

Approximate Formulae for the Delay in the Queueing System GI/G/1

Wolfgang Krämer and Manfred Lagenbach-Belz*
University of Stuttgart, Stuttgart, Federal Republic of Germany

ABSTRACT

The general single server system GI/G/1 has been treated manifold, but only for some special cases handy formulae are available. Very often exact calculations are too cumbersome and sophisticated for practical engineering, as well as upper and lower bounds generally are too gross approximations.

Therefore the need was felt to support the traffic engineer with simple explicit approximation formulae, based on a 2-moments approximation.

In this paper such formulae are derived heuristically for the mean waiting time and the probability of waiting.

The quality of the formulae, which have been checked by numerous comparisons with exact and simulation results, is such, that within the most interesting range of server utilizations from 0.2 to 0.9 the error is less than 20% (typically < 10%) for all combinations of the arrival and service processes characterized by the following distribution types: D, E_4, E_2, M, H_2 .

Besides these types, for validation purposes, also other distribution functions have been investigated, differing in the third and higher moments.

The formulae are also easily applicable with comparable accuracy to batch arrival systems by considering equivalent single arrival arrangements.

By known relations, also simple approximations are provided, e.g. for the variances of the associated output processes, the first two moments of the idle time distribution and the mean length of a busy period.

Beitrag des Instituts für
Nachrichtenvermittlung und Datenverarbeitung der
Universität Stuttgart zum 8th International Teletraffic
Congress Melbourne vom 10.-17. November 1976

* M. Lagenbach-Belz is now with Standard
Elektrik Lorenz AG, Stuttgart

1 INTRODUCTION

1.1 GENERAL REMARKS

In computers and communications systems very often queueing problems may be represented by queueing systems of the type GI/G/1 (general input and general service process, single server). For traffic engineers in particular, the mean waiting time and the probability of waiting are of interest for system analysis or design. In the literature, exact and explicit solutions for such queueing systems are available only for certain types of arrival or service processes (e.g. M/G/1).

Very often these different solutions for different traffic assumptions require the numerical evaluation of roots of transcendental equations by the aid of computers. For other types of arrival or service processes implicit solutions are known (e.g. based on Lindley's integral equation). But often these solutions are not straightforward and/or require a lot of evaluation work.

In many applications of traffic engineering, either the procedures and tools for these solutions are not available for the engineers in due time or the amount of evaluation work is not justified for quick estimates. On the other hand, known approximation formulae for the delay are limited strongly in application range (e.g. heavy traffic approximations).

Therefore, the urgent need was felt to support the traffic engineer with simple explicit but general approximation formulae for the mean waiting time and the probability of waiting.

The restriction to the first two moments of the interarrival and service time distribution functions (d.f.'s) was near at hand, since e.g.

- in case of Poisson input the mean waiting time only depends on the first two moments of the service time d.f.
- 2-moments approximations have been proved useful e.g. for overflow systems
- the models used for system analysis often are approximate models themselves, and often exact d.f.'s are not known at all

In addition, very encouraging results have been obtained with 2-moments approximations applied to queues in series [17], considering many different d.f.'s of service times (see also Chapter 3.3).

1.2 SOME NOTATIONS

Let be

- T_A = interarrival time between requests
- T_W = waiting time of a request (excluding service)
- T_H = service or holding time of a request
- T_F = flow time of a request ($=T_W+T_H$)
- T_{IP} = duration of an idle period
- T_{BP} = duration of a busy period
- T_D = interdeparture time between requests

Now, the traffic offered A is

$$A = \frac{E(T_H)}{E(T_A)} = \lambda \cdot E(T_H) = \lambda \cdot h, \quad (1.1)$$

with $E(T_H)=h$ as mean service time and λ as arrival rate of the requests. The traffic offered A here is identical with the server utilization, since there are no losses (pure waiting system). The queue discipline may be arbitrary as long as it is independent of the service times.

The d.f.'s of the interarrival and service times be $F_A(t)$ and $F_H(t)$, with associated coefficients of variation

$$c_A = \sqrt{\text{Var}(T_A)} / E(T_A), \quad c_H = \sqrt{\text{Var}(T_H)} / E(T_H) \quad (1.2)$$

1.3 MAIN RESULTS OBTAINED

The key results of this paper are simple explicit approximation formulae for the mean waiting time $E(T_W)$ and the probability of waiting W in a GI/G/1 system:

$$E(T_W) = \frac{A \cdot h}{2(1-A)} \cdot (c_A^2 + c_H^2) \cdot \begin{cases} e^{-\frac{2(1-A)}{3A} \cdot \frac{(1-c_A^2)^2}{c_A^2 + c_H^2}} & c_A^2 \leq 1 \\ e^{-(1-A) \cdot \frac{c_A^2 - 1}{c_A^2 + 4c_H^2}} & c_A^2 \geq 1 \end{cases} \quad (1.3)$$

and

$$W = A + (c_A^2 - 1) \cdot A(1-A) \cdot \begin{cases} \frac{1 + c_A^2 + A \cdot c_H^2}{1 + A(c_H^2 - 1) + A^2(4c_A^2 + c_H^2)} & c_A^2 \leq 1 \\ \frac{4A}{c_A^2 + A^2(4c_A^2 + c_H^2)} & c_A^2 \geq 1 \end{cases} \quad (1.4)$$

The accuracy of these formulae has been tested by comparisons with a large amount of exact and simulation results, including in particular systems with all combinations of D, E_4, E_2, M, H_2 d.f.'s.

From these results, further traffic characteristics can be obtained (e.g. mean queue lengths), also for batch input systems.

1.4 CHAPTER SURVEY

In Chapter 2 a short summary is given about explicit results, existing in the literature and being of special interest for the heuristic development of the formulae (1.3) and (1.4) in Chapter 3, where also validation results are included.

In Chapter 4 it is shown, that also batch input systems may be easily calculated approximately by considering equivalent single arrival systems.

Chapter 5 finally demonstrates that via known relations also explicit and simple formulae for further traffic characteristics (e.g. the variances of the associated output processes) now are available.

2 EXISTING EXPLICIT RESULTS

The substantial amount of publications relating to the general queueing system GI/G/1 demonstrates the state of the art of queueing theory, which nevertheless cannot reveal the gap between exact mathematical results and quick engineering applications.

Several calculation methods have been developed to solve GI/G/1 problems, including

- ERLANG's method of phases, see e.g. [1] - [3]
- LINDLEY's integral method [4]
- KENDALL's method of the imbedded Markov chain [5]

These methods and others have been proved very powerful to penetrate into the theoretical depth and to develop many exact results. But for practical applications there are severe disadvantages:

- 1) Some methods require special traffic assumptions in order to be applicable
- 2) Most of the existing exact results include Laplace transforms and generating functions as well as roots of partly very sophisticated equations
- 3) For each type of a special GI/G/1 system, the way of obtaining numerical results may be more or less different

In the essence, simple explicit and exact formulae for the mean waiting time and the probability of waiting exist only for the case of Poisson input (M/G/1). The well-known formula of POLLACZEK and KHINTCHINE (1930/32), cf. e.g. [3], is

$$E(T_W) = \frac{A(1+c_H^2)}{2(1-A)} \cdot h, \quad (2.1)$$

here only given for the first moment of the waiting time.

The associated probability of waiting simply is

$$W = A \quad (2.2)$$

The need for bridging the above mentioned gap has been recognized manifold and resulted e.g. in

- the application of numerical inversion techniques for Laplace transforms
- fitting observed distributions by step functions [6] or phase-type functions
- tables for queueing systems of different types, see e.g. [19]
- the derivation of upper and lower bounds for the mean waiting times [9], [10], [11].

KINGMAN [7] has derived an upper bound for the mean waiting time in GI/G/1

$$E(T_W) \leq \frac{A[(c_A^2/A^2) + c_H^2]}{2(1-A)} \cdot h, \quad (2.3)$$

being a good approximation for heavy traffic ($A \rightarrow 1$).

Some authors also derived lower bounds or low traffic approximations ([8], [9]) for single arrival GI/G/1 queues.

Also diffusion approximation methods have been applied [2], to obtain heavy traffic formulae even for transient conditions. The mean virtual delay for stationary conditions has been approximated recently in [12] by

$$E(T_W)_{\text{virtual}} = \frac{A(c_A^2 + c_H^2)}{2(1-A)} \cdot h \quad (2.4)$$

Unfortunately the upper and lower bounds for real mean waiting times are not very useful for the major interesting range of utilizations. E.g. the heavy traffic formula (2.3) applied to $A=0.7$ overestimates the mean waiting times up to 100%.

Therefore, the goal was set to derive purely heuristically an extension of the Pollaczek-Khintchine formula, allowing to give quick answers with a reasonable accuracy, though being restricted to the first 2 moments of interarrival and service time d.f.'s.

3 SYSTEMS WITH SINGLE ARRIVALS

The restriction to single arrivals within this chapter means that interarrival times $T_A=0$ are not allowed, i.e. $F_A(0)=0$.

3.1 APPROXIMATION OF THE MEAN WAITING TIME

3.1.1 BASIS OF SOLUTION

Preliminary investigations had shown, that a good starting point for the heuristic approximation of the mean waiting time was the formula

$$E(T_W) = \frac{A \cdot h}{2(1-A)} \cdot (c_A^2 + c_H^2) \cdot g(A, c_A^2, c_H^2), \quad (3.1)$$

containing a refinement function $g(\cdot)$, which is derived in the following.

The function $g(\cdot)$ now is restricted by several boundary conditions:

- a) For $c_A^2=1$ the Pollaczek-Khintchine formula must result $\rightarrow g(A, 1, c_H^2) = 1$ (3.2)
- b) For $A \rightarrow 1$ the heavy traffic formula (2.3) should result $\rightarrow g(1, c_A^2, c_H^2) = 1$ (3.3)
- c) For D/D/1 systems there is no waiting at all $\rightarrow g(A, 0, 0) < \infty$ (3.4)

3.1.2 FORMULA DEVELOPMENT

During formula development it proved useful to distinguish between two different ranges for the refinement function $g(\cdot)$, $c_A^2 \leq 1$ and $c_A^2 \geq 1$, respectively.

3.1.2.1 Hypoexponential Interarrival d.f.'s

From separate pre-investigations of D/M/1 systems a useful explicit approximation turned out to be

$$E(T_W)_{D/M/1} = \frac{A \cdot h}{2(1-A)} \cdot e^{-\frac{2(1-A)}{3A}} \quad (3.5)$$

The combination of all these boundary conditions (3.2)-(3.5) led to the form

$$g(A, c_A^2, c_H^2) = e^{-\frac{2(1-A)}{3A} \cdot \frac{(1-c_A^2)^b}{a \cdot c_A^2 + c_H^2}} \quad (c_A^2 \leq 1) \quad (3.6)$$

where a and b are free parameters. Regarding many other systems with $0 < c_A^2 < 1$ (e.g. E_4 and E_2 inputs), the best compromise turned out to be $a=1, b=2$, which led to the final result (1.3) for $c_A^2 \leq 1$.

3.1.2.2 Hyperexponential Interarrival d.f.'s

In order to fulfill (3.2) and (3.3) for $c_A^2 \geq 1$, the form

$$g(A, c_A^2, c_H^2) = e^{-\frac{(1-A) \cdot \frac{(c_A^2-1)c}{a \cdot c_A^2 + b \cdot c_H^2}}{3A}} \quad (c_A^2 \geq 1) \quad (3.7)$$

has been selected. Again a, b, c are still free parameters.

In case of $H_2/D/1$ systems with $c_A^2=2$, only a is relevant in this expression. Comparisons with simulation results led to $a=1$. Using this result and considering $H_2/D/1$ systems with $c_A^2=4$, it turned out, that $c=1$ is a reasonable choice.

Finally, $H_2/M/1$ systems comparisons led to $b=4$, thus resulting the final result (1.3) for $c_A^2 \geq 1$.

(Throughout this paper, H_2 is a hyperexponential d.f. with 2 alternative branches being utilized equally.)

Naturally, this heuristic procedure was not straightforward and required revision of a former parameter selection. Furthermore, the systematic finding of fitting expressions was facilitated by defining boundary conditions and utilizing the degrees of freedom within them.

3.1.3 COMPARISONS OF NUMERICAL RESULTS

Fig. 3.1 and 3.2 are some examples of comparisons between the approximate results according to eq. (1.3) and results of simulation runs.

As can be seen, the coincidence between approximate and simulation results is sufficiently good. General error considerations will be done in 3.3, also for other d.f.'s.

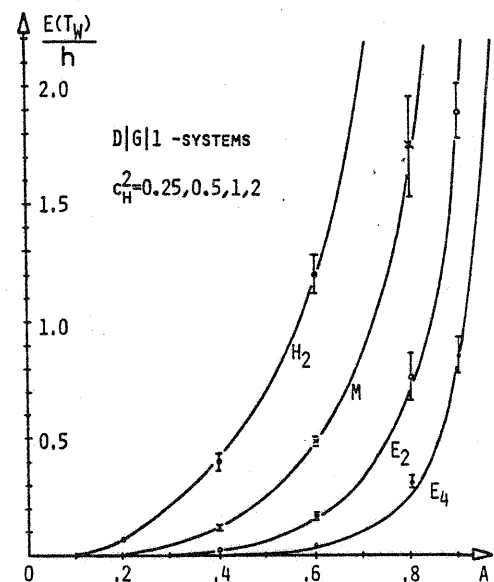


Fig. 3.1 Mean Waiting Times for D/G/1 Systems ([Simulation results with 95% confidence intervals])

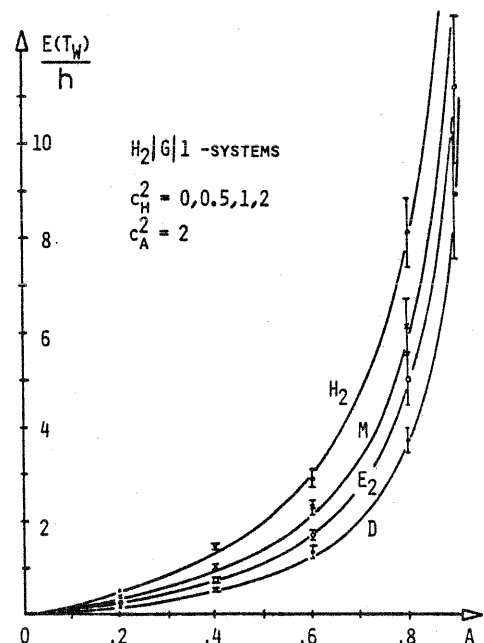


Fig. 3.2 Mean Waiting Times for $H_2/G/1$ Systems

3.2 APPROXIMATION OF THE PROBABILITY OF WAITING

3.2.1 BASIS OF SOLUTION

Similarly as in the previous case for the mean waiting time, the frame of the formula for the probability of waiting W had to satisfy different boundary conditions. As a reasonable base it was selected

$$W = A + (c_A^2 - 1) \cdot A \cdot (1-A) \cdot f(A, c_A^2, c_H^2) \quad (3.8)$$

with $f(\cdot)$ as another refinement function. Boundary conditions are

- a) For $c_A^2 = 1$ (e.g. Poisson input) it holds $W = A$
- b) For $A \rightarrow 1$ also $W \rightarrow 1$, for $A \rightarrow 0$ also $W \rightarrow 0$
- c) $W < A$ for $c_A^2 < 1$ and $W > A$ for $c_A^2 > 1$ are well-known general trends.

3.2.2 FORMULA DEVELOPMENT

As previously, hypo- and hyperexponential types of interarrival d.f.'s have been treated separately.

3.2.2.1 Hypoexponential Interarrival d.f.'s

Since in D/D/1 $W=0$ for stationary conditions, it is necessary that

$$f(A, 0, 0) = \frac{1}{1-A} \quad (3.9)$$

This led to the general form of

$$f(A, c_A^2, c_H^2) = \frac{1 + ac_A^2 + bc_H^2}{1 - A + cc_A^2 + dc_H^2} \quad (c_A^2 \neq 1) \quad (3.10)$$

where a, b, c, d again are free parameters, which in addition may depend on A .

In the first step, D/G/1 systems have been considered for the determination of b and d . Separate pre-investigations for D/M/1, where

$$A = -\frac{1-W}{\ln W} \quad (3.11)$$

holds, led to the approximation

$$W = \frac{2A^3}{1+A^2} \quad (3.12)$$

Now, (3.12) and (3.10) in (3.8) with $c_A^2 = 0$ and $c_H^2 = 1$ yielded

$$\frac{1+A}{1+A^2} = \frac{1+b}{1-A+d}$$

from which $b=A$ was adopted, rendering $d=A(1+A)$.

Fortunately, this selection of b and d proved to be useful for other D/G/1 systems, too.

Further considerations of E_2 and E_4 arrival processes similarly yielded $a=1$ and $c=4A^2$, which led to the final result (1.4) for $c_A^2 \leq 1$.

3.2.2.2 Hyperexponential Interarrival d.f.'s

For $c_A^2 > 1$ the following form of the refinement function $f(\cdot)$ was selected:

$$f(A, c_A^2, c_H^2) = \frac{a}{bc_A^2 + cc_H^2} \quad (3.13)$$

The parameters a, b, c again are possibly dependent on A .

Considering $H_2/D/1$ systems, $a=4A$ and $b=1+4A^2$ have been found to fit reasonably with simulation results. Since the influence of c_H^2 on W was relatively small, $c=A^2$ has been found quickly, which led to the final result (1.4) for $c_A^2 > 1$.

3.2.3 COMPARISONS OF NUMERICAL RESULTS

Fig. 3.3 shows as an example some approximation results for the probabilities of waiting for systems with E_4 input processes, compared with simulation results (small confidence intervals omitted).

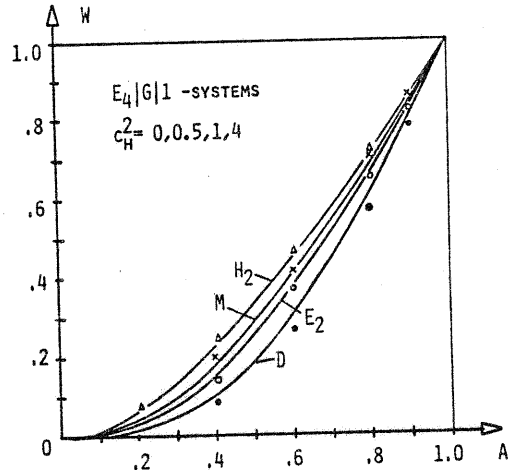


Fig. 3.3 Waiting Probability for $E_4/G/1$ Systems

3.3 ERROR CONSIDERATIONS AND LIMITS OF APPLICATION

3.3.1 'STANDARD' DISTRIBUTION FUNCTIONS

Of special interest for the application of the approximations is their accuracy. Since only the first two moments of both interarrival and service time d.f. are considered, certain errors have been tolerated from the start.

First of all, as a dedicated statement, for all GI/G/1 systems with arbitrary combinations of the d.f.'s

$$D, E_4, E_2, M, H_2 (c^2=2), H_2 (c^2=4)$$

(here called 'standard' types), the error in the mean waiting time is less than $\approx 20\%$ for traffics offered $A = 0.2$ up to 0.9 . Typically, the error is less than 10% , cf. fig. 3.1 and 3.2.

The same error limits are obtained even if higher variances of H_2 d.f.'s are included, only restricted by the condition

$$c_A^2 + c_H^2 \lesssim 12,$$

which seems to be uncritical for practical applications.

It is obvious that with respect to the mean waiting time approximation, the associated errors in flow times (waiting + service) are sensibly lower, especially for lower traffics A .

With respect to the probability of waiting these error limits can be maintained, if the 'boundary' combination D/ E_4 /1 is neglected.

3.3.2 OTHER DISTRIBUTION FUNCTIONS

To judge the influence of the 3'rd and higher moments, many other types of d.f.'s have been investigated.

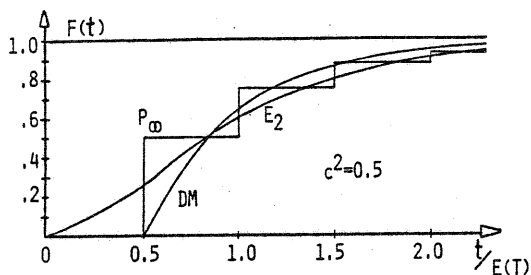


Fig. 3.4 Three Distribution Functions with Same First Two Moments

As an example, consider fig. 3.4 with 3 different d.f.'s $F(t)$, having same mean $E(T)$ and variation coefficient c .

Compared with the E_2 d.f. are the following types

- a step function, here called P_∞ , having geometrically distributed heights
- a function type DM, describing the sum of a constant and of a negative-exponentially distributed part (allowing to adopt all variation coefficients $0 \leq c \leq 1$, cf. [17]).

Figs 3.5a,b show as an example simulation results for the mean waiting time and the probability of waiting for all 9 systems $F_A/F_H/1$, having these 3 types E_2, P_∞, DM as types of interarrival distribution F_A and/or service time distribution F_H .

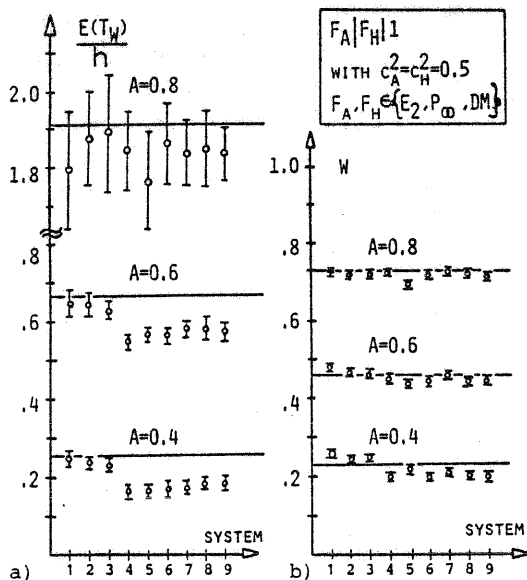


Fig. 3.5a,b Mean Waiting Times $E(T_W)$ and Probabilities of Waiting W for Different Systems $F_A/F_H/1$ ($1=E_2/E_2/1, 2=E_2/P_\infty/1, 3=E_2/DM/1, 4=P_\infty/E_2/1, 5=P_\infty/P_\infty/1, 6=P_\infty/DM/1, 7=DM/E_2/1, 8=DM/P_\infty/1, 9=DM/DM/1$)

In addition, the approximation results according to (1.3) and (1.4) are shown as horizontal lines.

As it can be seen, the errors in the probability of waiting are very low, whereas the mean waiting time sometimes is overestimated for these special types of d.f.'s.

There has only one critical case been observed, where the 2-moments approximation is inefficient even for estimates. This has been systems with $c_A^2 \approx 0$ and (degenerated) 2-step d.f.'s for the service times, allowing also service times $T_H=0$. For such $D/P_2/1$ systems for

$$A \leq 1/(1+c_H^2)$$

no waiting at all occurs, which naturally cannot be reflected in a 2-moments approximation. For this very special and degenerated case the approximations cannot be recommended generally.

4 SYSTEMS WITH BATCH ARRIVALS

In this chapter it is shown that the formulae for the mean waiting time and the probability of waiting also can be applied to systems with batch arrivals by defining equivalent single arrival systems.

4.1 GENERAL REMARKS

In communications systems, e.g., often the arrivals at centralized units occur in batches or groups of arbitrary size K , now being assumed to be an independent random variable. Let p_K ($K=0,1,\dots$) be the probability that at a possible arrival instant of a batch (characterized by a d.f. for the interarrival times T_{AB} of batches) a batch of size K arrives. For reasons of generality, let also 'batches' of size $K=0$ be included. This is useful especially in so-called sampled systems, where the batches only arrive at equidistantly distributed clock instants (see e.g. [13]).

Now, by standard calculations, general relations between characteristic values of the first request of a batch (indexed by an additional 1) and an arbitrary member of it can be established. With respect to the mean waiting time it results

$$E(T_W) = E(T_{W1}) + \left(\frac{\text{Var}(K)}{E(K)} + E(K) - 1 \right) \cdot \frac{h}{2} \quad (4.1)$$

This formula can be found e.g. in [10], [13] and [14].

With respect to the probability of waiting, it can be simply stated that with the probability q that an arbitrary request is the first request within its batch

$$q = \frac{1}{E(K|K>0)} = \frac{1-p_0}{E(K)} \quad (4.2)$$

it holds (zero service times excluded)

$$W = q \cdot W_1 + (1-q) \cdot 1 \quad (4.3)$$

This means

$$W = 1 - \frac{(1-p_0)}{E(K)} \cdot (1-W_1) \quad (4.4)$$

The two formulae (4.1) and (4.4) allow to determine values for single requests, if the corresponding ones of the first of a batch are known, see 4.2.

4.2 EQUIVALENT SYSTEMS WITH SINGLE ARRIVALS

The waiting time of the first request of a batch can be calculated by considering an equivalent system, defining a whole batch (>0) as a 'super-request' [2]. Then the new interarrival times are the times between the arrivals of batches >0 , the service times are the total times to serve a whole batch (>0).

If we denote the characteristics of the equivalent system by an additional asterisk *, then

$$E(T_H^*) = E(K|K>0) \cdot E(T_H) \quad (4.5)$$

With

$$\text{Var}(T_H^*) = E(T_H)^2 \text{Var}(K|K>0) + E(K|K>0) \text{Var}(T_H) \quad (4.6)$$

it can be shown that the resulting equivalent (squared) variation coefficient of service time is

$$c_H^{*2} = \frac{1-p_0}{E(K)} \cdot \left[\frac{\text{Var}(K)}{E(K)} + c_H^2 \right] - p_0 \quad (4.7)$$

Let now the arrival process be characterized by

- the interarrival times T_{AB} of batches (of size $K \geq 0$, i.e. the times between successive closings of an 'input switch', with mean $E(T_{AB})$ and variation coefficient c_{AB})
- the batch size probabilities p_K ($K \geq 0$)

Since the equivalent system is based on batches of size > 0 , it holds

$$E(T_A^*) = \frac{E(T_{AB})}{1-p_0} \quad (4.8)$$

and

$$c_A^{*2} = (1-p_0) \cdot c_{AB}^2 + p_0 \quad (4.9)$$

Summarizing, the equivalent system is characterized by

- the mean service time

$$E(T_H^*) = \frac{E(K)}{1-p_0} \cdot E(T_H) \quad (4.10)$$

- the same traffic offered as in the original batch input system

$$A^* = \frac{E(T_H^*)}{E(T_A^*)} = \frac{E(K) \cdot E(T_H)}{E(T_{AB})} = A \quad (4.11)$$

- and the variation coefficients c_A^* and c_H^* according to (4.9) and (4.7).

These values have to be calculated before using the GI/G/1 approximation formulae for the equivalent single arrival system.

4.3 EXAMPLES FOR BATCH ARRIVAL SYSTEMS

Since for M/G/1 systems the formulae (1.3) and (1.4) are exact, all batch arrival systems rendering equivalent M/G/1 systems will be calculated exactly.

This holds for negative exponentially distributed interarrival times between batches of arbitrary size $K \geq 0$ and distribution (i.e. $c_{AB}^2 = 1$, p_0 arbitrary). If $p_0 > 0$, this will also result in negative exponentially distributed interarrival times between batches of size $K > 0$, with mean increased by $1/(1-p_0)$ (compound Poisson arrival process).

Combining (4.1) with (1.3) and (4.7) to (4.11), it results for negative exponentially distributed interarrival times between batches of arbitrary size $K \geq 0$

$$E(T_W) = \frac{h}{2(1-A)} \cdot \left[A(1+c_H^{*2}) + \frac{\text{Var}(K)}{E(K)} + E(K) - 1 \right] \quad (4.12)$$

For the probability of waiting from (4.4) with $W_1 = A$

$$W = 1 - \frac{1-p_0}{E(K)} \cdot (1-A) \quad (4.13)$$

Eq. (4.12) is identical with a result of GAVER [20], who derived the generating function of the state probabilities of systems with compound Poisson arrival processes.

Since the formulae are exact for compound Poisson arrival processes, for validity purposes, it remains to show examples, where the times between two batch arrivals are not negative exponentially distributed.

To select an extreme but nevertheless important case, equidistantly distributed interarrival times are selected, i.e. so-called sampled systems. These systems may be conceived as having an input switch being closed periodically.

Geometrically distributed batch sizes:

The first example has geometrically distributed batch sizes, with

$$p_K = \begin{cases} 0 & K=0 \\ (1-p) \cdot p^{K-1} & K>0 \end{cases} \quad (4.14)$$

Thus

$$E(K) = \frac{1}{1-p} \quad \text{and} \quad \text{Var}(K) = \frac{p}{(1-p)^2} \quad (4.15a, b)$$

and with (4.7)

$$c_H^{*2} = (1-p) \cdot c_H^2 + p \quad (4.16)$$

Since $p_0 = 0$, $c_A^{*2} = c_{AB}^2 = 0$. The probability p could be interpreted as the probability, that the next single request belongs to the same batch (i.e. $p = F_A(0)$ in case of normal single arrival GI/G/1 notation).

The following examples adopt $p=1/3$, rendering $E(K)=1.5$ and $\text{Var}(K)=0.75$, also $E(T_H^*)=1.5 \cdot h$. With (4.1) and (4.4) approximate values for $E(T_W)$ and W can be calculated, here done for different service time d.f.'s, see figs 4.1 and 4.2.

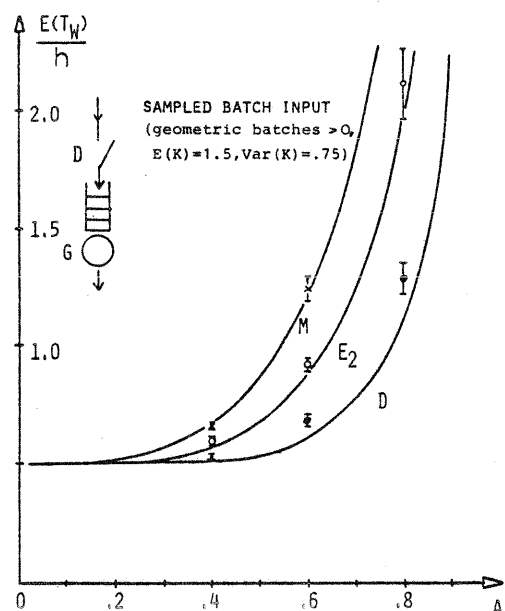


Fig. 4.1 Mean Waiting Times for Sampled Batch Arrival Systems

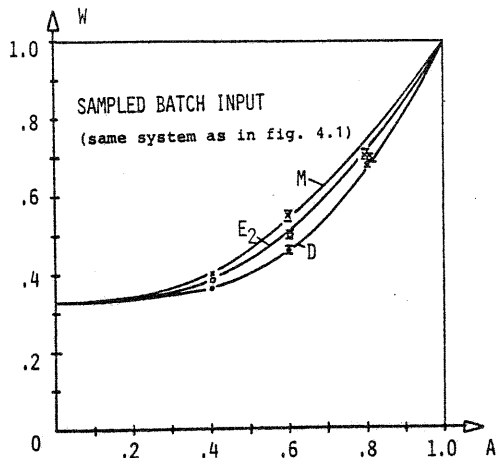


Fig. 4.2 Probability of Waiting for Sampled Batch Arrival Systems

Remember, that these results have been obtained with very low calculation effort.

Poisson distributed batch sizes:

If the single requests arrive in front of the sample switch with negative exponentially distributed interarrival times, the distribution of the batch sizes is Poisson, having $\text{Var}(K)/E(K)=1$.

For the examples shown in fig 4.3 the clock time was chosen to be equal to the mean service time ($E(T_{AB})=E(T_H)$), rendering $A=E(K)$ and thus determining the P_0 values via the Poisson distribution. Then from (4.9) $c_A^{*2}=P_0$, whereas c_H^{*2} is determined with (4.7).

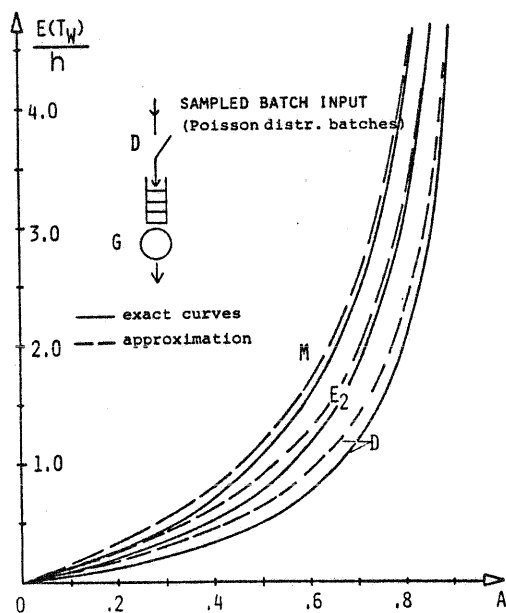


Fig. 4.3 Mean Waiting Times for Sampled Batch Arrival Systems (Poisson distributed batch sizes)

The solid lines are exact results, the curves for M and E_2 have been calculated with a program of WEISSCHUH and WIZGALL [15]. For constant service times it simply holds (LANGENBACH-BELZ [13])

$$E(T_W) = \frac{A}{2(1-A)} \cdot h \quad (4.17)$$

The associated curves for the probabilities of waiting W nearly have been identical for all 3 d.f.'s with the line $W=A$, both simulation and approximate calculation. Therefore they are not shown here.

5 DETERMINATION OF FURTHER TRAFFIC VALUES

Up to this chapter only the expected waiting time $E(T_W)$ and the probability of waiting W have been considered, both for single arrival and batch arrival systems. Based on relations for a wide class of stationary single server systems (MARSHALL [9], RICE [16]), the approximation formulae (1.3) and (1.4) can also be used to calculate approximately further systems characteristics.

5.1 OUTPUT VARIANCE

The following relation between the variance of the output process and the mean waiting time has been found in [9]

$$\text{Var}(T_D) = \text{Var}(T_A) + 2\text{Var}(T_H) - \frac{2}{\lambda}(1-A) \cdot E(T_W) \quad (5.1)$$

Using coefficients of variation and $E(T_D)=E(T_A)$

$$c_D^2 = c_A^2 + 2A^2 c_H^2 - 2A(1-A) \cdot \frac{E(T_W)}{h} \quad (5.2)$$

With $E(T_W)$ according to (1.3) this is a simple and explicit approximation for the variation coefficient c_D of the output or departure process.

This formula specializes for Poisson input to the well-known form [21]

$$c_D^2 = 1 - A^2(1 - c_H^2) \quad (5.3)$$

KÜHN [18] has used (5.2) for the output processes in general queueing networks and took over (1.3) (see also output figures in [18]).

5.2 IDLE AND BUSY PERIOD

According to an exact formula from [16]

$$(1-W) \cdot E(T_{IP}) = \frac{1}{\lambda} - h \quad (5.4)$$

also the mean time $E(T_{IP})$ of an idle period can be approximated, using (1.4):

$$E(T_{IP}) = \frac{1-A}{A(1-W)} \cdot h \quad (5.5)$$

The next relation (from [9]) includes the second moment of the idle period $E(T_{IP}^2)$, which therefore also could be calculated using the approximations derived:

$$E(T_W) = \frac{c_A^2 + A^2 c_H^2 + (1-A)^2}{2A(1-A)} \cdot h - \frac{E(T_{IP}^2)}{2E(T_{IP})} \quad (5.6)$$

There is also a possibility to approximate the mean length of a busy period T_{BP} via the relation [16]

$$E(T_{BP}) = \frac{1-P_0}{P_0} \cdot E(T_{IP}) \quad (5.7)$$

where P_0 is the absolute probability of an empty system (also allowing finite waiting storage). For pure waiting systems, as being considered here, $P_0=1-A$, such that with (5.5)

$$E(T_{BP}) = \frac{1}{1-W} \cdot h \quad (5.8)$$

In RIORDAN [2] eqns (5.4) and (5.7) are derived without special assumptions concerning the input process. Thus the equations for the mean idle period (5.5) and the mean busy period (5.8) are also applicable to batch input systems with W according to (4.4).

6 SUMMARY AND CONCLUSION

For the general single server system GI/G/1 simple 2-moments approximations have been derived for the mean waiting time and the probability of waiting. The formulae have been the gap between many complex exact results and a quick numerical calculation for engineering purposes. With the restriction to 2 moments, the formulae and application should be quick and simple, naturally thus inducing certain errors. By using explicit exact formulae as frames and by systematically utilizing the freedoms within those, the heuristic formulae development could be done more or less systematically.

The accuracy of the formulae has been investigated and proved to be very useful within a wide range of applications or traffic assumptions, also easily including systems with batch arrivals.

Furthermore it may be noted, that there exist simple possibilities to include also GI/G/1 systems with probabilistic feedback or even batch service systems, not being considered here.

In addition, also simple formulae are available for the variances of the output processes as well as the mean values for the idle and busy periods.

The usefulness of the approximation of the mean waiting times and of the output variances has been already demonstrated in a companion paper by KÜHN [18] in context with queueing networks. Also tables including these approximations are provided [19].

It is hoped that these heuristic approximations, which cannot and will not replace GI/G/1 investigations with detailed reflection of the inter-arrival and service time d.f.'s, will be helpful for the traffic engineer to obtain very quickly and simply useful estimates for the delay in his special single server models.

ACKNOWLEDGEMENT

The authors wish to thank G. Kröner for his valuable assistance during formulae development and his programming support. They are also grateful to the Federal Ministry of Research and Technology (BMFT) of the Fed. Rep. of Germany and to the German Research Association (DFG) for supporting this work.

REFERENCES

- [1] SYSKI, R. Introduction to Congestion Theory in Telephone Systems. Oliver & Boyd, London/Edinburgh, 1960
- [2] COHEN, J.W. The Single Server Queue. North Holland, Amsterdam/London, 1969
- [3] KLEINROCK, L. Queueing Systems, Vol I: Theory. JWS, NY, London, 1975
- [4] LINDLEY, D.V. The Theory of Queues with a Single server. Proc. Camb. Phil. Soc. 48 (1952), 277-289
- [5] KENDALL, D.G. Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. Ann. Math. Stat. 24 (1953), 338-354
- [6] GREENBERG, I. Distribution-Free Analysis of M/G/1 and G/M/1 Queues. J. Op. Res. 21 (1973), 629-635
- [7] KINGMAN, J.F.C. Some Inequalities for the GI/G/1 Queue. Biometrika, 49 (1962), 315-324
- [8] BLOOMFIELD, P., COX, D.R. A Low Traffic Approximation for Queues. J. Appl. Prob. 9 (1972), 832-840
- [9] MARSHALL, K.T. Some Inequalities in Queueing. J. Op. Res. 16 (1968), 651-668
- [10] MARSHALL, K.T. Bounds for Some Generalizations of the GI/G/1 Queue. J. Op. Res. 16 (1968), 841-848
- [11] KOBAYASHI, H. Bounds for the Waiting Time in Queueing Systems. IBM Research Rpt RC4718, Yorktown Heights, February 1974
- [12] HEYMAN, D.P. A Diffusion Model Approximation for the GI/G/1 Queue in Heavy Traffic. BSTJ 54 (1975), 1637-1646
- [13] LANGENBACH-BELZ, M. Sampled Queueing Systems. Proc. Symp. on Computer Communications Networks & Teletraffic, Polytechnic Press, Brooklyn, New York, 1972
- [14] BURKE, P.J. Delays in Single Server Queues with Batch Input. J. Op. Res. 23 (1975), 830-833
- [15] WEISSCHUH, H., WIZGALL, M. Investigations of the Traffic Behaviour of the Common Control in SPC Switching Systems. 8th ITC, Melbourne, 1976 (this volume)
- [16] RICE, S.O. Single Server Systems-I. Relations between some Averages. BSTJ 41 (1962), 269-278
- [17] KRÄMER, W. Investigations of Systems with Queues in Series (Untersuchung von Systemen mit seriellem Warten) Dissertation University of Stuttgart 1975 and 22nd Report on Studies in Congestion Theory, Institute of Switching and Data Technics 1975 (In German)
- [18] KÜHN, P. Analysis of Complex Queueing Networks by Decomposition. 8th ITC, Melbourne, 1976 (this volume)
- [19] KÜHN, P. Tables of Delay Systems. Institute of Switching and Data Technics, University of Stuttgart, 1976
- [20] GAVER, D.P. Jr. Imbedded Markov Chain Analysis of a Waiting-Line Process in Continuous Time. Ann. Math. Stat. 30 (1959), 698-720
- [21] MAKINO, T. On a Study of Output Distribution. J. Op. Res. Japan 8 (1966), 109-133
- [22] RIORDAN, J. Stochastic Service Systems. JWS, NY, London, 1962