)}, 'y', POWER\_TERMS\right), polynom\Big)\Big)$$

```
> # Sum K terms.  The expansions (in x) are not computed at this stage.
> transient_sum:=(i,n,x,nu,K)->subs(y=x,sum(''expand('y'^(n-i)*alpha('y',nu,m,K)
>     *M_in(i,n,'y',m,K),'y')/(1-2*'y'*cos(m*Pi/(K+1))+'y'^2 )'','m'=1..K));
```

$$transient\_sum := (\, i, n, x, \nu, K \,) \to \mathrm{subs}\Bigg(y = x,$$

$$\sum_{'m'=1}^{K} {}'' \frac{\mathrm{expand}\left('y'^{(\,n-i\,)}\, \alpha(\,'y', \nu, m, K\,)\, \mathrm{M\_in}(\,i, n, 'y', m, K\,), 'y' \right)''}{1 - 2\,'y' \cos\left(\dfrac{m\,\pi}{K+1}\right) + 'y'^2} \Bigg)$$

```
> # Expand terms in transient_sum and convert the sum to a rational
> # polynomial in x.
> transient:=(i,n,x,nu,K)->eval(subs(y=x,2/(K+1)
>     *normal(eval(transient_sum(i,n,'y',nu,K)),expanded)));
```

$$transient := (\, i, n, x, \nu, K \,) \to \mathrm{eval}\Bigg($$

$$\mathrm{subs}\left(y = x, 2\, \frac{\mathrm{normal}(\,\mathrm{eval}(\,\mathrm{transient\_sum}(\,i, n, 'y', \nu, K\,)\,), \mathit{expanded}\,)}{K+1} \right) \Bigg)$$

```
> # Evaluate the rational polynomial using floating point arithmetic.
> p_in:=(i,n,lambda,mu,t,K)->evalf(transient(i,n,sqrt(lambda/mu),mu*t,K),10);
```

$$p\_in := (\, i, n, \lambda, \mu, t, K \,) \to \mathrm{evalf}\left(\mathrm{transient}\left(i, n, \mathrm{sqrt}\left(\frac{\lambda}{\mu}\right), \mu\, t, K \right), 10 \right)$$

```
> # Compute the transient component of the M/M/1/K probability P_in(t)
> # with lambda = 0.2, mu = 1, t = 0.5, K = 8, i = 8 and n = 0,
> # Note that the steady-state component P_n is 0.8000004096 .
> p_in(8,0,0.2,1.0,0.5,8);
```

$$-.8000001470$$

Appendix C

Derivations for Moore's nonstationary model

This appendix derives both the distribution for the number of arrivals during a service $P(\mathcal{A}(t) = j)$ and the expected length of time from an arbitrary instant t to the first arrival $E[\mathcal{Z}(t)]$ for Moore's nonstationary arrival model[1].

Moore's *nonstationary arrival model* for the $M_t^X/E_Y/1/K$ queue approximates a non-homogeneous arrival process by dividing the time axis into $N+1$ intervals $I_i, 1 \leq i \leq N+1$ and the arrival process is considered to be homogeneous in each interval.

Let the intervals be defined by

$$I_i = \begin{cases} [t_0, t_1], & \text{if } i = 1, \\ (t_{i-1}, t_i], & \text{if } 1 < i \leq N+1, \end{cases}$$

where $0 = t_0 < t_1 < t_2 < \cdots < t_N \leq t_{N+1} = \infty$. The arrival process in each interval is modelled by the compound Poisson distribution (equation (29)) with inter-arrival parameter λ_i and group size probability mass function $g_{ij} = P(X_i = j)$. The service

[1]The original derivation is given by Moore in [Moo75]. Several errors found in the original paper have been corrected.

process is homogeneous and identical to the service process for the *stationary arrival model.*

C.1 The number of arrivals during a service

Let $S(t)$ be the time at which a customer arriving at time t completes its service. Let $A_{I_i}(t,s)$ be the number of customers arriving in the interval $(t,s]$ for t and s in interval I_i. Let $A(t,s)$ be the number of customers arriving in the interval $(t,s]$ for $0 \leq t$ and $s \leq \infty$. Finally let $\mathcal{A}(t)$ be the number of customers arriving in the interval $(t, S(t)]$ for $t \geq 0$.

Now $P(A_{I_i}(a,b) = m)$ is given by the compound Poisson distribution

$$P(A_{I_i}(a,b) = m) \quad = \quad f_{A|T}(m \mid b-a) \; = \; \sum_{r=0}^{m} e^{-\lambda_i(b-a)} \frac{(\lambda_i(b-a))^r}{r!} g_{im}^{(r*)}$$

where $g_{im}^{(r*)}$ is the mth component of the r-fold convolution of distribution g_i with itself.

The distribution of $A(t,s)$ is calculated from the distribution of $A_{I_i}(t,s)$ as follows. Let the index $\hat{\imath}$ be defined such that for a given t, $t \in I_{\hat{\imath}}$. Then given an instant t and an interval I_i such that $0 \leq t \leq t_i$, the distribution $P(A(t,t_i) = m)$ is determined by calculating convolutions of the distributions of $A_{I_i}(\cdot,\cdot)$ over the appropriate intervals, that is

$$P(A(t,t_i) = m) \quad = \quad \begin{cases} 0, & \text{if } i < \hat{\imath}, \\ P(A_{I_i}(t,t_{\hat{\imath}}) = m), & \text{if } i = \hat{\imath}, \\ P(A_{I_i}(t,t_{\hat{\imath}}) + \sum_{k=\hat{\imath}+1}^{i} A_{I_i}(t_{k-1},t_k) = m), & \text{if } i > \hat{\imath}. \end{cases} \tag{87}$$

The distribution that we require, namely $P(\mathcal{A} = j)$ for $0 \leq j \leq K-1$ is determined from the conditional probabilities $P(\mathcal{A}(t) = j, S(t) \in I_i)$ using the equation

$$P(\mathcal{A}(t) = j) \quad = \quad \sum_{i=\hat{\imath}}^{N+1} P(\mathcal{A}(t) = j, S(t) \in I_i) \tag{88}$$

Let the interval $(t, S(t)]$ be partitioned into two intervals $(t, t_{i-1}]$ and $(t_{i-1}, S(t)]$ where $S(t) \in I_i$. Since the arrival and service distributions processes are independent we have

$$P(\mathcal{A}(t) = j, S(t) \in I_i) \quad = \quad 0 \tag{89}$$

for $i < \hat{\imath}$,

$$P(\mathcal{A}(t) = j, S(t) \in I_i) \quad = \quad \int_t^{t_i} P(A(t,s) = j)\, dP(S(t) \leq s) \tag{90}$$

for $i = \hat{\imath}$ and

$$\begin{aligned} P(\mathcal{A}(t) = j, S(t) \in I_i) &= \\ \sum_{n=0}^{j} P(A(t, t_{i-1}) = j - n) &\int_{t_{i-1}}^{t_i} P(A(t_{i-1}, s) = n)\, dP(S(t) \leq s) \end{aligned} \tag{91}$$

for $i > \hat{\imath}$.

Define

$$R_n(t, y, t_i) = \int_y^{t_i} P(A_{I_i}(y, s) = n)\, dP(S(t) \leq s) \tag{92}$$

for $t \leq y, y \in \{I_i \cup t_{i-1}\}$ and $\hat{\imath} < i$. (That is, t is in an interval prior to the particular interval containing y and t_i.)

The service time of a virtual customer entering service at time t, namely $X(t) = S(t) - t$ is independent of t since the service process is stationary. Therefore

$$P(S(t) \leq s) = P(X(t) \leq s - t) = H_T(s - t)$$

where $H_T(t)$ is the cumulative service time distribution. Substituting equation (29) into equation (92) yields

$$\begin{aligned} R_n(t, y, t_i) &= \int_y^{t_i} \sum_{r=0}^{n} e^{-\lambda_i (s-y)} \frac{(\lambda_i (s-y))^r}{r!} g_{in}^{(r*)}\, dH_T(s-t) \\ &= \sum_{r=0}^{n} g_{in}^{(r*)} \int_{y-t}^{t_i - t} p_r(\lambda_i (s + t - y))\, dH_T(s) \end{aligned} \tag{93}$$

where $p_r(u)$ is the Poisson probability mass function given by

$$p_r(u) \quad = \quad e^{-u}\frac{u^r}{r!}$$

for $r \geq 0$. Substituting the weighted-sum Erlang distribution function (equation (30)) into equation (93) yields

$$\begin{aligned} R_n(t,y,t_i) \quad &= \quad \sum_{r=0}^{n} g_{in}^{(r*)} p_r(\lambda_i(t-y)) \sum_{k=1}^{\infty} \frac{c_k \mu^k}{(k-1)!} \\ &\quad \cdot \int_{y-t}^{t_i-t} e^{-(\lambda_i+\mu)s}(s+t-y)^r s^{k-1} ds \end{aligned} \tag{94}$$

Using the binomial theorem, we can write the integral portion of equation (94) as

$$\begin{aligned} \int_{y-t}^{t_i-t} e^{-(\lambda_i+\mu)s}(s+t-y)^r s^{k-1} ds \ &= \int_{y-t}^{t_i-t} e^{-(\lambda_i+\mu)s} s^{r+k-1} ds \\ + \sum_{m=0}^{r-1} \binom{r}{m} (t-y)^{r-m} &\int_{y-t}^{t_i-t} e^{-(\lambda_i+\mu)s} s^{m+k-1} ds \end{aligned} \tag{95}$$

Using the identity

$$\int_0^T e^{-au} u^n du \quad = \quad \frac{n!}{a^{n+1}} - n!\, e^{-aT} \sum_{i=0}^{n} \frac{T^i}{i!} \frac{1}{a^{n-i+1}}$$

(which is obtained by integrating by parts), we have

$$\begin{aligned} &\int_{y-t}^{t_i-t} e^{-(\lambda_i+\mu)s} s^{m+k-1} ds = \\ &\frac{(m+k-1)!}{(\lambda_i+\mu)^{m+k}} \sum_{j=0}^{m+k+1} \Big(p_j\left((\lambda_i+\mu)(y-t)\right) - p_j\left((\lambda_i+\mu)(t_i-t)\right) \Big) \end{aligned} \tag{96}$$

Combining the equations (93) through (96) yields

$$\begin{aligned} R_n(t,y,t_i) \quad &= \quad e^{-\lambda_i(t-y)} \sum_{r=0}^{n} g_{in}^{(r*)} \lambda_i^r \sum_{k=1}^{\infty} c_k \mu^k \\ &\cdot \left(\binom{r+k-1}{r} \frac{d_{r+k-1}(\lambda_i+\mu,t,y,t_i)}{(\lambda_i+\mu)^{r+k}} \right. \\ &\left. + \sum_{m=0}^{r-1} \binom{m+k-1}{m} \frac{(t-y)^{r-m} d_{m+k-1}(\lambda_i+\mu,t,y,t_i)}{(r-m)!\,(\lambda_i+\mu)^{m+k}} \right) \end{aligned} \tag{97}$$

where

$$d_n(a,t,y,t_i) \quad = \quad \sum_{m=0}^{n} \Big(p_m\left(a(y-t)\right) - p_m\left(a(t_i-t)\right) \Big)$$

Equations (89), (90) and (91) are rewritten as

$$P(\mathcal{A}(t)=j, S(t)\in I_i) \; = \\ \begin{cases} 0, & \text{if } i<\hat{\imath}, \\ R_j(t,t,t_{\hat{\imath}}), & \text{if } i=\hat{\imath}, \\ \sum_{n=0}^{j} P(A(t,t_{i-1})=j-n)R_n(t,t_{i-1},t_i), & \text{if } i>\hat{\imath}. \end{cases} \tag{98}$$

Combining equations (87), (88), (97) and (98) yields an expression for evaluating the probabilities $P(\mathcal{A}(t)=j)$.

Moore suggests that in practice $P(\mathcal{A}(t)=j, S(t)\in I_i)$ need not be calculated for every interval I_i. Because

$$P(\mathcal{A}(t)=j, S(t)\in I_i) \quad \le \quad P(S(t)\in I_i)$$

where

$$P(S(t)\in I_i) \quad = \quad \begin{cases} 0, & \text{if } i<\hat{\imath}, \\ F(t_i-t), & \text{if } i=\hat{\imath}, \\ F(t_i-t)-F(t_{i-1}-t), & \text{if } i>\hat{\imath}, \end{cases} \tag{99}$$

and because the evaluation of $P(S(t) \in I_i)$ requires less computation than equation (98), equation (98) is calculated only if $P(S(t)\in I_i)$ has significant magnitude. In particular, Moore suggests that equation (98) is calculated only if $P(S(t)\in I_i)\ge 0.001$.

C.2 The time until an arrival

We wish to calculate $E[\mathcal{Z}(t)]$, the expected length of time from an arbitrary instant t to the first arrival.

Define $Y(t)$ to be the time of the first arrival starting from an arbitrary instant t, therefore $\mathcal{Z}(t) = Y(t) - t$ with mean

$$E[\mathcal{Z}(t)] \quad = \quad E[Y(t)] - t \tag{100}$$

We have

$$\begin{aligned} P(Y(t) \leq s) &= \sum_{i=1}^{N+1} P(Y(t) \leq s \mid Y(t) \in I_i)\, P(Y(t) \in I_i) \\ &= \sum_{i=1}^{N+1} P(Y(t) \leq s, Y(t) \in I_i) \end{aligned}$$

therefore

$$E[Y(t)] \quad = \quad \sum_{i=1}^{N+1} \int_{Y(t) \in I_i} s\, dP(Y(t) \leq s)$$

Because $Y(t) \geq t$ and $t \in I_{\hat{\imath}}$, we have

$$E[Y(t)] \quad = \quad \int_t^{t_{\hat{\imath}}} s\, dP(Y(t) \leq s) \tag{101}$$

for $\hat{\imath} = N + 1$ and

$$E[Y(t)] \quad = \quad \int_t^{t_{\hat{\imath}}} s\, dP(Y(t) \leq s) + \sum_{i=\hat{\imath}+1}^{N+1} \int_{t_{i-1}}^{t_i} s\, dP(Y(t) \leq s) \tag{102}$$

for $1 \leq \hat{\imath} \leq N$. We require expressions for the integral portions of equations (101) and (102).

Define $B_i(t)$ to be the event that no arrival occurs in the interval $(t, t_i]$ and let

$$P(B_i(t)) \quad = \quad 1 \tag{103}$$

for $t_i \leq t$. Given that $t \leq s, s \in I_i$, we have

$$\begin{aligned} &P(\text{first customer after } t \text{ arrives in } \{\tau : s \leq \tau \leq s + ds\}) \; = \\ &\qquad P(\text{no customers arrive in } \{\tau : t \leq \tau \leq t_{i-1}\}) \\ &\quad \times \;\; P(\text{first customer after } t_{i-1} \text{ arrives in } \{\tau : s \leq \tau \leq s + ds\}) \end{aligned}$$

(because the arrival process is Markovian) and is written as

$$dP(Y(t) \leq s) \quad = \quad P(B_{i-1}(t))\, dF_i(s - t_{i-1}) \tag{104}$$

Now, $F_i(t) = 1 - f_{A|T}(0|t)$ where $f_{A|T}(j|t)$ is the compound Poisson distribution and consequently $F_i(t)$ is the negative exponential distribution given by

$$F_i(t) = 1 - e^{-\lambda_i t}$$

for $t \geq 0$. Therefore equation (104) is written as

$$dP(Y(t) \leq s) \quad = \quad P(B_{i-1}(t))\, \lambda_i e^{-\lambda_i(s-t_{i-1})} ds \tag{105}$$

for $t \leq s, s \in I_i$.

The probability that no arrival occurs in the interval $[t, t_i]$ is given by the product of probabilities that no arrival occurs within each interval from $I_{\hat{\imath}}$ to I_i, that is,

$$P(B_i(t)) \quad = \quad \begin{cases} 1, & \text{if } t_i < t, \\ \bar{F}_{\hat{\imath}}(t_{\hat{\imath}} - t), & \text{if } i = \hat{\imath}, \\ \bar{F}_{\hat{\imath}}(t_{\hat{\imath}} - t) \prod_{j=\hat{\imath}+1}^{i} \bar{F}_j(t_j - t_{j-1}), & \text{if } i > \hat{\imath}, \end{cases} \tag{106}$$

where $\bar{F}_i(x) = 1 - F_i(x)$. Integrating equation (105) from t to t_i and using equation (106) yields

$$\begin{aligned} \int_t^{t_{\hat{\imath}}} s\, dP(Y(t) \leq s) \quad &= \quad \int_t^{t_{\hat{\imath}}} s\, dF_{\hat{\imath}}(s-t) \\ &= \quad t + (1/\lambda_{\hat{\imath}}) - (1/\lambda_{\hat{\imath}} + t_{\hat{\imath}}) e^{-\lambda_{\hat{\imath}}(t_{\hat{\imath}} - t)} \end{aligned} \tag{107}$$

Similarly, integrating equation (105) from t_{i-1} to t_i, yields

$$\begin{aligned} &\int_{t_{i-1}}^{t_i} s\, dP(Y(t) \leq s) \\ &= \quad P(B_{i-1}(t)) \left(t_{i-1} + (1/\lambda_i) - (1/\lambda_i + t_i) e^{-\lambda_i(t_i - t_{i-1})}\right) \end{aligned} \tag{108}$$

Substituting equations (107) and (108) into equations (101) and (102) and using equation (100) yields

$$E[\mathcal{Z}(t)] \quad = \quad 1/\lambda_{\hat{\imath}} - (1/\lambda_{\hat{\imath}} + t_{\hat{\imath}}) e^{-\lambda_{\hat{\imath}}(t_{\hat{\imath}} - t)} \tag{109}$$

for $\hat{\imath} = N + 1$ and

$$\begin{aligned} E[\mathcal{Z}(t)] &= 1/\lambda_{\hat{\imath}} - (1/\lambda_{\hat{\imath}} + t_{\hat{\imath}})e^{-\lambda_{\hat{\imath}}(t_{\hat{\imath}}-t)} + \left(\sum_{i=\hat{\imath}+1}^{N+1} P(B_{i-1}(t)) \right. \\ &\quad \left. \cdot \left(t_{i-1} + (1/\lambda_i) - (1/\lambda_i + t_i)e^{-\lambda_i(t_i - t_{i-1})} \right) \right) \end{aligned} \tag{110}$$

for $1 \leq \hat{\imath} \leq N$.

Bibliography

[AAK93] M. O. Abou-el-ata, R. O. Al-seedy, and K. A. M. Kotb. Transient solution of the state-dependent queue: M/M/1/N with balking and reflecting barrier. *Microelectronics & Reliability*, 33(5):681–688, 1993.

[AC89] J. Michael Andrewartha and Pierce E. Cantrell. Transient approximations for Jackson networks. In *IEEE Global Telecommunications Conference and Exhibition (Dallas, Texas)*, volume 2, pages 954–958, November 1989.

[Ack82] Martin H. Ackroyd. M/M/1 transition state occupancy probabilities via the discrete fourier transform. *IEEE Transactions on Communications*, COM 30(3):557–559, 1982.

[Agn75] Carson E. Agnew. Dynamic modelling and control of congestion-prone systems. *Operations Research*, 24(3):400–419, 1975.

[AW88] Joseph Abate and Ward Whitt. Simple spectral representations for the M/M/1 queue. *Queueing Systems*, 3:321,346, 1988.

[AW89] Joseph Abate and Ward Whitt. Calculating time-dependent performance measures of the M/M/1 queue. *Queueing Systems*, 37(10):1102–1104, October 1989.

[AW92] Joseph Abate and Ward Whitt. The fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10:5–88, 1992.

[Bai54] Norman T. J. Bailey. A continuous time treatment of a simple queue using generating functions. *Journal of the Royal Statistical Society (Series B)*, 16:288–291, 1954.

[Bai64] Norman T. J. Bailey. *The elements of Stochastic Processes with applications to the natural sciences.* John Wiley & Sons, Inc., 1964.

[BT93] R. Boucherie and P. G. Taylor. Transient product form distributions in queueing networks. *Disc. Event Dyn. Sys.*, 3:375–396, 1993.

[Can86] Pierce E. Cantrell. Computation of the transient M/M/1 queue cdf, pdf, and mean generalized Q-Functions. *IEEE Transactions on Communications*, COM-34(8):814–817, August 1986.

[Ç67] E. Çinlar. Queues with semi-markovian arrivals. *Journal of Applied Probability*, 4:365–379, 1967.

[Cha56] D. G. Champernowne. An elementary method of solution of the queueing problem with a single server and constant parameters. *Journal of the Royal Statistical Society (Series B)*, 18(1):125–128, 1956.

[Cha77] S. S. L. Chang. Simulation of transient and time varying conditions in queueing networks. In *7th Annual conference on modeling and simulation (University of Pittsburg)*, pages 1075–1078, 1977.

[Cla53] A. B. Clarke. The time-dependent waiting line problem. Technical report, University of Michigan Engineering Research Institute, 1953.

[Cla56] A. B. Clarke. A waiting line process of Markov type. *The Annals of Mathematical Statistics*, 27:452–459, 1956.

[Cla81] Gordon M. Clark. Use of Polya distributions in approximate solutions to nonstationary M/M/s queues. *Communication of the ACM*, 24(4):206–217, April 1981.

[Coo81] Robert B. Cooper. *Introduction to Queueing Theory.* Elsevier North Holland, Inc. (New York), 1981.

[CS61] D. R. Cox and Walter L. Smith. *Queues.* Spottiswoode, Ballantyne & Co. Ltd., 1961.

[DJ62] A. R. DiDonata and M. P. Jarnagin. A method for computing the circular coverage function. *Mathematics of Computation*, 16:347–355, July 1962.

[Fil88] Janusz Filipiak. *Modelling and Control of Dynamic Flows in Communication Networks.* Springer-Verlag, 1988. ISBN 3-540-18292-6.

[Fol82] R.D. Foley. The non-homogeneous $m/g/\infty$ queue. *Opsearch*, 19:40–48, 1982.

[Gav54] D. P. Gaver. The influence of servicing times in queueing processes. *Operations Research*, 2:139–149, 1954.

[Gav66] D. P. Gaver. Observing stochastic processes, and approximate transform inversion. *Operations Research*, 14:444–459, 1966.

[GH74] Donald Gross and Carl M. Harris. *Fundamentals of Queueing Theory.* John Wiley & Sons, Inc., 1974.

[Gif78] Walter C. Giffin. *Queueing, Basic Theory and Applications.* Grid Inc. Columbus, Ohio, 1978.

[GW58] Herbert P. Galliher and R. Clyde Wheeler. Nonstationary queueing probabilities for landing congestion of aircraft. *Operations Research*, 6:607–623, 1958.

[Ham73] R. W. Hamming. *Numerical Methods for Scientists and Engineers.* Mc Graw-Hill, second edition, 1973.

[Har81] P. G. Harrison. Transient behaviour of queueing networks. *Journal of Applied Probability*, 18:482–490, 1981.

[Hec93] André Heck. *Introduction to Maple*. Springer-Verlag, 1993.

[HL81] P. G. Harrison and A. J. Lemoine. A note on networks of infinite-server queues. *Journal of Applied Probability*, 18:561–567, 1981.

[JCJ80] Stephen K. Jones, R. K. Cavin, and D. A. Johnston. An efficient computational procedure for the evaluation of the $M/M/1$ transient state occupancy probabilities. *IEEE Transactions on Communications*, COM-28(12):2019–2020, December 1980.

[JK69] Norman L. Johnson and Samuel Kotz. *Distributions in statistics: Discrete distributions*. Houghton Mifflin Company (Boston, MA), 1969.

[Kan92] Krishna Kant. *Introduction to Computer System Performance Evaluation*. Mc Graw-Hill, 1992.

[Ken51] David G. Kendall. Some problems in the theory of queues. *Journal of the Royal Statistical Society (Series B)*, 13:151–185, 1951.

[Ken53] David G. Kendall. Stochastic processes occuring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 24:338–354, 1953.

[Kin69] J. F. C. Kingman. Markov population processes. *Journal of Applied Probability*, 6:1–18, 1969.

[Kle75] Leonard Kleinrock. *Queueing Systems, Volume I: Theory*. Wiley-Interscience, 1975.

[Koo72] Bernard O. Koopman. Air-terminal queues under time-dependent conditions. *Operations Research*, 20:1089–1114, 1972.

[Kot78] T. C. T. Kotiah. Approximate transient analysis of some queueing systems. *Operations Research*, 26(2):333–346, March-April 1978.

[Lac93] Martin-Daniel Lacasse. Fudgit, a multi-purpose data-processing and fitting program (users manual). Technical report, Center for the Physics of Materials and Department of Physics, McGill University, 1993.

[LB66] Eric L. Leese and D. W. Boyd. Numerical methods of determining the transient behaviour of queues. *Journal of the Operational Research Society*, 4:1–13, 1966.

[LCW93] David M. Lucantoni, Gagan L. Choudhury, and Ward Whitt. Computing transient distributions in general single-server queues. In *IEEE Global Telecommunications Conference (Houston, Texas)*, pages 1045–1050, 1993.

[LGB83] F Le Gall and J Bernussou. An analytical formulation for grade of service determination in telephone networks. *IEEE Transactions on Communications*, 31(3):420–424, March 1983.

[LR54] W. Ledermann and G. E. H. Reuter. Spectral theory for the differential equations of simple birth and death processes. *Royal Society of London. Philosophical transactions A*, 246:321–369, 1954.

[Luc56] George Luchak. The solution of the single-channel queueing equations characterized by a time-dependent Poisson-distributed arrival rate and a general class of holding times. *Operations Research*, 4:711–732, 1956.

[Mar60] J. I. Marcum. A statistical theory of target detection by pulsed radar. *IRE Transactions on Information Theory*, IT-6:259–267, April 1960.

[Moo75] S. C. Moore. Approximating the behavior of nonstationary single server queues. *Operations Research*, 23(5):1011–1023, September-October 1975.

[Mor55] P. M. Morse. Stochastic properties of waiting lines. *Operations Research*, 3:255–261, 1955.

[Mor58] P. M. Morse. *Queues, Inventories and Maintenance*. John Wiley & Sons, Inc., 1958.

[Par87] P. R. Parthasarathy. A transient solution to a M/M1 queue: A simple approach. *Advances in Applied Probability*, 19:997–998, 1987.

[PFTV88] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C.* ISBN 0-521-35465-X. Press Syndicate of the University of Cambridge, 1988.

[Rid76] K. L. Rider. A simple approximation to the average queue size in the time-dependent queue. *Journal of the ACM*, 23(2):361–367, 1976.

[RO79] Michael H. Rothkopf and Shmuel S. Oren. A closure approximation for the nonstationary M/M/s queue. *Management Science*, pages 522–534, June 1979.

[Rob94] Thomas G. Robertazzi. *Computer Networks and Systems: Queueing Theory and Performance Evaluation.* Springer-Verlag, second edition, 1994.

[Saa61] Tomas L. Saaty. *Elements of Queueing Theory.* McGraw-Hill, 1961.

[SBS66] M. Schwartz, W. R. Bennett, and S. Stein. *Communication Systems and Techniques.* Mc Graw-Hill, 1966.

[Sch68] Francis Scheid. *Schaum's Outline Series - Theory and Problems of Numerical Analysis.* Mc Graw-Hill, 1968.

[Ste79] Thomas E. Stern. Approximations of queue dynamics and their application to adaptive routing in computer communication networks. *IEEE Transactions on Communications*, 27(9):1331–1335, September 1979.

[Taa82] M. R. Taafe. *Approximating nonstationary queueing systems.* PhD thesis, Ohio State University, Columbus, OH, 1982.

[Tak61] Lajos Takács. *The transient behavior of a single server queueing process with a Poisson input*, pages 535–567. University of California Press, 1961.

[Tak62] Lajos Takács. *Introduction to the Theory of Queues.* Oxford University Press, 1962.

[TS89] David Tipper and Malur K. Sundareshan. Modeling of computer networks for performance evaluation and control under nonstationary conditions. In *Proceedings of the IEEE, INFOCOM vIII*, 1989.

[TS90] David Tipper and Malur K. Sundareshan. Numerical methods for modeling computer networks under nonstationary conditions. *Proceedings of the IEEE*, 8(9):1682–1695, December 1990.

[Wra63] A. Wragg. The solution of an infinite set of differential-difference equations occuring in polymerization and queueing problems. *Proceedings of the Cambridge Philosophical Society*, 59:117–124, 1963.