QUEUEING SYSTEMS WITH PRIORITIES AND INTERVALS OF SATURATION

Werner Wagner

Badische Anilin- und Soda-Fabrik A.G.

Ludwigshafen am Rhein, Federal Republic of Germany

ABSTRACT

If the traffic offered to a delay system is greater than the number n of trunks, the mean number of calls which are present in the delay system grows continuously. When considering a delay system with non-preemptive priorities, the situation may be multifarious. If the traffic offered summed up for the highest priorities is less than n, the calls of these priority classes 1,2,...l form a stationary traffic. Calls of lower priority than the limit priority l (classes l+1,...m) are added to the queue of waiting calls. They cause the waiting traffic to grow. In order to study the expectation values for the number of waiting calls of priority classes 1...k, it is assumed that all n trunks are continuously occupied by calls of priority classes 1...k. POISSON arrival of calls and exponential distribution of holding times are assumed.

For priorities lower than l, it follows that the mean number of waiting k-calls and their mean waiting times increase linearly with the time, until the traffic offered decreases to a value which is less than n. Then the "peak" of waiting calls will be reduced gradually. Using the same assumption we find: At first the mean number of waiting 1-calls linearly decreases. This is done consecutively for all priorities. The mean waiting times decrease linearly, too. When all expectation values of waiting calls have reached the stationary values which are valid for intervals, in which all trunks are busy, a second ebbing period follows. It is assumed that the mean number of calls, which are present in the system, decreases exponentially to the ordinary stationary values.

It must be admitted that this method can give but a rough approximation both of the time-dependent expectation values of the number of occupied waiting places, and of the mean waiting times. One special result, confirmed by simulation tests, is to be mentioned: A peak of waiting calls exists. In the following interval the total number of calls decreases. Nevertheless, the mean number of calls of low priorities increases, and decreases only after a period of growing.

This investigation was performed at the University Stuttgart Germany when the author was with the Institute of Switching and Data Techniques. I should like to thank Professor Dr.-Ing. A. Lotze for interesting discussions and the Deutsche Forschungsgemeinschaft (German Research Society) for the support granted. Furthermore I appreciate the kind permission of Direktor Dr. Thurn, BASF M-Gruppe, to attend the 6th ITC and to present this paper.

1. INTRODUCTION

1.1 Intervals of saturation

In delay systems each call, which cannot occupy an outlet of the connecting system at once, is put to a waiting place. It is waiting until it can occupy a trunk. No waiting call leaves the queue before service. When the mean number of calls, which arrive during a unit of time, surpasses the service capacity of the delay system, then the mean number of waiting calls grows continuously. The state probabilities of the system and the expectation value of the waiting times are no longer constant and stationary, but they are time-dependent. This situation was called "saturation" /5/.

In real telephone connecting systems, the traffic offered does not remain constant during long periods. Intervals with a high mean rate of incoming calls are followed by periods of smaller traffic offered. Therefore the queue does not grow beyond all limits, but it grows only as long as the traffic offered is large, and then it begins to decrease.

1.2 Delay system without priorities

In order to give the time-dependent expectation values, it is necessary to calculate the time-dependent probabilities P(x,t) that the total number of calls in the system will be x at time t. Exact solutions were given for the case of POISSON input and exponential holding times. The formulae include series or integrals on BESSEL functions /1,2,9,11/ or integrals on trigonometric or TSCHEBYSCHEFF functions /4,9/. Most of the solutions consider 1 outgoing trunk /1,2,11/, others include the case of n trunks and full availability /4/. Some of these formulae are well suited for the growing queue, others are adapted for the decreasing queue. H. STÖRMER investigated the mean waiting times and the probabilities of delay, which are valid for the first call, the second etc., finally for the i-th call /15/. The assumptions of this study were POISSON input, constant holding times and 1 outgoing trunk.

An approximation method was introduced by D.R. COX and W.L. SMITH for 1 outgoing trunk and general input and service time distributions /3/. For the growing as well as for the decreasing waiting traffic they found a linear time-dependency. A similar method was described by J.M. ROBINS /10/. The corresponding formulae for POISSON input and exponential service times were deduced as limiting formulae for mean and variance by A.B. CLARKE and by other authors, too /2,11/.

1.3 Delay system with priorities

Demands which are connected from inlets to outlets of an exchange are not of the same importance. On account of their different importance and urgency, the calls may be arranged in priority classes. Calls of higher priority take precedence when waiting. The operation of the delay system is specified by the following rules:

1. Calls having occupied a trunk are never interrupted, not even if calls of higher priority must wait (nonpre-emptive priority).
2. Calls of higher priority queue up in front of calls of lower priority.
3. The order of service is "first come - first served" or strict queueing within each priority class.

A first step to use an approximation method for the mean number of waiting calls was done by F.P. RANDAZZO /8/. He calculated each final value at the end of a saturation period and used each of them, as if it were constant during the whole period.

In this paper, the time-dependent expectation values of the number of waiting calls and of their waiting times, as well as their variances, are deduced for a delay system with nonpre-emptive priorities. The method introduced by D.R. COX and W.L. SMITH /3/ is used to obtain formulae for the growing and the decreasing queue. Examples are given in order to compare this approximation to results of simulation tests.

1.4 Holding times and service capacity

All calls belonging to any priority class have the same exponential probability distribution of holding times (respectively service times). The holding times are mutually independent random variables. The mean holding time is to be unity. Thus, the time t is a normalized variable throughout this investigation.

The trunk group of n trunks is fully available. The number of waiting places is as large as necessary, so that at any moment and at any state of the delay system each new call can occupy a free waiting place, i.e. a pure delay system is considered, and no call is lost. No call leaves the queue without service; there are no defections.

1.5 The priorities and the traffic offered

The priority classes are usually numbered k=1,2,...m according to decreasing urgency. Higher priorities correspond to smaller indices k. The system considered has m priority classes. In the sequel calls of priority k will be termed <u>k-calls</u> /5/. The average number of k-calls arriving per unit of time is A_k, i.e. the traffic offered of priority k. For each priority a distinct POISSON input is assumed.

$$A(\leq k) = A_1 + A_2 + \dots + A_k \qquad (1)$$

2. THE PERIOD OF SATURATION AND THE INCREASING NUMBER OF WAITING CALLS

2.1 The basic assumption

Let us suppose that all n trunks of the trunk group will be busy at time t=0. Moreover it is supposed that, from that time t=0, the traffic offered $A(\leq k)$ of the priorities 1,2,...k will be much larger than the mean number of occupations which end during a mean holding time.

$$A(\leq k) \gg n \qquad (2)$$

Then the probability, that at any moment after t=0 a trunk is idle, is very small and may be neglected.

> The approximation is based on the assumption that during the whole interval, when $A(\leq k)>n$, all n trunks are always occupied by calls of the priorities 1,2,...k. (3)

The approximation disregards, that a trunk might be free, or that a trunk might be occupied by a call of lower priority, i.e. of any of the priority classes k+1, k+2, ... m. The growing queue is treated in detail because the same formulae are valid for the decreasing queue, too.

2.2 The rate of calls arriving and of occupations ending

During an interval of length d, $\underline{r}$ new calls of the priorities 1,2,...k arrive and s occupations end. (Underlining indicates random variables.) The <u>expectation</u> values and the <u>variances</u> are

$$\text{EXP}\ \underline{r} = A(\leq k)\cdot d \qquad (4)$$

$$\text{VAR}\ \underline{r} = A(\leq k)\cdot d \qquad (5)$$

$$\text{EXP}\ \underline{s} = n\cdot d \qquad (6)$$

$$\text{VAR}\ \underline{s} = n\cdot d \qquad (7)$$

2.3 The growing number of waiting calls

The random number of calls of the priorities 1,2,...k, which are waiting at the time t, is called $\underline{z}(t)$. From (3) it follows

$$\underline{z}(t) = \underline{z}(t-d) + \underline{r} - \underline{s} \qquad (8)$$

where $\underline{z}(t-d)$ is the number of waiting calls at the beginning of the interval d. The waiting traffic $\Omega(\leq k;t)$ of priorities 1,2,...k at time t is given by the expectation value EXP $\underline{z}(0)$ at time 0 and by EXP $\underline{r}$ and EXP $\underline{s}$ for an Interval of length t:

$$\text{EXP}\ \underline{z}(t) = \text{EXP}\ \underline{z}(0) + A(\leq k)\cdot t - n\cdot t \qquad (9)$$

The random variables $\underline{r}$, $\underline{s}$, and $\underline{z}(0)$ are independent of each other.

$$\text{VAR}\ \underline{z}(t) = \text{VAR}\ \underline{z}(0) + A(\leq k)\cdot t + n\cdot t \qquad (10)$$

In order to check whether the waiting traffic grows <u>linearly</u> and whether its standard deviation $\sqrt{\text{VAR}\ \underline{z}(t)}$ increases parabolicly, we compare the values calculated from eqs. (9) and (10) to the results of a simulation, cf. fig. 1. The programme written by Mrs. I. KUHNLE and the author is based on the time-true model /17/ or event-by-event simulation /16/. It is written in ALGOL. With the AEG-Telefunken computer TR4, it achieves a test speed of about 28 000 to 90 000 calls per hour.

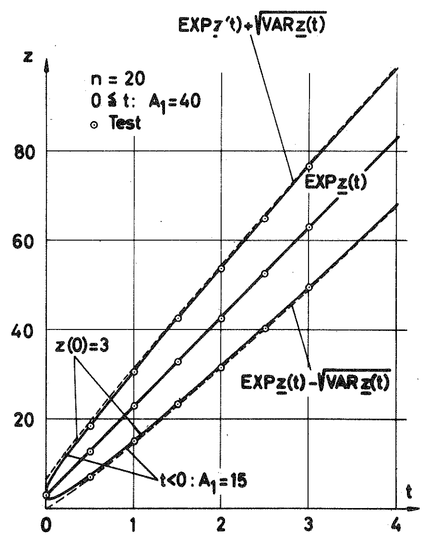

Fig. 1 The mean number of waiting 1-calls at the time t, calculated for a traffic offered A_1=40 to n=20 trunks. The initial conditions: a) z(0)=3 waiting 1-calls (continuous curves), b) stationary values, when A_1=15 for t<0 (dotted curves). The results of a simulation run are indicated by points encircled.

2.4 The limit priority l

If $A(\leq k) > n$, the mean number of waiting calls, comprehending all calls of priorities from 1 up to k, is not constant, but is growing. But for calls of priorities from 1 up to l the stochastic traffic process reaches

$$\text{stationarity, if } \quad A(\leq l) < n \quad \text{but} \quad A(\leq l+1) \geq n \qquad (11)$$

though the total traffic process of all calls of arbitrary priority classes does not reach stationarity /e.g. 2,10,11/. The increase of queue length results, on the average, only from less urgent calls of priorities l+1, l+2,...m. But the mean numbers of calls (k=1,2,...l) which are present in the trunk group or in the queue line are independent of time.

2.5 The waiting calls of priorities 1,2,...l

The mean number of waiting calls of priorities 1,2,...k is

$$\Omega(\leq k) = \frac{A(\leq k)}{n-A(\leq k)} \cdot E \qquad \text{for } k=1,2,\ldots l \text{ and } A(\leq l)<n \qquad (12)$$

The probability of delay and of blocking is

$$E = 1 \qquad \text{because of } A(\leq m)>n \qquad (13)$$

The probabilities that there are z waiting calls of priorities 1,2,...k

$$\text{PROB}(\underline{z}=z) = \left[\frac{A(\leq k)}{n}\right]^z \cdot \left[1 - \frac{A(\leq k)}{n}\right] \cdot E \qquad (14)$$

form a geometric probability distribution. Therefore the variance of the random number of waiting calls of priorities 1,2,...k is

$$\text{VAR } \underline{z} = \frac{n+A(\leq k)}{[n-A(\leq k)]^2} \cdot A(\leq k) \cdot E - \left[\frac{A(\leq k)}{n-A(\leq k)}\right]^2 \cdot E^2 \qquad (15)$$

This equation was used to calculate the initial variance for t=0 and $A_1=15$ in fig. 1 (case b). For E=1, VAR $\underline{z}$ turns out to be

$$\text{VAR } \underline{z} = \frac{n \cdot A(\leq k)}{[n-A(\leq k)]^2} \qquad \text{for } k=1,2,\ldots l \text{ and } A(\leq l)<n \qquad (16)$$

In order to dimension a delay system and to determine the number of waiting places which are necessary, the interesting quantity is given by the number of waiting calls summed up from priority 1 to priority k. On the other hand, the grade of service of a queueing system is given by the waiting times, which must be calculated distinctly for each priority. For that, we quote the mean number of waiting k-calls, if stationarity is reached:

$$\Omega_k = \frac{n \cdot A_k}{[n-A(\leq k)] \cdot [n-A(\leq k-1)]} \cdot E \qquad (17)$$

2.6 The waiting calls of priorities l+1, l+2, ... m

The waiting traffic $\Omega(\leq k;t)$ at the time t is given by eq. (9) for k=l+1 and by eq. (12) for k=l. Therefore, the waiting traffic of priority l+1 seperately is

$$\begin{aligned} \Omega_{l+1}(t) &= \Omega(\leq l+1;t) - \Omega(\leq l) \\ &= \Omega_{l+1}(0) + [A(\leq l+1)-n] \cdot t \end{aligned} \qquad (18)$$

The waiting traffic of priorities lower than l+1 is, cf. eqs. (1) and (9):

$$\Omega_k(t) = \Omega(\leq k;t) - \Omega(\leq k-1;t) = \Omega_k(0) + A_k \cdot t \qquad (19)$$

All arriving k-calls (k=l+2,...m) increase the queue of waiting k-calls; on the average no one of these calls is served during the period of saturation.

2.7 The mean waiting time of 1-calls

At first we investigate the special case that there exists no limit priority (l=0) – or, stated otherwise, $A_1>n$. In order to find the mean waiting time, we regard one single 1-call which arrives just at the time t, cf. fig. 2.

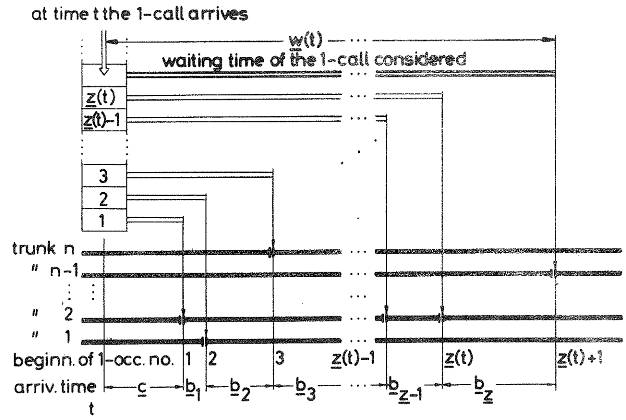

Fig. 2 A time chart which shows the random waiting time of one single 1-call

The time interval from an arbitrary starting point to the next time, when any of the occupations in the trunk group ends, is the random time $\underline{c}$. The time interval between two consecutive ends of occupations with numbers j and j+1 in the trunk group is $\underline{b}_j$. Thus, the waiting time of the 1-call considered is

$$\underline{w}(t) = \underline{c} + \underline{b}_1 + \underline{b}_2 + \ldots + \underline{b}_{z(t)-1} + \underline{b}_{z(t)} \qquad (20)$$

The sum of the random variables $\underline{b}_j$ contains a random number $\underline{z}$ of terms. Its expectation value is EXP $\underline{z}$ · EXP $\underline{b}$. If the holding times are exponentially distributed, time intervals $\underline{b}_j$ and $\underline{c}$ follow identical exponential distributions.

$$\text{EXP } \underline{b} = \text{EXP } \underline{c} = \frac{1}{n} \qquad (21)$$

$$\text{VAR } \underline{b} = \text{VAR } \underline{c} = \frac{1}{n^2} \qquad (22)$$

The mean waiting time of an 1-call which arrives at the time t is

$$\text{EXP } \underline{w}(t) = t_{w1}(t) = \frac{\text{EXP } \underline{z}(t) + 1}{n} \qquad (23)$$

This formula was used by A. LOTZE to derive the mean waiting time /6,7/. With eq. (9) for k=1 we get

$$t_{w1}(t) = \frac{\text{EXP } \underline{z}(0)+1}{n} + \left(\frac{A_1}{n} - 1\right) \cdot t \qquad (24)$$

2.8 The variance of waiting times of 1-calls

In order to determine the variance of the waiting times, one starts with the assumption, that the sum (20) consists of exactly z terms. The z random intervals between ends of occupations are independent from each other. Therefore, the variance is z·VAR $\underline{b}$. Then, the variance of the sum with a random number $\underline{z}$ of terms is found to be

$$\text{VAR} \sum_{j=1}^{\underline{z}} \underline{b}_j = \text{EXP } \underline{z} \cdot \text{VAR } \underline{b} + \text{VAR } \underline{z} \cdot [\text{EXP } \underline{b}]^2 \qquad (25)$$

Using this formula (25) together with eq. (20) the variance of the waiting times of 1-calls arriving at the time t is

$$\text{VAR } \underline{w}(t) = \frac{1}{n^2} \cdot [\text{EXP } \underline{z}(0) + 1 + \text{VAR } \underline{z}(0) + 2A_1 \cdot t] \qquad \text{for } A_1>n \qquad (26)$$

Again we compare the calculated values for the mean waiting time $t_{w1}(t)$ and its standard deviation $\sqrt{VAR\ \underline{w}(t)}$ to the results of the same simulation, which had been used for figure 1.

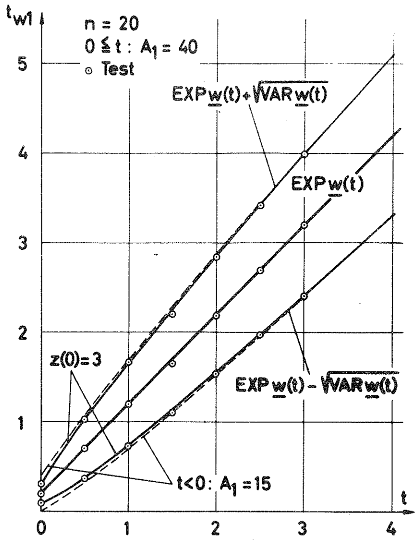

Fig. 3 The mean waiting time and its standard deviation. See the notes given above at fig. 1.

2.9 The mean waiting time of (l+1)-calls

Now we consider the case that the stochastic traffic process is stationary for priorities 1,2,...l. The mean waiting time during intervals of fully occupied trunk group is obtained from $\Omega_k = A_k \cdot t_{wk}$ with eq. (17)

$$EXP\ \underline{w} = t_{wk} = \frac{n}{[n-A(\leq k)]\,[n-A(\leq k-1)]} \qquad (27)$$

for k=1,2,...l and $A(\leq l)<n$

On the average, all calls of priorities 1,2, ...l will occupy the trunks. During gaps between these occupations, some of the (l+1)-calls will seize a trunk. Now, the mean waiting time for priority k=l+1 is deduced. On the average $A_{l+1}\cdot d$ calls of priority l+1 arrive between time 0 and d. The increase of queue length is on the average, cf. eq. (18)

$$\Delta\Omega_{l+1}(d) = \Omega_{l+1}(d) - \Omega_{l+1}(0) = [A(\leq l)+A_{l+1}-n]\cdot d \qquad (28)$$

On the average $A_{l+1}\cdot d - \Delta\Omega_{l+1}(d)$ calls of priority l+1 have occupied a trunk between time 0 and d.

$$A_{l+1}\cdot d - \Delta\Omega_{l+1}(d) = [n-A(\leq l)]\cdot d$$

As the interval between ends of occupations is 1/n, on the average, n·d new occupations begin between 0 and d. During this interval, $A(\leq l)\cdot d$ occupations of priorities 1,2,...l start; the intervals between beginning times of these occupations are $1/A(\leq l)$, on the average. The beginning times of the $[n-A(\leq l)]\cdot d$ occupations of (l+1)-calls have an average distance of $1/[n-A(\leq l)]$.

The interval from an arbitrary starting point t to the first end of an occupation, when this trunk is seized by a waiting (l+1)-call, is called $\underline{c}$; its unknown expectation value is EXP $\underline{c}$. From fig. 4, it may be seen, that the random waiting time $\underline{w}(t)$ of a (l+1)-call consists of the random interval $\underline{c}$, and of $\Omega_{l+1}(t)$ intervals between beginning times of (l+1)-occupations. Hence

$$EXP\ \underline{w}(t) = t_{w,l+1}(t) = EXP\ \underline{c} + \frac{\Omega_{l+1}(t)}{n-A(\leq l)} \qquad (29)$$

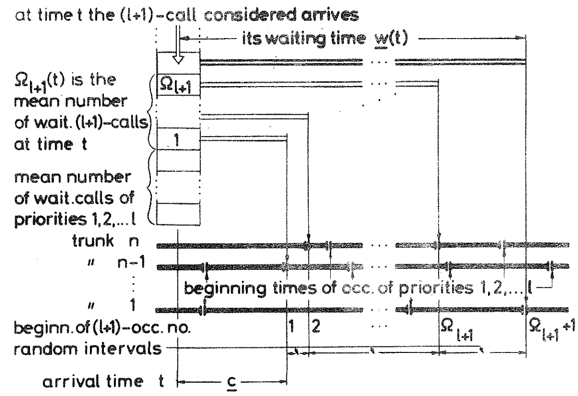

Fig. 4 A time chart which shows the random waiting time of a (l+1)-call

In order to determine the unknown value EXP $\underline{c}$, we investigate the stationary waiting time $t_{w,l+1}$ during a period of blocking, which is given by eq. (27). Introducing eq. (17) with E=1 into eq. (29), EXP $\underline{c}$ turns out to be

$$EXP\ \underline{c} = t_{w,l+1} - \frac{\Omega_{l+1}}{n-A(\leq l)} = \frac{n}{[n-A(\leq l)]^2} \qquad (30)$$

For the purpose of calculating the time-dependent mean waiting time, we introduce eqs. (9) and (30) into eq. (29)

$$t_{w,l+1}(t) = \frac{n}{[n-A(\leq l)]^2} + \frac{\Omega_{l+1}(0)}{n-A(\leq l)} + \frac{A(\leq l+1)-n}{n-A(\leq l)}\cdot t \qquad (31)$$

2.10 The waiting times of calls of priorities l+2, l+3, ... m

Because all trunks are continuously occupied by calls of priorities 1,2,...l+1, on the average, all calls of priorities l+2, l+3, ... m must wait as long as the traffic offered surpasses the service capacity: $A(\leq l+1)>n$. Then, the waiting times continue into the period when the peak of waiting calls will be reduced gradually. The time-dependent expectation values during the interval of cutting down the peak are determined in the next section.

3. THE PERIOD OF CUTTING DOWN A PEAK OF WAITING CALLS

3.1 The decreasing number of waiting 1-calls

The origin t=0 of the time scale is shifted to the beginning of the period considered. At the time regarded t=0 many 1-calls are waiting in front of the totally occupied trunk group. It may be possible, that all $n+\Omega_1(0)$ occupations will be very short and that no new 1-call will arrive during the duration of these occupations. But we consider mean holding times of occupations to find an approximation for the time-dependent expectation value EXP $\underline{z}(t)$. Again, it is assumed that all n trunks are continuously occupied by 1-calls. Then, the mean number of waiting 1-calls and its variance is given by eqs. (9,10,24,26). For $A(\leq 1)=A_1<n$, it can be seen, that the mean number of waiting 1-calls and the mean waiting time decrease linearly whereas the variances increase.

During the period, when the queue decreases, all arriving 1-calls must wait. The period of cutting down the peak of waiting 1-calls ends, when the influence of the high initial value $\Omega_1(0)$ has disappeared and when EXP $\underline{z}(t)$ has decreased to the stationary value EXP $\underline{z}$, which is given by eq. (12) with E=1 and k=1. The time-dependent waiting traffic reaches this value, which is valid during blocking periods, at the time T_1. Therefore, using eqs. (9) and (12)

$$\text{EXP } \underline{z}(T_1) = \Omega_1(T_1) = \Omega_1(0) - (n-A_1)\cdot T_1 = \frac{A_1}{n-A_1}$$

The duration T_1 of this time interval is found to be

$$T_1 = \frac{\Omega_1(0) - \frac{A_1}{n-A_1}}{n - A_1} \qquad (32)$$

Inserting eq. (32) into eq. (24), the mean waiting time $t_{w1}(T_1)$ at the time T_1 turns out to be

$$t_{w1}(T_1) = \frac{\Omega_1(0)+1}{n} - \frac{n-A_1}{n}\cdot T_1 = \frac{1}{n-A_1} = t_{w1} \qquad (33)$$

This is the well-known expression for the mean waiting time, referred to delayed 1-calls.

3.2 The mean number of waiting calls of other priorities and their mean times

At the time t=0 many calls of different priority classes are waiting. The initial values $\Omega_1(0)$, $\Omega_2(0)$, ... $\Omega_m(0)$ may have been reached at the end of a preceding saturation period; or they may be fixed prescribed numbers.

From the time t=0, the total traffic offered is sufficiently small that the traffic process can reach stationarity. The long queue, whatever its origin might have been, begins to decrease, when mean values are considered. The passage to stationarity is investigated for all priorities k with traffic offered $A(\leq k)<n$ from t=0.

During the period, when the waiting traffic decreases from the large initial value $\Omega(\leq k;0)$ to the vicinity of the stationary value $\Omega(\leq k)/E$, it is assumed, that all n trunks are continuously occupied by calls of priorities 1,2,...k, cf. the assumption (3). The mean number of waiting calls of priorities 1,2,...k decreases linearly with time t and is given by eq.(9); eq.(10) holds for the corresponding variance.

a) During a <u>first section of time</u> the queue of waiting 1-calls is reduced. This period of cutting down the peak of waiting 1-calls was discussed in the preceding section 3.1. All arriving calls of priorities 2,3,...m must wait. Therefore, the mean number of these waiting k-calls increases.

$$\Omega_k(t) = \Omega(\leq k;t) - \Omega(\leq k-1;t) = \Omega_k(0) + A_k\cdot t \qquad \text{for } k=2,3,\ldots m \qquad (34)$$

b) During a <u>second section of time</u> the mean number of waiting 1-calls remains constant. During the gaps between 1-occupations 2-calls are served. From eq. (9) the mean number of waiting 1-calls and 2-calls at the time t is

$$\Omega(\leq 2;t) = \Omega(\leq 2;T_1) - [n - A(\leq 2)]\cdot(t-T_1) \qquad \text{for } T_1\leq t\leq T_2$$

As the mean number of waiting 1-calls is constant

$$\Omega_1(t) = \frac{A_1}{n-A_1} \qquad \text{for } T_1\leq t\leq T_2$$

the mean number of waiting 2-calls turns out to be

$$\Omega_2(t) = \Omega_2(T_1) - (n-A_1-A_2)\cdot(t-T_1) \qquad (35)$$
$$\text{for } T_1\leq t\leq T_2$$

The mean number of waiting 2-calls increased from the time 0 to T_1, as given by eq. (34) and it decreases from the time T_1 to T_2, as given by eq. (35). The interval of cutting down the peak of waiting 2-calls ends at the time T_2, when the final value of waiting traffic during blocking intervals is reached, cf. eq. (17). The length of the second section of time is

$$T_2 - T_1 = \frac{\Omega_2(T_1) - \Omega_2(T_2)}{n-A(\leq 2)} = \frac{\Omega_2(0)+A_2\cdot T_1 - \frac{n\cdot A_2}{[n-A(\leq 2)](n-A_1)}}{n-A(\leq 2)} \qquad (36)$$

Then, $\Omega(\leq 2)/E$ corresponding to eq. (12) is reached, too. Therefore, T_2 may be obtained alternatively by the following formula:

$$T_2 = \frac{\Omega(\leq 2;0) - \frac{A(\leq 2)}{n-A(\leq 2)}}{n-A(\leq 2)} \qquad (37)$$

This expression is analogous to eq. (32). Following the same considerations as in section 2.9, but for l=2, we find:

$$t_{w2}(t) = \frac{n}{(n-A_1)^2} + \frac{\Omega_2(0)+A_2\cdot T_1}{n-A_1} - \frac{n-A(\leq 2)}{n-A_1}\cdot(t-T_1)$$
$$\text{for } T_1\leq t\leq T_2 \qquad (38)$$

c) Now a <u>third section of time</u> follows. During this interval on the average only the queue of waiting 3-calls will be diminished. Looking at the trunk group, an observer will see only 1-, 2- and 3-occupations, on the average. Following the same reasoning as above, the equations (35), (37) and (38) may be generalized.

$$\Omega_k(t) = \Omega_k(0) + A_k\cdot T_{k-1} - [n-A(\leq k)]\cdot(t-T_{k-1}) \qquad (39)$$
$$\text{for } T_{k-1}\leq t\leq T_k \text{ and } k=1,2,\ldots m$$

$$T_k = \frac{\Omega(\leq k;0) - \frac{A(\leq k)}{n-A(\leq k)}}{n-A(\leq k)} \qquad \text{for } k=1,2,\ldots m \quad (40)$$

$$t_{wk}(t) = \frac{n}{[n-A(\leq k-1)]^2} + \frac{\Omega_k(T_{k-1})}{n-A(\leq k-1)} - \frac{n-A(\leq k)}{n-A(\leq k-1)}\cdot(t-T_{k-1}) \qquad (41)$$
$$\text{for } T_{k-1}\leq t\leq T_k \text{ and } k=2,3,\ldots m$$

3.3 An example

The fig. 5 shows that the mean number $\Omega(\leq k;t)$ of waiting calls of priorities 1,2,...k decreases linearly with the time t. The example given is described by the following parameters: At the time t=0, 2 calls of priority 1, 3 calls of priority 2 and 5 calls of priority 3 are waiting in front of n=20 occupied trunks. From t=0, the values of traffic offered are A_1=1.5, A_2=6.0, and A_3=7.5 Erl.
The formula (40) gives the following results: T_1=0.104, T_2=0.352 and T_3=1.4. The approximation for $T_3<t$ will be deduced in the following section (cf. fig. 9, which shows the continuation of the curve $\Omega(\leq 3;t)$ of fig. 5). The test results show, that the real curves smooth the sharp angles between straight lines calculated with this approximation method.

When regarding the waiting calls of each priority seperately, we find that the mean number $\Omega_k(t)$ increases till the time T_{k-1}, then it decreases. This is shown in fig. 6, for the same example as above.

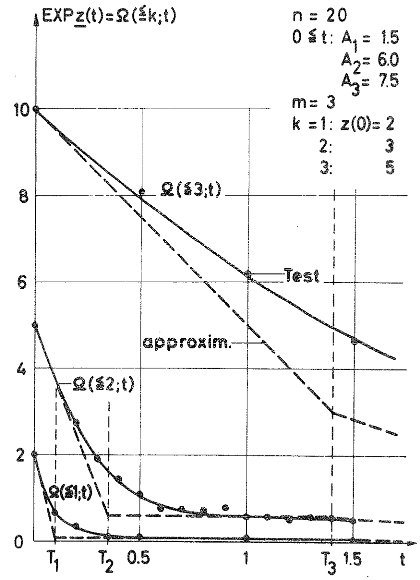

Fig. 5 The mean number of waiting calls of priority k and of higher priorities

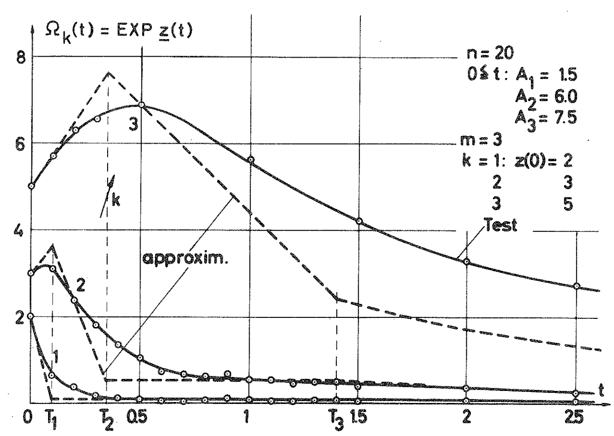

Fig. 6 The mean number of waiting k-calls seperately

Finally, we are going to examine the mean waiting times of the same example. Eq. (41) yields the straight lines plotted in fig. 7. The test results lie on smooth curves. The mean waiting times of 2-calls and 3-calls, which arrived at the time t=0, are smaller than the calculated values. The reason is: the assumption that all trunks are only occupied by 1-calls and 2-calls till the time T_2, is too unfavourable. In reality a non-vanishing probability exists, that a trunk might be occupied by a 3-call even before T_2, and that a trunk might be idle.

3.4 The ebbing period

At the time T_m, all expectation values of waiting calls will reach the stationary values, which are valid during blocking periods. The approximation for the second ebbing period is based on three assumptions for its beginning. They express the conditions of continuity. The expected total number of calls, the velocity of its decrease, and the probability for delay cannot jump. Two further assumptions for the whole ebbing period are: The expected total number of calls decreases exponentially with a

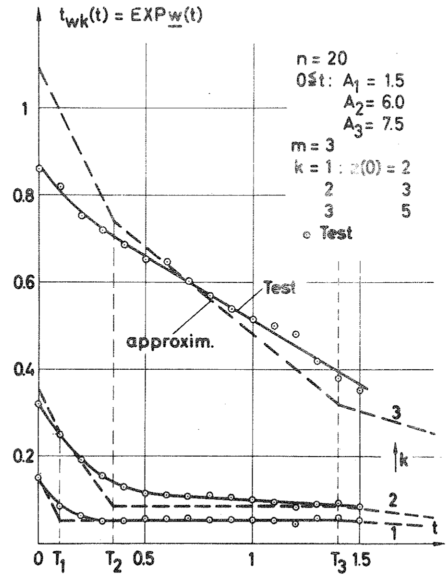

Fig. 7 The mean waiting times of k-calls

time-constant D. Furthermore, the probability for delay is supposed to decrease exponentially with the same time-constant D. The expected number of busy trunks approaches to the value $A(\leq m)$, which is less than n. If the number of calls were always less than n, no difference could be observed between n finite and $n \to \infty$. For $n \to \infty$ the exponential decrease is valid. Therefore we use exponential decrease as an approximation.

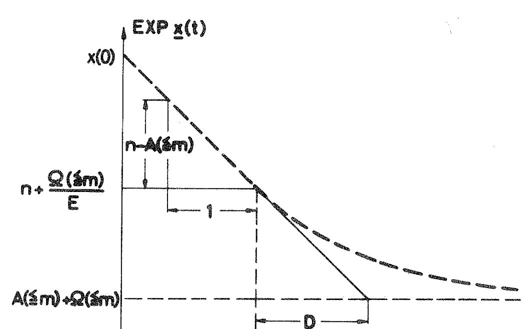

Fig. 8 The decrease of the expectation value EXP $\underline{x}(t)$ of the total number of calls which are present in the delay system at the time t

From the first and second assumption the time-constant is obtained by means of the fig. 8:

$$D = \frac{n + \frac{\Omega(\leq m)}{E} - A(\leq m) - \Omega(\leq m)}{n - A(\leq m)}$$

$$= 1 + A(\leq m)\,\frac{1 - E_{2,n}\,[A(\leq m)]}{[n - A(\leq m)]^2} \qquad (42)$$

The example given in figures 5 to 7 leads to D=1.504. The expectation values EXP $\underline{x}(t)$ and EXP $\underline{z}(t)$ found both by simulation and by this approximation are plotted in fig. 9.

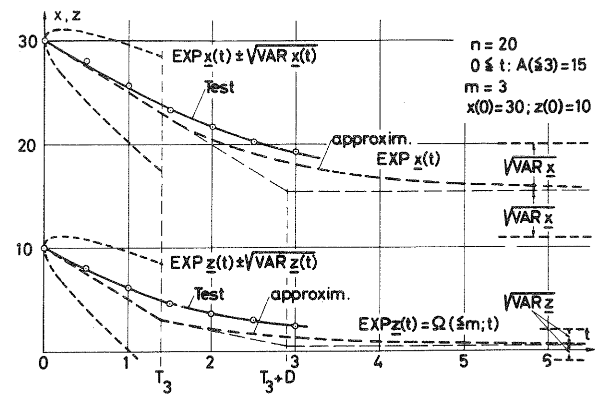

Fig. 9 The decreasing expectation values of the total number of calls EXP $\underline{x}(t)$ and of the waiting calls EXP $\underline{z}(t)$

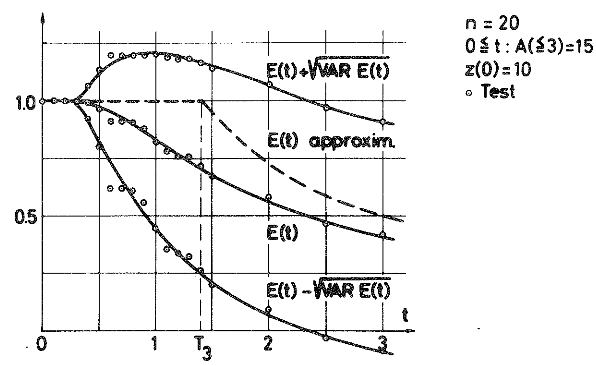

Fig. 10 The probability of delay calculated by the approximation method, and found by a simulation test

4. CONCLUSION

The method deduced in this paper is but an approximation. It gives, however, the final values, which the waiting traffic will reach, and its variance. From these figures the teletraffic engineer can estimate the number of waiting places which is necessary for the queue line of the delay system. The approximation method gives the mean waiting times during the periods both of saturation and of reducing a peak. Thus the grade of service of the delay system is known. This analysis may be useful, to vary the number of trunks or the priority classification in order to achieve a higher grade of service.

In order to improve the calculation, the assumption that all n trunks are always busy should be substituted by a special assumption on the temporary course of the blocking probability. Some of the results of this paper can be generalized to a certain extent: Constant holding times may be treated. In stead of strict queueing, "last come - first served" discipline within each priority may be regarded.

Literature

I. Theory

/1/ CHAMPERNOWNE, D.G., An elementary method of solution of the queueing problem with a single server and constant parameters. Journ. Roy. Statist. Soc. Ser. B, 18 (1956), 125 - 128.

/2/ CLARKE, A.B., A waiting line process of Markov type. Annals of Mathem. Statist. 27 (1956), 452 - 459.

/3/ COX, D.R. and SMITH, W.L., Queues, Methuen London, J. Wiley New York, 1961.

/4/ KARLIN, S. and MC GREGOR, J., Many server queueing processes with Poisson input and exponential service times. Pacific Journ. Mathem. 8 (1958), 87 - 118.

/5/ KESTEN, H. and RUNNENBURG, J.T., Priority in waiting line problems, Proc. Kon. Ned. Akad. van Weten. Ser. A, 60 (1957), 312 - 336.

/6/ LOTZE, A., Berechnung der Verkehrsgrößen im Wartezeitsystem aus den Verkehrsgrößen eines Verlustsystems. Fernmeldetechn. Zeitschr. 7 (1954), 443 - 453.

/7/ LOTZE, A. and SCHWIDERSKI, E., Wartezeitprobleme im Fernsprechverkehr. Nachrichtentechn. Zeitschr. 8 (1955), 646 - 649.

/8/ RANDAZZO, F.P., Optimale Auslegung von Nachrichtennetzen mit Hilfe der Wartezeittheorie. Elektr. Nachrichtenwesen 38 (1963), 524 - 537.

/9/ RIORDAN, J., Stochastic service systems. J. Wiley New York, London, 1962.

/10/ ROBINS, J.M., Message switching traffic formulae for engineers. 5th Internat. Teletraffic Congress, New York 1967, Preprints p. 551 - 555.

/11/ STANGE, K., Die Anlauflösung für den einfachen exponentiellen Bedienungskanal (mit beliebig vielen Warteplätzen), der für t=0 leer ist. Unternehmensforschung 8 (1964), 1 - 24.

/12/ SYSKI, R., Introduction to congestion theory in telephone systems. Oliver and Boyd, Edinburgh, London, 1960.

/13/ WAGNER, W., On combined delay and loss systems with nonpre-emptive priority service. 5th Internat. Teletraffic Congress, New York 1967, Preprints p. 73 - 84.

/14/ WAGNER, W., Über ein kombiniertes Warte-Verlust-System mit Prioritäten. Dissertation Universität Stuttgart, 1968.

/15/ ZIMMERMANN, G.O., and STÖRMER, H., Wartezeiten in Nachrichtenvermittlungen mit Speichern. Oldenbourg München, 1961.

II. Simulation

/16/ COLE, A.C., THOMSON, W.E., and WILLIAMS, J.W.J., Digital computer simulation of switching systems using the ALGOL automatic programming language. 4th Internat. Teletraffic Congress, London 1964, Doc. 50, and Post Office Telecommunications Journal 1964 Special Issue, 35 - 37.

/17/ WAGNER, H., Traffic simulation according to the time-true model (methods and results). 4th Internat. Teletraffic Congress, London 1964, Doc. 52, and Post Office Telecommunications Journal 1964, Special Issue, 39.