OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

An Exact FCFS Waiting-Time Analysis for
a General Class of G/G/s Queueing Systems

by

Dimitris Bertsimas

OR 157-87 March 1937

An exact FCFS waiting-time analysis for a general class of $G/G/s$ queueing systems

by

Dimitris Bertsimas

Mathematics department and Operarions Research Center

Massachussetts Institute of Technology

Rm 2-342, Cambridge, Mass. 02139, USA

Abstract

A closed form expression for the waiting-time distribution under FCFS is derived for the queueing system $C_k/C_m/s$, where C_k is the class of Coxian probability density functions (pdfs) of order k, which is a subset of the pdfs that have rational Laplace transform (R). Using the calculus of difference equations and based on previous results of the author it is proved that the waiting-time distribution is a mixture of $\binom{s+m-1}{s}$ exponential terms. Our approach offers qualitative insight by providing exact and asymptotic expressions, generalizes and unifies the theories developed for the well known $G/G/1$, $G/M/s$ QS and leads to an $O(k^3\binom{s+m-1}{s}^3)$ algorithm, which is polynomial if only one of the parameters s or m varies, but it is exponential if both parameters vary. As an example numerical results for the waiting-time distribution of the $C_2/C_2/s$ queueing system are presented.

Keywords: Multichannel queues, Coxian pdf, Waiting-time distribution

Acknowledgments

I would like to express my appreciation to my advisor Professor Amedeo Odoni for his encouragement and support throughout the course of my Master thesis, a part of which is this paper. I would also like to thank the Mathematics department of MIT for awarding me an Applied Mathematics Fellowship.

1 Introduction

The explicit evaluation, either by analytic or by numerical means, of the waiting-time distribution in a general multi-server queueing system (QS) is by now well known to present substantial difficulties. Up to the beginning of the last decade, the only revelant results available in the literature were those for the $M/D/s$ system, found in the first half of the century by Erlang and Crommelin (as reviewed by Takacs [22]) and for the $G/M/s$ system, first given independently by Pollaczek [15] and Kendall [10]. In this respect, it is reasonable to expect that the more we depart beyond the $M/D/s$ and $G/M/s$ assumptions, (which are seldom validated in practice), the harder the relevant problem becomes.

In this paper based on previous results of the author (Bertsimas [2]), which are summarized in section 2.3, we derive closed-form expressions for the waiting-time distribution under FCFS for the $C_n/C_m/s$ QS, where C_k is the class of Coxian probability density functions (pdfs) of order k, which is a subset of the pdfs that have rational Laplace transform (R).

1.1 Related work

Although a historical review and description of related work is given in Papaconstantinou and Bertsimas [14] we include for completeness the following very short commentary on the application of solution approaches to multiserver models, which assesses fairly their inherent potential to produce theoretical and, especially, exact numerical results (which should usually be their ultimate goal) for the waiting-time distribution. In the area of exact waiting-time analysis for multiserver QS's with non-constant or non-exponential service-times, the only "potentially able to provide numerical results" approaches (always restricted to FCFS) are, up to date, due to:

1. Pollaczek [16],[17](see also Syski [21])who presents for the $G/G/s$ model a system of equations and describes how they should be solved for the $G/R_m/s$

case, but restricts his analysis to the s=2 case, which still necessitates considerable efforts to obtain numerical results. He also claims that the waiting-time distribution is a mixture of at most $\binom{s+m-1}{s}$ exponential terms. His derivations, being long and complicated, lack any probabilistic interpretation, need deep arguments from complex variable theory and for $s > 2$ lead to almost intractable mathematics. It should be stressed, however, that Pollaczek's ideas were pioneering and his work exerted tremendous impact on the development of queueing theory.

2. Avis [1], who develops for the virtual waiting time of the $G/E_m/s$ system a computational procedure based on Yu's [25] theoretical treatise. Because of the inherent complexity of the algorithmic method used, this procedure is applicable to $M/E_2/2, E_k/E_2/2$ and $D/E_2/2$ systems.

3. Cohen [3] who (constructing a 2-dimensional functional equation and applying several times a Wiener-Hopf decomposition) investigates the $M/H_m/2$ system, but derives an explicit Laplace-Stieltjes transform which he admits to be very complicated and not suitable for further analysis (even for studying its asymptotic properties).

4. De Smit [19], who, (continuing and expanding the work of Pollaczek), considers the waiting-time distribution of $G/H_m/s$ and reports numerical results for $G/H_2/s$ in [20]. Yet, this approach leads to an algorithm of very high computational complexity, which for the $G/H_2/s$ is $O(s^6)$ compared with $O(s^3)$ of our method.

5. Ishikawa [9], who (proceeding through an intricate analysis based on supplementary variables), gives numerical results for the waiting-time distribution of the $G/E_m/s$ system for very small values of m, s $(m \leq 3, s \leq 3)$.

6. Papaconstantinou and Bertsimas [14], who describe a probabilistic method based on the post-departure probabilities for the $E_k/G/s$ QS. Using the method of stages, they show that if the probability distribution for the number of stages in the system is known, then the waiting-time distribution can be easily obtained and apply this method to derive explicit formulas and numerical results for the waiting-time distribution of the $E_k/C_2/s$ QS.

Regarding approximations of the exact waiting-time distribution for multi-server QS, Hokstad [5] presented an approximate expression for the waiting-time distribution of a $M/G/s$ QS by extending the formulae of the $M/G/1$ QS and in [6] gave bounds for the mean queue length of the $M/C_2/s$ QS. Takahashi [23] and Neuts and Takahashi [13] gave an expression for the tail of the distribution for $PH/PH/s$ and $G/PH/s$ QS's respectively. We discuss the relation between their and our result in section 3.3 of this paper. Van Hoorn and Tijms [8] presented an approximation (in the form of a defective renewal equation) for the waiting-time distribution of the $M/G/s$ QS. Seelen [18], using the previous results of Takahashi, proposed for the $PH/PH/s$ QS a relatively simple approximation, for which though, there is no guarantee of being a proper pdf and of preserving the correct asymptotic behaviour. The interested reader should also try to read van Hoorn's [7] lucid discussion and typical results of various exact and approximate algorithmic methods on QS's.

1.2 Organization of the paper

In section 2, we describe the system, explain its structure and the notation used and review the results for the steady-state probability distribution for the number of customers in the QS from Bertsimas [2] that we will use. In section 3, which is the most important part of the paper, we write down the equations that describe the system and then use the calculus of difference equations to derive a closed form expression for the waiting-time distribution. Furthermore, in section 3.3 we examine the asymptotic behavior of the waiting-time distribution. In section 3.4 we show that the results for two seemingly very different QS, $G/G/1$ and $G/M/s$, which are traditionally analysed with very different methods (Wiener-Hopf decomposition and Imbedded Markov Chain respectively) are unified using the results of the present paper. In section 4, the established theoretical results, which are considerably simpler than those mentioned in section 1.1, are used to write a computationally efficient algorithm in order to extract numerical results. As an example, the algorithm is applied to the general class of C_2 distributions, i.e. for the $C_2/C_2/s$ QS, for which numerical results are reported.

2 Description of the system and previous results

We shall examine, henceforth, an s identical-single-waiting-line QS with interarrival and service time distributions of Coxian type of order k and m respectively. The queue discipline is First-Come-First-Served (FCFS).

2.1 Probabilistic interpretation of the QS

The general Coxian class C_n was introduced in Cox's [4] pioneering paper and is clearly presented in Kleinrock [11]. Graphically it is presented in figure 1. It is remarkable that even if we permit transitions from a stage with rate μ_i to a stage with rate μ_j in figure 1 we do not obtain a new class of distributions. We can still formulate this situation with a C_n distribution with different transition rates. The salient feature of the class C_n is its high versatility (see e.g. Neuts [12], Whitt [24]) based on its ability to:

1. Generalize well-known distributions such as the exponential, the hyperexponential and all the forms (i.e., special, general, weighted, compound, etc.) of the Erlangian.

2. Be dense in the set of all probability distributions concentrated on $(0, \infty)$ and thus to be able to approximate a general pdf.

3. Permit coefficients of variation V_s^2 greater than 1.

Figure 1: The C_m class of distributions

To analyse the model we conceive of the arrival process as an arrival timing channel (ATC) consisting of k consecutive stages with rates $\lambda_1, \lambda_2, ..., \lambda_k$ and with probabilities $p_1, p_2, ..., p_k \triangleq 1$ of entering the QS after the completion of the 1st, 2nd, ... kth stage. We remark that as soon as a customer in the ATC enters the QS a new customer arrives at stage 1 of the ATC. For the service time distribution we consider as above a service timing channel (STC) consisting of m consecutive stages with rates $\mu_1, \mu_2, ..., \mu_m$ and with probabilities $q_1, q_2, ..., q_m \triangleq 1$ of leaving the system.

2.2 Notation

For the steady-state we introduce the random variables:

1. $N \triangleq$ The number of customers in the system.

2. $N^- \triangleq$ The number of customers seen by an arriving customer just before his arrival.

3. $R_a \triangleq$ The number of the ATC stage currently occupied by the arriving customer.

4. $R_j \triangleq$ The number of customers being served at the jth STC stage ($j = 1, 2, ..., m$).

5. $R_j^- \triangleq$ The number of customers being served at the jth STC stage ($j = 1, 2, ..., m$), just before the arrival of an entering customer.

6. $T_q \triangleq$ The waiting-time of an arriving customer.

7. $L_q \triangleq$ The length of the queue.

For simplicity of notation we introduce the vectors of random variables

$$\vec{R} \triangleq (R_1, ..., R_m), \quad \vec{R^-} \triangleq (R_1^-, ..., R_m^-)$$

and also we will use the notation:

$$\vec{\delta_j} \triangleq (0, ..., 0, 1, 0..., 0) \qquad a(s, m) \triangleq \binom{s+m-1}{s}$$

$|\vec{i}| = s$ meaning that $\sum_{j=1}^{m} i_j = s$.

With the above definitions the system can be formulated as a continuous time Markov chain with infinite state space:

$$\{(N, R_a, R_1, ..., R_m), N = 0, 1, ..., \quad R_a = 1, 2, ...k, \quad \sum_{j=1}^{m} R_j = min(N, s)\}$$

where the states with $N < s$ (i.e., the states with at least one server free) and $N \geq s$ (or all servers busy) will be termed "unsaturated" and "saturated" respectively. We now introduce the following set of probabilities:

$P_{n,l,\vec{i}} \triangleq Pr\{N = n, R_a = l, \vec{R} = \vec{i}\}$
$P_{n,\vec{i}}^- \triangleq Pr\{N = n, \vec{R}^- = \vec{i}\}$
$P_n \triangleq Pr\{N = n\}$
$P_n^- \triangleq Pr\{N^- = n\}$

We also define:

$f_{T_a}^*(\theta)$, $f_{T_s}^*(\theta) \triangleq$ The Laplace transform of the interarrival and service time distributions respectively.

$\frac{1}{\lambda} \triangleq$ The mean interarrival time.

$\frac{1}{\mu} \triangleq$ The mean service time.

$\rho \triangleq \frac{\lambda}{s\mu}$ = The traffic intensity.

$V_a^2 \triangleq$ The squared coefficient of variation of the interarrival distribution .

$V_s^2 \triangleq$ The squared coefficient of variation of the service-time distribution .

2.3 Review of previous results

In Bertsimas [2], after writing the equations for $P_{n,l,\vec{i}}$ and using seperation of variables and a generating function technique, the following closed form expressions were established for $\rho < 1$:

$$P_{n,l,\vec{i}} = \sum_{j=1}^{a(s,m)} B_j \, (\prod_{r=1}^{l-1} \frac{(1-p_r)\lambda_r}{x(w_j)+\lambda_{r+1}}) \, f(\vec{i}, w_j) \, w_j^n \quad n \geq s, \;\; l = 1, \ldots, k, \;\; |\vec{i}| = s \tag{1}$$

where B_j are computed by solving a linear system of $a(s,m)$ equations with $a(s,m)$ unknowns, $f(\vec{i}, w_j)$ are computed from the following recursion :

$$f((0,0,\ldots,s), w_j) = 1$$

$$w_j \sum_{r=1}^{m} q_r \mu_r (i_r+1) f(\vec{i}+\vec{\delta_r}-\vec{\delta_1}, w_j) + \sum_{r=1}^{m} (1-q_r)\mu_r(i_r+1) f(\vec{i}+\vec{\delta_r}-\vec{\delta_{r+1}}, w_j) =$$

$$f(\vec{i}, w_j)\{\sum_{r=1}^{m} i_r \mu_r - x(w_j)\} \;, \;\; |\vec{i}| = s, \; j = 1, \ldots, a(s,m) \tag{2}$$

Each of the $a(s,m)$ roots w_j satisfy the following system of nonlinear equations (the subscript j corresponds to one of the $a(s,m)$ combinations of the vector $\vec{i} = (i_1, i_2, \ldots, i_m)$, $\sum_{r=1}^{m} i_r = s$ and for simplicity of notation $x(w_j)$ is simply written x):

$$\phi_{\vec{i}}(x) \triangleq i_1\theta_1(x) + i_2\theta_2(x) + \ldots + i_m\theta_m(x) = x, \;\; i_1 + i_2 + \ldots + i_m = s \tag{3}$$

where $\theta_j(x) \quad (j = 1, \ldots, m)$ are the m roots of the polynomial equation of degree m:

$$f^*_{T_a}(x) f^*_{T_s}(-\theta_j(x)) = 1 \tag{4}$$

and

$$w = f^*_{T_a}(x) \tag{5}$$

The generating function of the coefficients $f(\vec{i}, w_j)$

$$G_j(\vec{z}) \triangleq \sum_{|\vec{i}|=s} f(\vec{i}, w_j) z_1^{i_1} \dots z_m^{i_m}, \quad \vec{i} = (i_1, \dots, i_m), \quad \vec{z} = (z_1, \dots, z_m)$$

satisfies the following linear partial differential equation

$$\sum_{r=1}^{m} \frac{\partial G_j(\vec{z})}{\partial z_r} (\mu_r z_r - w_j z_1 q_r \mu_r - (1 - q_r)\mu_r z_{r+1}) = x(w_j) G_j(\vec{z}) \tag{6}$$

was found to be (each j corresponding to a vector $\vec{i}$) :

$$G_j(\vec{z}) = \prod_{r=1}^{m} \left(\frac{b_{1,r}(w_j) z_1 + \dots + b_{m,r}(w_j) z_m}{b_{m,r}(w_j)}\right)^{i_r} \tag{7}$$

and the coefficients $b_{i,r}(w_j)$ are computed from the expansion of a determinant. The "unsaturated" probabilities $P_{n,l,\vec{i}}$ are of the form:

$$P_{n,l,\vec{i}} = \sum_{j=1}^{a(s,m)} B_j \, g(n, l, \vec{i}, w_j) \quad n < s, \; l = 1, \dots, k, \; |\vec{i}| = n \tag{8}$$

where the coefficients $g(n, l, \vec{i}, w_j)$ are computed recursively. Furthermore, the pre-arrival probabilities $P^-_{n,\vec{i}}$ were found to be:

$$P^-_{n,\vec{i}} = \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j \; f(\vec{i}, w_j) \; (\lambda_1 + x(w_j)) w_j^{n+1} \quad n \geq s, \; |\vec{i}| = s \tag{9}$$

3 Waiting time analysis under FCFS

We denote with

$$W(t) \triangleq Pr\{0 < T_q \leq t\}$$

$$F_{T_q}(t) \triangleq Pr\{T_q \leq t\}$$

$$F_{n,\vec{i}}(t) \triangleq Pr\{0 < T_q \leq t | N^- = n + s, \; \vec{R}^- = \vec{i}\}.$$

In this section we shall derive a closed form expression for $W(t)$ and $F_{T_q}(t)$, i.e. the probability distribution for the waiting-time under FCFS of an arriving customer. Then by conditioning on N^- and $\vec{R}^-$ we easily find that

$$W(t) = \sum_{|\vec{i}|=s} \sum_{n=0}^{\infty} F_{n,\vec{i}}(t) P^-_{n+s,\vec{i}}$$

Using (9) we then find

$$W(t) = \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j \ (\lambda_1 + x(w_j)) w_j^{s+1} \{ \sum_{|\vec{i}|=s} f(\vec{i}, w_j) \sum_{n=0}^{\infty} F_{n,\vec{i}}(t) w_j^n \} \qquad (10)$$

3.1 The equations

By conditioning on the next event in the interval $(t, t+\delta t)$ and taking the limit as $\delta t \to 0$ we can write the equations that $F_{n,\vec{i}}(t)$ satisfy

$$\frac{d}{dt} F_{n,\vec{i}}(t) + F_{n,\vec{i}}(t) \sum_{r=1}^{m} i_r \mu_r =$$

$$\sum_{r=1}^{m} i_r q_r \mu_r F_{n-1,\vec{i}+\vec{\delta}_r-\vec{\delta}_1}(t) + \sum_{r=1}^{m} (1-q_r) \mu_r i_r F_{n,\vec{i}+\vec{\delta}_r-\vec{\delta}_{r+1}}(t) \quad n \geq 0, \ \ |\vec{i}| = s \qquad (11)$$

where $F_{-1,\vec{i}}(t) \triangleq 1$, by definition. We define the Laplace transforms

$$W^*(\theta) \triangleq \mathcal{L}(W(t)) \ \ and \ \ \Phi^*_{n,\vec{i}}(\theta) \triangleq \mathcal{L}(F_{n,\vec{i}}(t))$$

and the quantities

$$A_{\vec{i},j}(\theta) \triangleq \sum_{n=0}^{\infty} w_j^n \Phi^*_{n,i}(\theta)$$

$$H_j(\theta) \triangleq \sum_{|\vec{i}|=s} f(\vec{i}, w_j) A_{\vec{i},j}(\theta)$$

Then from (10) we take

$$W^*(\theta) = \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j \ (\lambda_1 + x(w_j)) w_j^{s+1} \ H_j(\theta) \qquad (12)$$

Although (12) seems dissapointing we will find a very simple formula for $H_j(\theta)$. We now transform (11) and obtain

$$\{\theta + \sum_{r=1}^{m} i_r\mu_r\}\Phi^*_{n,\vec{i}}(\theta) = \sum_{r=1}^{m} i_r q_r \mu_r \Phi^*_{n-1,\vec{i}+\vec{\delta_r}-\vec{\delta_1}}(\theta) +$$

$$\sum_{r=1}^{m}(1-q_r)\mu_r i_r \Phi^*_{n,\vec{i}+\vec{\delta_r}-\vec{\delta}_{r+1}}(\theta) \quad n \geq 0, \ |\vec{i}| = s \tag{13}$$

where

$$\Phi^*_{-1,\vec{i}}(\theta) \triangleq \mathcal{L}\{F_{-1,\vec{i}}(t)\} = \mathcal{L}\{1\} = \frac{1}{\theta}$$

3.2 The basic method

The strategy for obtaining a closed form expression for $W^*(\theta)$ is the following:

1. We obtain a difference equation for $A_{\vec{i},j}(\theta)$.
2. We multiply the equation for $A_{\vec{i},j}(\theta)$ by the coefficients $f(\vec{i}, w_j)$, add with respect to $\vec{i}$ and using the equations (2), we are able to solve for $H_j(\theta)$.

We multiply (13) with w_j^n and add with respect to n to take

$$\{\theta + \sum_{r=1}^{m} i_r\mu_r\}A_{\vec{i},j}(\theta) = \frac{1}{\theta}\sum_{r=1}^{m} i_r q_r \mu_r + w_j \sum_{r=1}^{m} i_r q_r \mu_r A_{\vec{i}+\vec{\delta_r}-\vec{\delta_1},j}(\theta) +$$

$$\sum_{r=1}^{m}(1-q_r)\mu_r i_r A_{\vec{i}+\vec{\delta_r}-\vec{\delta}_{r+1},j}(\theta) \tag{14}$$

Performing now the second step we get

$$\theta H_j(\theta) + \sum_{|\vec{i}|=s} f(\vec{i}, w_j) A_{\vec{i},j}(\theta) \sum_{r=1}^{m} i_r\mu_r = \frac{1}{\theta}\sum_{r=1}^{m} q_r\mu_r \sum_{|\vec{i}|=s} i_r f(\vec{i}, w_j) +$$

$$\sum_{|\vec{i}|=s} A_{\vec{i},j}(\theta)\{w_j \sum_{r=1}^{m} q_r\mu_r(i_r+1) f(\vec{i}+\vec{\delta_r}-\vec{\delta_1}, w_j) + \sum_{r=1}^{m}(1-q_r)\mu_r(i_r+1) f(\vec{i}+\vec{\delta_r}-\vec{\delta}_{r+1}, w_j)\}$$

(15)

Substituting (2) into the rhs of (15) we find

$$\theta H_j(\theta) + \sum_{|\vec{i}|=s} f(\vec{i}, w_j) A_{\vec{i},j}(\theta) \sum_{r=1}^{m} i_r \mu_r = \frac{1}{\theta} \sum_{r=1}^{m} q_r \mu_r \sum_{|\vec{i}|=s} i_r f(\vec{i}, w_j) +$$

$$\sum_{|\vec{i}|=s} A_{\vec{i},j}(\theta)\ f(\vec{i}, w_j)\ (\sum_{r=1}^{m} i_r \mu_r - x(w_j)) \tag{16}$$

Since in (16) the term

$$\sum_{|\vec{i}|=s} f(\vec{i}, w_j)\ A_{\vec{i},j}(\theta) \sum_{r=1}^{m} i_r \mu_r$$

cancels from both sides we get

$$\theta H_j(\theta) = \frac{1}{\theta} \sum_{r=1}^{m} q_r \mu_r \sum_{|\vec{i}|=s} i_r f(\vec{i}, w_j) - x(w_j) H_j(\theta)$$

Hence we can solve for $H_j(\theta)$

$$H_j(\theta) = \frac{\sum_{r=1}^{m} q_r \mu_r \sum_{|\vec{i}|=s} i_r f(\vec{i}, w_j)}{\theta(\theta + x(w_j))} \tag{17}$$

From (6) for $\vec{z} = \vec{1} \triangleq (1, 1, \ldots, 1)$ we find that

$$\sum_{r=1}^{m} q_r \mu_r \sum_{|\vec{i}|=s} i_r f(\vec{i}, w_j) = \sum_{r=1}^{m} q_r \mu_r \frac{\partial G_j(\vec{z})}{\partial z_r}|_{\vec{z}=\vec{1}} = \frac{x(w_j)}{1 - w_j} G_j(\vec{1})$$

from where (17) becomes

$$H_j(\theta) = \frac{x(w_j)\ G_j(\vec{1})}{1 - w_j} \frac{1}{\theta(\theta + x(w_j))} \tag{18}$$

Substituting (18) into (12) we obtain

$$W^*(\theta) = \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j\ G_j(\vec{1})(\lambda_1 + x(w_j)) \frac{w_j^{s+1}}{1 - w_j} \frac{x(w_j)}{\theta(\theta + x(w_j))}$$

Using partial fractions we find

$$W^*(\theta) = \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j\ G_j(\vec{1})(\lambda_1 + x(w_j)) \frac{w_j^{s+1}}{1 - w_j} (\frac{1}{\theta} - \frac{1}{\theta + x(w_j)}) \tag{19}$$

Now the inversion of (19) is an easy task. Thus

$$W(t) \triangleq Pr\{0 < T_q \le t\} = \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j \, G_j(\vec{1}) \, (\lambda_1 + x(w_j)) \frac{w_j^{s+1}}{1 - w_j} (1 - e^{-x(w_j)t}) \quad (20)$$

As a check on the algebra we verify that in the limit $t \to \infty$ we find

$$\lim_{t \to \infty} W(t) = Pr\{T_q > 0\} = \sum_{n=s}^{\infty} P_n^- = \sum_{n=s}^{\infty} \sum_{|\vec{i}|=s} P_{n,\vec{i}}^-$$

Also

$$F_{T_q}(t) = 1 - \frac{1}{\lambda} \sum_{j=1}^{a(s,m)} B_j \, G_j(\vec{1}) \frac{w_j^{s+1}}{1 - w_j} (\lambda_1 + x(w_j)) e^{-x(w_j)t} \quad (21)$$

where from (5) $w_j = f_{T_s}^*(x(w_j))$. It is remarkable that the waiting-time pdf has also rational Laplace transform, i.e. it belongs to the class $R_{a(s,m)}$ of distributions. From (21) it is easy to find the following compact expression for the rth moment of T_q

$$E\{T_q^r\} = \frac{r!}{\lambda} \sum_{j=1}^{a(s,m)} B_j \, G_j(\vec{1}) \frac{w_j^{s+1}}{1 - w_j} \frac{(\lambda_1 + x(w_j))}{[x(w_j)]^r}$$

As an additional check on the algebra we calculate the factorial moments of L_q

$$E\{L_q(L_q - 1) \dots (L_q - r + 1)\} \triangleq \sum_{n=s}^{\infty} (n - s)(n - s - 1) \dots (n - s - r + 1) \; P_n$$

Since for $n \ge s$

$$P_n = \sum_{l=1}^{k} \sum_{|\vec{i}|=s} P_{n,l,\vec{i}}$$

then from (1) we find

$$P_n = \sum_{j=1}^{a(s,m)} B_j \, G_j(\vec{1}) \frac{\lambda_1 + x(w_j)}{x(w_j)} w_j^n (1 - w_j) \quad n \ge s$$

from where, after algebraic manipulations and using the identity

$$\sum_{n=0}^{\infty} n(n-1) \dots (n - r + 1) \; a^{n-r} = \frac{r!}{(1-a)^{r+1}}$$

we find

$$E\{L_q(L_q - 1) \dots (L_q - r + 1)\} = r! \sum_{j=1}^{a(s,m)} B_j G_j(\vec{1}) \frac{w_j^{s+r}}{(1 - w_j)^r} \frac{(\lambda_1 + x(w_j))}{x(w_j)}$$

For $r = 1$ we verify Little's formula

$$E\{L_q\} = \lambda E\{T_q\}$$

For $k = 1$ the model becomes $M/C_m/s$ and then we verify the well known result which holds for $M/G/s$

$$E\{L_q(L_q - 1)\dots(L_q - r + 1)\} = \lambda^r \; E\{T_q^r\}$$

where we used the fact that in this case $w_j = f^*_{T_a}(x(w_j)) = \frac{\lambda}{\lambda + x(w_j)}$ $(\lambda_1 = \lambda)$.

3.3 Asymptotic results

Takahashi [23] proved that in a $PH/PH/s$ QS the stationary probability Π_m that there are more than m customers waiting in the queue behaves asymptotically as $m \to \infty$:

$$\Pi_m \sim K_1 \eta^m$$

where the constant K_1 was not computed and $\eta = f^*_{T_a}(sy)$ where y is the unique positive root of the equation:

$$f^*_{T_a}(sy) f^*_{T_s}(-y) = 1$$

Furtermore, he proved that as $t \to \infty$:

$$1 - F_{T_q}(t) \sim K_2 e^{-syt}$$

where $\frac{K_1}{K_2} = \frac{\lambda(1-\eta)}{sy}$.

In order to see the connection of the above results and the results of the present paper, we observe that

$$\Pi_m = \sum_{n=m+s+1}^{\infty} P_n = \sum_{j=1}^{a(s,m)} B_j \; G_j(\vec{1}) \frac{\lambda_1 + x(w_j)}{x(w_j)} w_j^{s+1} w_j^m$$

Thus asymptotically as $m \to \infty$ Π_m behaves as

$$\Pi_m \sim B_1 \, G_1(\vec{1}) \frac{\lambda_1 + x(w_1)}{x(w_1)} w_1^{s+1} w_1^m$$

where w_1 is the root corresponding to the combination of $\vec{i} = (0, \ldots, s)$. Specializing (3),(4) and (5) we find that $w_1 = f^*_{T_a}(x(w_1))$ where $x(w_1)$ is the unique positive root of the equation

$$f^*_{T_a}(x) f^*_{T_s}(-\frac{x}{s}) = 1$$

Thus we see that by letting $y = \frac{x(w_1)}{s}$, $\eta = w_1$ the two results are identical. Furthermore, we observe from (21) that as $t \to \infty$

$$1 - F_{T_q}(t) \sim B_1 \, G_1(\vec{1})(\lambda_1 + x(w_1)) \frac{w_1^{s+1}}{1 - w_1} e^{-x(w_1)t}$$

Again the two results are identical and we can also easily check that $\frac{K_1}{K_2} = \frac{\lambda(1-w_1)}{x(w_1)}$. It should also be stressed that in our expressions we are able to compute expicitely the constants K_1, K_2.

3.4 Two special cases

1. $C_n/C_m/1$

   Since the only combination of $\vec{i}$ for $s = 1$ are of the type $\vec{i} = (0, \ldots, 0, 1, 0, \ldots, 0)$ $(a(1,m) = m)$ we verify a well-known result from $G/G/1$ theory that $F_{T_q}(t)$ is a mixture of m exponential terms of the form of (21) where in this case $x(w_j), j = 1, \ldots, m$ are the m roots of the equation

   $$f^*_{T_a}(x) f^*_{T_s}(-x) = 1$$

   subject to the constraint $|f^*_{T_a}(x)| < 1$.

2. $C_m/M/s$

   For $m = 1$ (4) becomes

   $$f^*_{T_a}(x) f^*_{T_s}(-\frac{x}{s}) = 1 \Rightarrow x = s\mu(1 - f^*_{T_a}(x))$$

Since $a(s,1) = 1$ we find the well known result from $G/M/s$ theory that

$$1 - F_{T_q}(t) = K_3 e^{-xt}$$

where x is the unique root of the above equation.

4 Computational and complexity considerations

4.1 The algorithm

In order to extract numerical results from the formulae presented in the previous section the author in [2] has proposed an algorithm with complexity $O(k^3\binom{s+m-1}{s}^3)$, which is polynomial if only one of the parameters s or m varies, but it is exponential if both parameters vary. In other words, for an arbitrary interarrival distribution and a given service-time distribution the problem of determining the waiting-time distribution under FCFS can be solved in time polynomial in the number of servers. This algorithm is summarized as follows:

1. Determination of the $\binom{s+m-1}{s}$ roots w_j of the system of equations (3), (4),(5).
2. Recursive determination of the coefficients $f(\vec{i}, w_j)$ from (2).
3. Recursive determination of the coefficients $g(n, l, \vec{i}, w_j)$ in (8) from the equations that $P_{n,l,\vec{i}}$ satisfy for $n < s$.
4. Determination of the $\binom{s+m-1}{s}$ unknowns B_j as a solution of a non-homogeneous linear system with $\binom{s+m-1}{s}$ equations.

4.2 The numerical solution of the $C_2/C_2/s$ QS

To fully gauge the performance of the proposed algorithm we programmed it in FORTRAN, because of its inherent superiority in accuracy and speed and in BASIC because of its greater availability in microsystems. The first program has been run on a CYBER 171 and the second on a SPECTRUM 48K, in order to prove that even for such a "difficult" model exact numerical results can be obtained by a practitioner on a small personal computer.

The reasons we selected this model are :

1. It is representative of the general behavior of the algorithm for more general models.

2. It is in real arithmetic.

3. Its complexity is $O(s^3)$.

4. This model allows the determination of exact results when the coefficients of variation of the interarrival and of the service time pdf are both bigger than 1.

Merely as an illustration of the stability and accuracy of the present algorithm, we present in figure 2 some results for the waiting-time complementary distribution $\hat{F}_{T_q}(t) \triangleq 1 - F_{T_q}(t)$ for the $C_2/C_2/s$ QS as V_a^2 varies ($s = 10,\ \rho = 0.9,\ V_s^2 = 5.0$). In figure 3 we show the dependence of $\hat{F}_{T_q}(t)$ for the $C_2/C_2/s$ QS as V_s^2 varies ($s = 5,\ \rho = 0.8,\ V_a^2 = 5.0$).

Figure 2: $\hat{F}_{T_q}(t)$ as a function of V_a^2 for the $C_2/C_2/10$ QS

Figure 3: $\hat{F}_{T_q}(t)$ as a function of V_s^2 for the $C_2/C_2/5$ QS

References

[1] AVIS,D. (1977). "Computing waiting times in $GI/E_k/s$ queueing system", TIMS studies in Management Science, 7, 215-232.

[2] BERTSIMAS,D. (1986). "An analytic approach to a general class of queueing systems", Oper. Res., submitted for publication.

[3] COHEN,J.W. (1982). "On the $M/G/2$ queueing model", Stoch. proc. and their applications, 12, 231-248.

[4] COX,D.R. (1955). "A use of complex probabilities in the theory of stochastic processes" Proceedings of the Cambridge Philosophical Society, 51, 313-319.

[5] HOKSTAD,P. (1978). "Approximations for the $M/G/m$ queue", Oper. Res., 26, 511-523.

[6] HOKSTAD,P. (1986). "Bounds for the mean queue length of the $M/K_2/m$ queue", Europ. Jour. Oper. Res., 23, 108-117.

[7] van HOORN,M.H. (1983). "Algorithms and approximations for queueing systems", Mathematisch Centrum, Amsterdam.

[8] van HOORN,M.H. and TIJMS,H.C. (1982) . "Approximations of the waiting-time distribution of the $M/G/c$ queue", Performance Evaluation, 2, 22-28.

[9] ISHIKAWA,A. (1984). "Stationary waiting time distribution in a $GI/E_k/m$ queue", Journal of Operation Research Sosiety of Japan, 27, 130-149.

[10] KENDALL,D.G. (1953). "Stochastic processes occuring in the theory of queues and their analysis by means of the imbedded Markov chain", The Annals of Mathematical Statistics, 24, 338-354.

[11] KLEINROCK,L. (1975). "Queueing systems; Vol. 1:Theory", John Wiley and Sons, New York.

[12] NEUTS,M.F. (1981). "Matrix geometric solutions in stochastic models :an algorithmic approach",The John Hopkins University Press.

[13] NEUTS,M.F. and TAKAHASHI,Y. (1981),. "Asymptotic behavior of the stationary distribution in the $GI/Ph/c$ queue with heterogeneous servers", Z. Wahrscheinlichkeits-theorie Verw. Gebiete, 57, 441-452.

[14] PAPACONSTANTINOU X. and BERTSIMAS D. (1986) "An exact probabilistic FCFS waiting-time analysis for the $E_k/G/s$ queue with application to the general class of C_2 service-time distributions", Naval Res. Log. Quart., submitted for publication.

[15] POLLACZEK,F. (1961). "Thèorie analytique des problèmes stochastiques relatifs à un groupe de lignes téléphoniques avec dispositif d' attente", Gauthier, Paris.

[16] POLLACZEK,F.(1963). "Summary of "Thèorie analytique des problèmes stochastiques relatifs à un groupe de lignes téléphoniques avec dispositif d' attente", The institute of fluid dynamics and applied mathematics, University of Maryland, Techn. Note BN-323.

[17] POLLACZEK,F. (1964). "Concerning an analytic method for the treatment of queueing problems", 1-42 in W.L.Smith and R.I.Willkinson, (eds), Carolina press.

[18] SEELEN,L.P. (1986). "An algorithm for $PH/PH/c$ queues", Europ. Jour. Oper. Res., 23, 118-127.

[19] de SMIT,J.H.A. (1983). "The queue $GI/M/s$ with customers of different types or the queue $GI/H_m/s$", Adv. Appl. Prob., 15, 392-419.

[20] de SMIT,J.H.A. (1983). "A numerical solution for the multiserver queue with hyper-exponential service times", Oper. Res. Let., 2, 217-224.

[21] SYSKI,R. (1967). "Pollaczek method in queueing theory", in "Queueing theory; recent developments and applications", edited by R.Cruon, The English Univ. Press Ltd. 33-67.

[22] TAKACS,L. (1976). "On fluctuation problems in the theory of queues", Adv. Appl. Prob. 8, 548-583.

[23] TAKAHASHI,Y. (1981). "Asymptotic exponentiality of the tail of the waiting-time distribution in a $PH/PH/c$ queue", Adv. Appl. Prob., 13, 619-630.

[24] WHITT,W. (1982). "Approximating a point process by a renewal process, 1: Two basic methods', Oper. Res., 30, 125-147.

[25] YU,O. (1977). "The steady-state solution of a heterogeneous server queue with Erlang service time", Tims studies in management science, 7, 199-213.

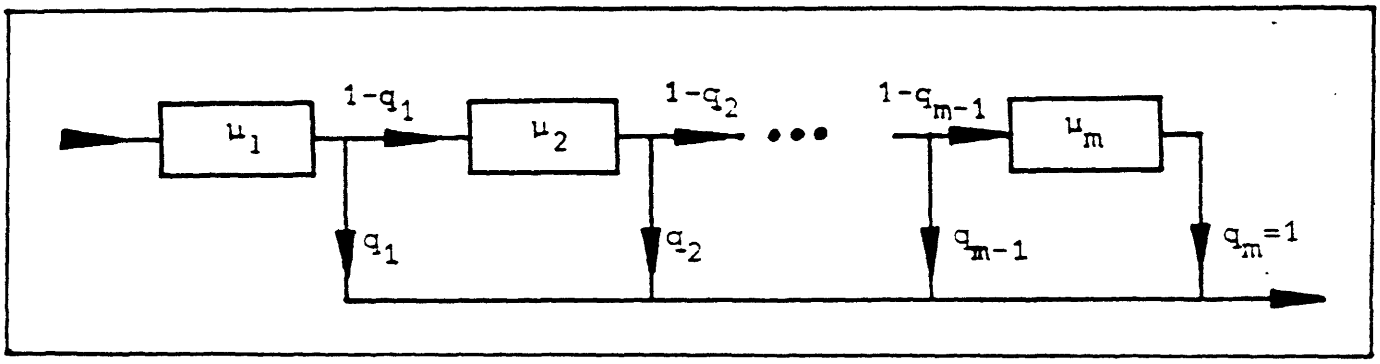

Figure 1: The C_m class of distributions

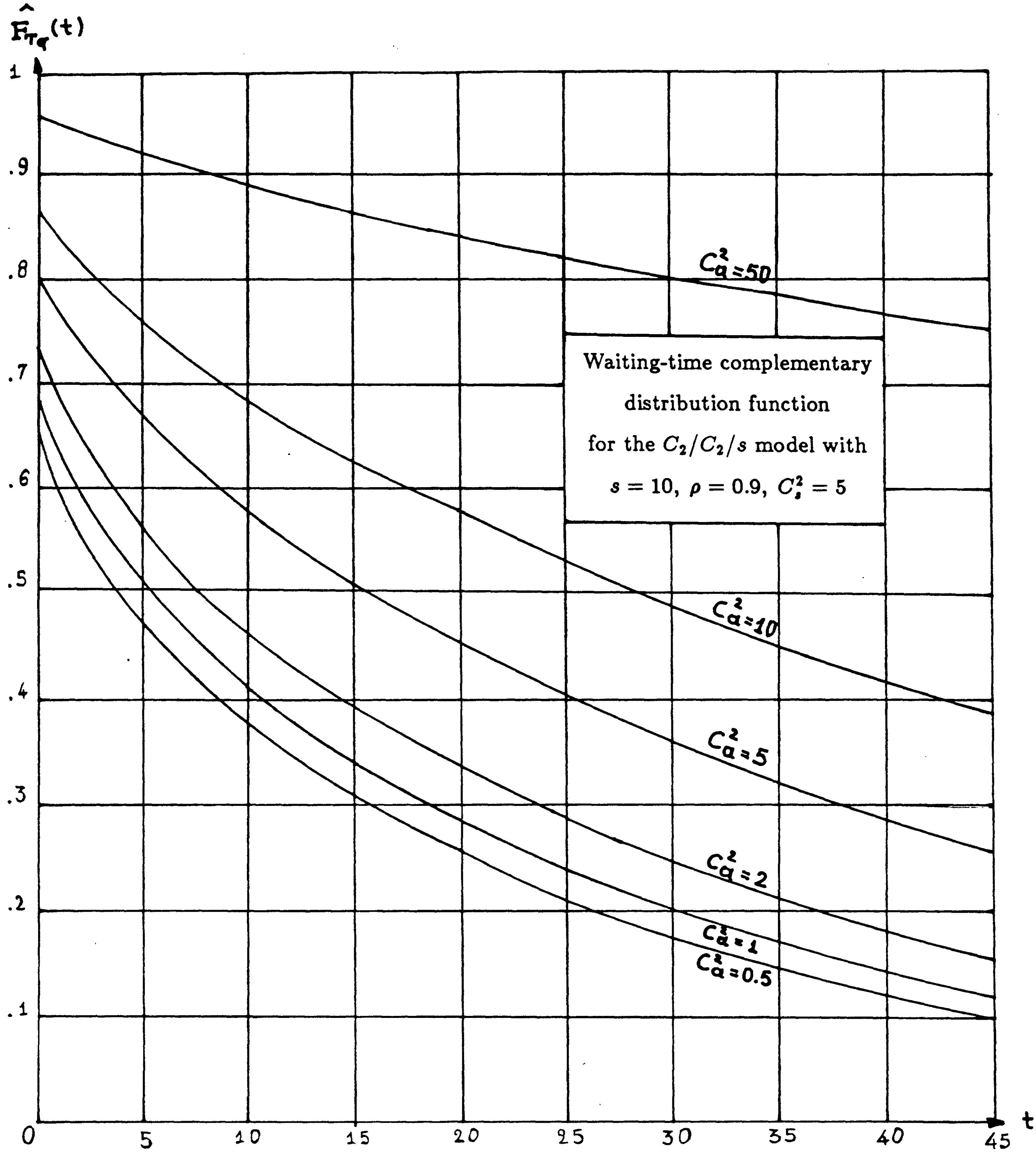

Figure 2: $\hat{F}_{T_q}\ (t)$ as a function of V_a^2 for the $C_2/C_2/10$ QS

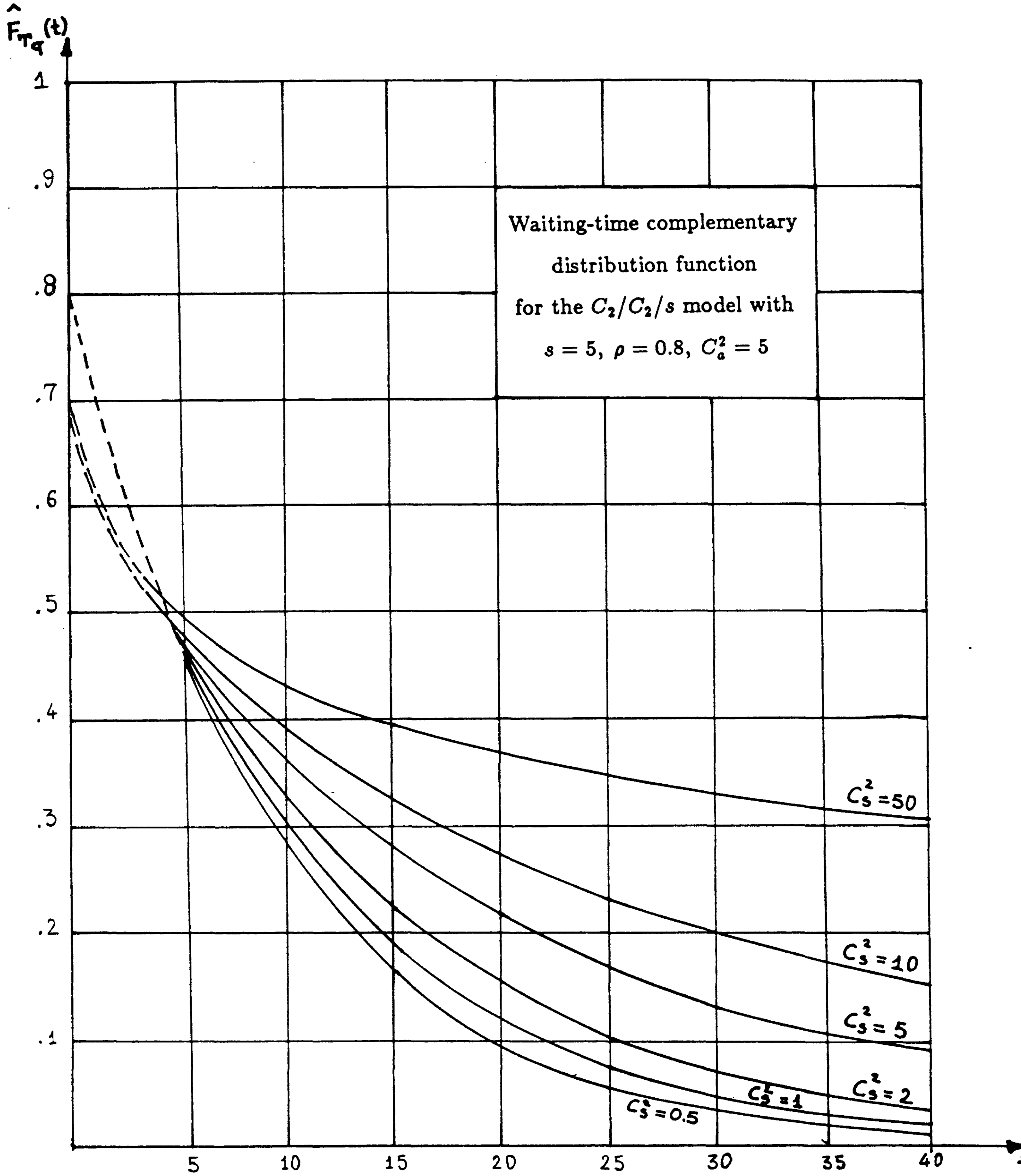

Figure 3: $\hat{F}_{T_q}(t)$ as a function of V_s^2 for the $C_2/C_2/5$ QS