

# COMBINED DELAY AND LOSS SYSTEMS WITH SEVERAL INPUT QUEUES, FULL AND LIMITED ACCESSIBILITY

Paul Kühn

Technical University Stuttgart

Stuttgart, Federal Republic of Germany

## ABSTRACT

Queuing theory deals mostly with systems having only one input queue and a group of fully accessible servers. In telephone exchanges, data switching networks, and computer systems there are often structures which cannot be treated by a single queue model.

This paper deals with systems having in general  $g > 1$  input queues. The  $j$ -th input queue has  $s_j$  waiting places,  $j = 1, 2, \dots, g$ . There is a group of  $n$  servers which can be reached with full or limited accessibility. The traffic offered is Poissonian with mean arrival rate  $\lambda_j$  for the  $j$ -th input,  $j = 1, 2, \dots, g$ . The service times are negative exponentially distributed with different means for the different servers, i.e. the mean terminating rate for the  $i$ -th server is  $\epsilon_i$ ,  $i = 1, 2, \dots, n$ .

As disciplines for the service within the queues (queue disciplines)

- D1 : first-come, first-served service,
- D2 : random order of service,
- D3 : last-come, first-served service

will be distinguished. In addition, the following assumptions are made for the service of a certain queue (interqueue disciplines):

- A : service within the whole waiting room which has access to a certain server,
- B : service of the  $j$ -th queue with an arbitrary but constant probability,  $j = 1, 2, \dots, g$ ,
- C : service of the  $j$ -th queue with a probability which depends on the different queue lengths,  $j = 1, 2, \dots, g$ .

The linear equation system for the stationary probabilities of state will be derived. The characteristic values, such as the probabilities of waiting and loss, the traffic carried, the waiting traffic, and the mean waiting time, will be calculated from the stationary probabilities of state. Curves are given to demonstrate the influence of traffic offered and interqueue disciplines on the characteristic values.

The method for the exact calculation of the distribution function of waiting times (d.f.w.t.) will be treated in general and applied to some selected examples. Differential equations for the conditional d.f.w.t. are derived and dealt with in matrix notations. Algorithms for the solution are discussed.

For the special cases of symmetrical systems, which are characterized by identical traffic loads and identical numbers of waiting places for all incoming grading groups, algorithms are given for approximate calculation in case of ideal gradings and standardized gradings. Furthermore, it can be shown that the obtained results hold exactly for fully accessible servers. By these methods it is made possible to calculate relatively large systems which occur in practice. Results of the calculation of large systems will be given and compared with those results obtained by a traffic test on a digital computer.

## 1. STATIONARY PROBABILITIES OF STATE

### 1.1 Full Accessibility

In service systems with full accessibility, calls of each input have access to each server. The rule for selection of a waiting call from a certain queue is given by the interqueue discipline. Fig. 1a shows a simple example of a double-queue system having 3 servers and  $s_j$  waiting places,  $j = 1, 2$ .

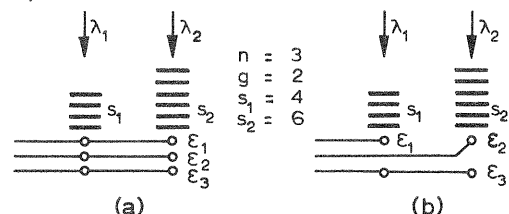


Fig. 1. Example of a multiqueuing system with (a) full accessibility and (b) limited accessibility

Let  $(x_1, x_2, x_3; z_1, z_2)$  be the state defined by

$$x_i = \begin{cases} 0 & \text{server number } i \text{ is idle} \\ 1 & \text{server number } i \text{ is busy} \end{cases}, \quad i = 1, 2, 3,$$

$z_j$  waiting places are occupied within the  $j$ -th queue,  $j = 1, 2$ .

The interqueue discipline can be described by a probability  $p_j$ ,  $j = 1, 2$ .  $p_j$  is the conditional probability that the queue  $j$  will be served, under the condition, that the occupation of a server terminates. For the described model the disciplines within the queues need not be distinguished for the stationary probabilities of state but for the d.f.w.t..

Fig. 2 shows the state space and the transition coefficients, where the abbreviations are:

$$\lambda = \lambda_1 + \lambda_2 \quad (1a)$$

$$\mu = \epsilon_1 + \epsilon_2 + \epsilon_3 \quad (1b)$$

$$\mu_j = \mu \cdot p_j, \quad j = 1, 2. \quad (1c)$$

The stationary probabilities of state - for the above example  $p(x_1, x_2, x_3; z_1, z_2)$  - can be calculated from a linear equation system, the so-called Kolmogorov-forward-equation [1]

$$q_{\nu} p(\nu) - \sum_{\kappa \neq \nu} q_{\kappa \nu} p(\kappa) = 0, \quad (2a)$$

where  $q_{\nu} = \sum_{\kappa \neq \nu} q_{\nu \kappa}$ .

In eq. (2a)  $p(\nu)$  denotes the stationary probability of the state  $(\nu)$ ,  $q_{\kappa \nu}$  denotes the transition coefficient for the transition from state  $(\kappa)$  to state  $(\nu)$ .

Eq. (2a), applied to the state  $(1, 1, 1; 1, 1)$  in the above example, leads to

$$(\lambda_1 + \lambda_2 + \mu_1 + \mu_2) p(1, 1, 1; 1, 1) - \lambda_1 p(1, 1, 0; 1, 1) - \lambda_2 p(1, 1, 1; 0, 1) - \mu_1 p(1, 1, 1; 2, 1) - \mu_2 p(1, 1, 1; 1, 2) = 0.$$

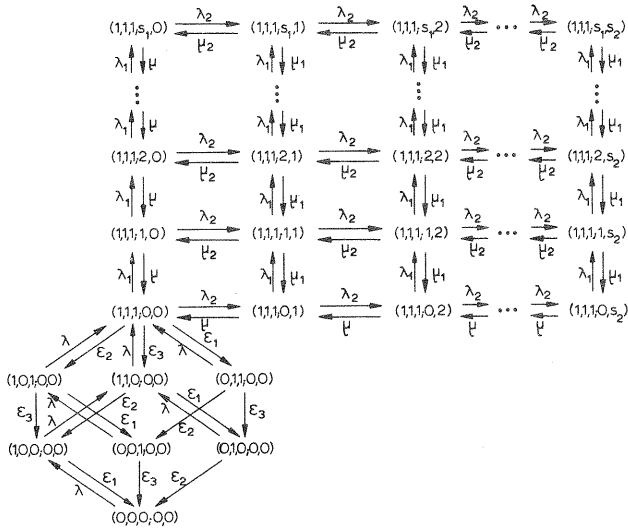


Fig. 2. State space and transition coefficients for a fully accessible 3-server system with two queues

The normalizing condition is

$$\sum_{\nu} p(\nu) = 1. \quad (2b)$$

In the general case the linear equation system (2a,b) will be solved by iterative methods on a digital computer. In some special cases, however, the solution can be given either in terms of recursion formulas or explicitly.

Special cases

a) Identical servers

In this case

$$\epsilon_i = \epsilon, \quad i = 1, 2, \dots, n. \quad (3)$$

Then, the subset of states with no waiting calls can be reduced to a one-dimensional state space defining a state (x) by "x servers are busy". For one-dimensional state spaces the solution of the probabilities of state can be given explicitly by means of recursion. Thus, only the linear equation system for the remaining subset of multi-dimensional states is to be solved.

b) Priority type discipline

The case of nonpre-emptive priority type discipline with a fixed number of waiting places for calls of each priority class is obtained if

$$p_1 = 1, \quad p_2 = 0 \quad \text{for } z_1 > 0, \quad z_2 \geq 0, \quad (4)$$

$$p_1 = 0, \quad p_2 = 1 \quad \text{for } z_1 = 0, \quad z_2 > 0.$$

In this special case of the interqueue discipline B, calls of type 1 have nonpre-emptive priority over calls of type 2. For the two-dimensional subspace of states in fig. 2, a recursion algorithm can be given to calculate the probabilities of state because all quantities  $\mu_2$  vanish. Taking  $p(1,1,1;0,z_2)$  as unknown, all the probabilities  $p(1,1,1; z_1, z_2 - 1)$ ,  $z_1 = 1, 2, \dots, s_1$ , can be expressed in terms of  $p(1,1,1;0,z_2)$ . The equilibrium for the state  $(1,1,1; s_1, z_2 - 1)$  allows the calculation of the unknown  $p(1,1,1;0,z_2)$  and thus, the probabilities  $p(1,1,1; z_1, z_2 - 1)$ ,  $z_1 = 1, 2, \dots, s_1$ . This holds for  $z_2 = 1, 2, \dots, s_2$ . By this method all probabilities of the two-dimensional state space can be expressed in terms of  $p(1,1,1;0,0)$ . In the special case of identical servers, all probabilities of state can be calculated by recursion.

c) Queue length type discipline

This discipline is defined by

$$p_j = \frac{z_j}{z_1 + z_2}, \quad j = 1, 2, \quad z_1 + z_2 > 0. \quad (5)$$

The longer queue will be served with a greater probability. This interqueue discipline is a

special case of C. Additionally, it can be shown that eq. (5) holds also for the interqueue discipline A with respect to D1,2,3 within the whole waiting room.

In case of identical servers the probabilities of state are known explicitly:

$$p(x;0,0) = p(0;0,0) \frac{A^x}{x!}, \quad x = 0, 1, 2, \dots, n \quad (6a)$$

$$p(n; z_1, z_2) = p(0;0,0) \frac{A^n}{n!} \left(\frac{A_1}{n}\right)^{z_1} \left(\frac{A_2}{n}\right)^{z_2} \frac{(z_1 + z_2)!}{z_1! z_2!}, \quad (6b)$$

$$z_j = 0, 1, 2, \dots, s_j, \quad j = 1, 2$$

$$p(0;0,0)^{-1} = \sum_{x=0}^{n-1} \frac{A^x}{x!} + \frac{A^n}{n!} \sum_{z_1=0}^{s_1} \sum_{z_2=0}^{s_2} \left(\frac{A_1}{n}\right)^{z_1} \left(\frac{A_2}{n}\right)^{z_2} \frac{(z_1 + z_2)!}{z_1! z_2!}, \quad (6c)$$

where  $A = \lambda/\epsilon$ ,  $A_j = \lambda_j/\epsilon$ ,  $j = 1, 2$ . (7)

The formulas (6a-c) can be extended easily to the general case of an arbitrary number of queues.

d) Cyclic type discipline

When queues are served in cyclic order approximate solutions for the probabilities of state can also be calculated by the aid of the above model [2].

1.2 Limited Accessibility

Limited accessibility exists in a service system when calls of at least one input have only access to a number  $k < n$  of servers. The interqueue discipline now refers to each individual server. In fig. 1b the simplest example of a double-queue system with limited accessibility is shown. Server 1 serves only 1-calls,  $i = 1, 2$ , where server 3 serves both calls of type 1 and 2.

In general, the interqueue discipline for the  $i$ -th server is given by the probabilities  $p_{ij}$ ,  $j = 1, 2, \dots, g$ , which stand for service of the  $j$ -th queue if server  $i$  terminates,  $i = 1, 2, \dots, n$ . In the above example there only exists an interqueue discipline  $p_{3j}$ ,  $j = 1, 2$ , for server 3.

In fig. 3 a part of the state space is shown for the system of fig. 1b. The definition of states and transition coefficients is the same as in the case of full accessibility.

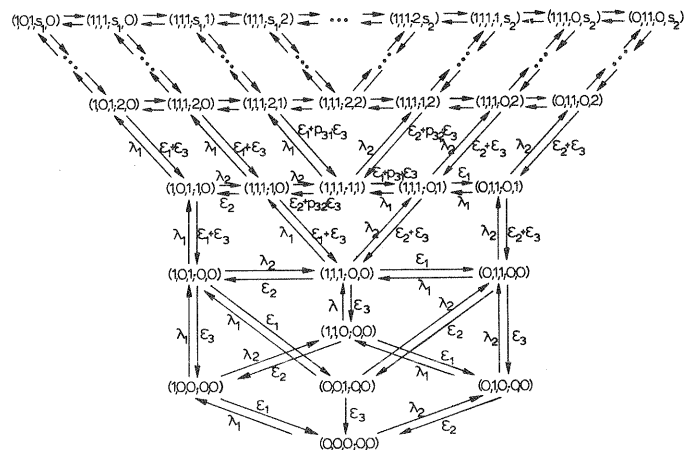


Fig. 3. State space and transition coefficients for a limited accessible 3-server system with two queues

The equations of state can be written in a similar way as in the case of fully accessible servers. In any case, the whole linear equation system is to be solved, e.g. by iterative methods.

### 1.3 The Number of States - Programming

The computability is closely related with the number of states. Multiqueuing systems can lead to an immense number of states so that the exact calculation will be either impossible (because of too many unknowns) or outside the scope of economy (because of too long computation times). There are three possibilities to cope with that problem:

1. Exact description of the system by a smaller number of states (macrostates),
2. Development of efficient approximation methods,
3. Simulation on a digital computer.

In any case, the solution should be tried in this order. Simulation techniques, however, have gained another importance in checking results won by approximate calculation methods. For this purpose the flexibility of special simulation languages (as SIMULA, SIMSCRIPT, GPSS) can be of advantage.

#### a) Number of states in case of full accessibility

For the general system consisting of  $g$  input queues and  $n$  servers the number of states is

$$Z = 2^n - 1 + \prod_{j=1}^g (s_j + 1) \quad (8a)$$

By declaration of an array according to the definition of states the programme needs

$$Z' = 2^n \cdot \prod_{j=1}^g (s_j + 1) > Z \quad (8b)$$

places of working storage.

#### b) Number of states in case of limited accessibility

The number of states cannot be given in general because it depends on the type of grading. For the example given by fig.1b we have

$$Z = 2^3 - 3 + (s_1 + 1) + (s_2 + 1) + (s_1 + 1) \cdot (s_2 + 1).$$

Computer programmes have been developed for calculation of arbitrary systems with full and limited accessibility from the input data

$$\{g, n, s[j], \text{type of grading, type of discipline, } \lambda[j], \epsilon[i]\}.$$

An example of the exact calculation will be given below. Examples for approximate calculations for systems with ideal and standardized gradings are treated in chapter 4.

## 2. CHARACTERISTIC VALUES

The characteristic values are quantities which give an answer to the grade of service of a service system. They are used to dimension such systems. The grade of service can be characterized by the probabilities of waiting and loss, the traffic carried, the mean queue length, the mean waiting time, and the distribution function of waiting times (d.f.w.t.).

As an example, the characteristic values will be defined for a  $n$ -server system with full accessibility and 2 queues (queue disciplines D1 or D2).

### 2.1 The probability of waiting (1-calls)

$$W_1 = \sum_{z_1=0}^{s_1-1} \sum_{z_2=0}^{s_2} p(n; z_1, z_2) \quad 1) \quad (9)$$

### 2.2 The probability of loss or overflow (1-calls)

$$B_1 = \sum_{z_2=0}^{s_2} p(n; s_1, z_2) \quad (10)$$

1)  $p(n; z_1, z_2)$  is an abbreviation f.  $p(1, \dots, 1; z_1, z_2)$ .

### 2.3 The traffic carried

$$Y = \sum_{x_1=0}^1 \dots \sum_{x_n=0}^1 \sum_{z_1=0}^{s_1} \sum_{z_2=0}^{s_2} (x_1 + \dots + x_n) p(x_1, \dots, x_n; z_1, z_2) \quad (11)$$

### 2.4 The mean queue length (queue 1)

$$\Omega_1 = \sum_{z_1=0}^{s_1} \sum_{z_2=0}^{s_2} z_1 \cdot p(n; z_1, z_2) \quad (12)$$

### 2.5 The mean waiting time (1-calls)

$$t_{W_1} = \frac{\Omega_1}{W_1 \cdot \lambda_1} \quad (13)$$

The results hold also for 2-calls when the index numbers are changed. The d.f.w.t. will be treated separately in chapter 3.

For systems with limited accessibility the characteristic values can be obtained from the probabilities of state in a similar way by a somewhat more complicated summation as above.

### 2.6 Example of a 3-server system with limited accessibility

Fig.4 shows the characteristic values for the system of fig.1b. The traffic offered in the first incoming grading group is variable, in the second one constant. The solid curves are shown for priority type discipline according to eq.(4) (queue 1 has priority over queue 2), the dashed curves are shown for queue length type discipline according to eq.(5).

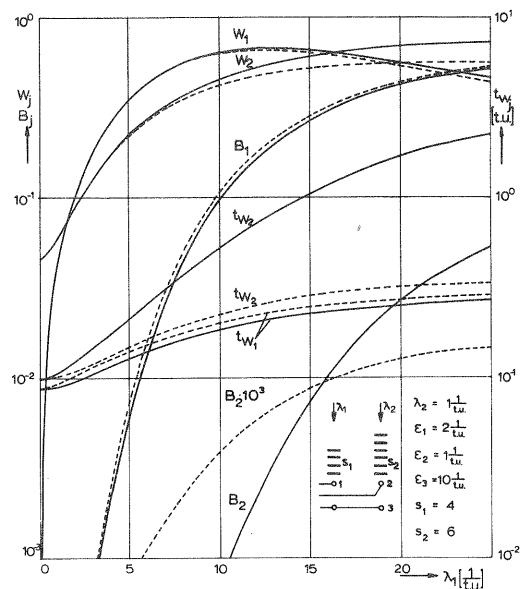


Fig. 4. Characteristic values for a 3-server system with limited accessibility  
 ——— priority type discipline  
 - - - - queue length type discipline

By means of such curves the influences of

1. system parameters
2. system disciplines
3. system loads (input and service time parameters)

on the characteristic values, i.e. the grade of service, can be studied. Vice versa, from a prescribed grade of service one can dimension a system to meet all requirements.

### 3. DISTRIBUTION FUNCTION OF WAITING TIMES

#### 3.1 General Theory

For Markovian systems with only one queue and fully accessible servers R. Syski [1] has given a general theory to calculate the d.f.w.t.. Further investigations on such systems were made by the author [3]. This theory can be extended to multiqueuing systems.

A j-test call arrives at the j-th queue and starts a special waiting process. This process terminates immediately when the j-test call gets either service or is pushed out. A random variable  $\mathcal{Y}_j(t)$  is defined as a random occupation pattern which exists after the waiting time t of the j-test call.  $\mathcal{Y}_j(t)$  contains only those calls in the system which may influence the waiting time of the j-test call, the j-test call being excluded. The definition of  $\mathcal{Y}_j$  depends on the system structure, and on the queue and inter-queue discipline. The  $\mathcal{Y}_j(t)$ -process is a Markovian process.

A (complementary) conditional d.f.w.t. for the j-test call which starts waiting from the pattern i is defined by

$$w_j(t|i) = P\{t_j > t | \mathcal{Y}_j(0) = i\}, \quad i \in H_j, j=1,2,\dots,g, \quad (14)$$

where  $t_j$  denotes the waiting time of the j-test call, and  $H_j$  the set of absorbing states for the j-test call, i.e. all those occupation patterns where the j-test call will either be served or pushed out.

The differential equation system, the so-called Kolmogorov-backward-equation, is given by the following theorem:

$$\frac{dw_j(t|i)}{dt} = -q_j(i)w_j(t|i) + \sum_{\substack{k \neq i \\ k \in H_j}} q_j(i,k)w_j(t|k), \quad i \in H_j, j=1,2,\dots,g, \quad (15a)$$

where  $q_j(i)$ ,  $q_j(i,k)$  are the conditional transition coefficients for the  $\mathcal{Y}_j(t)$ -process. The initial conditions  $w_j(0|i)$ ,  $i \in H_j$ , can also be calculated from (15a) for  $t = 0$  if

$$-\lim_{t \rightarrow 0^+} \frac{dw_j(t|i)}{dt} = \mathcal{E}_j(i), \quad i \in H_j, j=1,2,\dots,g, \quad (15b)$$

where  $\mathcal{E}_j(i)$  denotes the conditional transition coefficient for instantaneous termination of the  $\mathcal{Y}_j(t)$ -process, under the condition, that the state i is reached. Eq.(15a) then reduces to an inhomogeneous linear equation system.

It should be noted that the equation systems (15a,b) hold first of all for j-test calls which wait after their arrival, under the condition, that state i had been found. Defining the conditional d.f.w.t.  $w_j^*(t|i)$  and  $w_j^{**}(t|i)$  for j-test calls which wait after their arrival successfully and in vain, under the condition, that state i had been found, the equation systems (15a,b) hold also for  $w_j^*(t|i)$  and  $w_j^{**}(t|i)$  when  $\mathcal{E}_j(i)$  is substituted by  $\mathcal{E}_j^*(i)$  and  $\mathcal{E}_j^{**}(i)$ , respectively.

From the conditional d.f.w.t. the total d.f.w.t. for all j-calls,  $W_j(t)$ , can be calculated by

$$W_j(t) = \sum_{i \in H_j} P_j(i)w_j(t|i), \quad (16a)$$

where  $P_j(i) = P\{\mathcal{Y}_j(0) = i\}$ ,  $j=1,2,\dots,g$ . (16b)

$P_j(i)$  can be calculated from the stationary probabilities of state. Eq.(16a) holds also for the corresponding quantities  $W_j^*(t)$ ,  $w_j^*(t|i)$  and  $W_j^{**}(t)$ ,  $w_j^{**}(t|i)$ ,  $i \in H_j$ .

To treat eq.(15a,b) the tool of Laplace-transformation is used. From eq.(15a,b) we obtain by that transformation

$$[s + q_j(i)]W_j(s|i) - \sum_{\substack{k \neq i \\ k \in H_j}} q_j(i,k)W_j(s|k) = w_j(0|i), \quad i \in H_j, \quad (17) \\ j=1,2,\dots,g,$$

where  $W_j(s|i)$  is the Laplace-transform of  $w_j(t|i)$ ; s denotes a complex variable. Eq.(17) contains all the information about the waiting time process, the initial conditions included.

Solving of the system (17) is mainly an eigenvalue problem. The eigenvalues of (17) prove to be negative-real. As solution for the  $W_j(s|i)$ ,  $i \in H_j$ , rational functions of s are found. With the aid of the eigenvalues the  $W_j(s|i)$  can be written in form of partial fraction expansion. The partial fractions can be transformed into the time domain by the inverse Laplace transformation. The expressions for the conditional d.f.w.t. are weighted sums of exponential distribution functions (this depends on whether the eigenvalues are single and different or multiple).

For the single-queue system with fully accessible servers the total d.f.w.t. is known for queue discipline D1 [4,5]. For D2 and D3 approximations are also known [6]. Further investigations on the exact calculation of the d.f.w.t. for D2 and D3 were made by the author which yield the exact d.f.w.t. explicitly for D3 and proofs about the location of the eigenvalues for D2 [3].

The conditional mean waiting times for waiting of a j-call from initial state i,  $t_{W_j}(i)$ , from initial state i successfully,  $t_{W_j}^*(i)$ , and from the initial state i in vain,  $t_{W_j}^{**}(i)$ , can be obtained by integration of  $w_j(t|i)$ ,  $w_j^*(t|i)$ , and  $w_j^{**}(t|i)$ , respectively. By this we get

$$t_{W_j}(i) = W_j(0|i), \quad i \in H_j, j=1,2,\dots,g. \quad (18)$$

(For  $t_{W_j}^*(i)$ ,  $t_{W_j}^{**}(i)$  the corresponding relation holds). Hence, the conditional mean waiting times can be calculated from eq.(17) at  $s = 0$ .

The total mean waiting time for j-calls,  $t_{W_j}$ , is obtained by integration of eq.(16a). Thus,

$$t_{W_j} = \sum_{i \in H_j} P_j(i)t_{W_j}(i), \quad j=1,2,\dots,g. \quad (19)$$

Eq.(19), again, holds also for  $t_{W_j}^*$  and  $t_{W_j}^{**}$ , when  $t_{W_j}(i)$  is replaced by  $t_{W_j}^*(i)$  and  $t_{W_j}^{**}(i)$ , respectively.

#### 3.2 Application to Interqueue Discipline A

The interqueue discipline A serves waiting calls throughout the whole waiting room which has access to a certain server according to the dispatching rule (D1,D2,D3). In case of full accessibility the waiting time problem can be reduced principally to that of a single-queue system, because the information about the order of arrival must be stored in a central memory. Thus, the solution method will be the same as for a single-queue system [1,3]. For limited accessibility the waiting time problem can also be solved when the order of arrival is known for all those calls which have access to a certain server.

The interqueue discipline A needs a maximum of control information so that this discipline is rarely realized (except of D2, the random order of service).

#### 3.3 Application to Interqueue Discipline B

The interqueue discipline B serves a certain queue with a certain probability without regard to the actual number of waiting calls within this queue.

a) Full accessibility, queue discipline D1

For example, a double-queue system will be considered now. For 1-calls  $w_1(t|z_1, z_2)$  is defined, where  $(n; z_1, z_2)$  is the state which the 1-test call found on its arrival. For this discipline all customers which were accepted at their arrival will wait successfully. From the equation system (15a,b) follows that  $w_1(0|z_1, z_2) = 1$ , for  $z_1=0, 1, \dots, s_1-1, z_2=0, 1, \dots, s_2$ . The equation system (17) - in matrix notation - is as follows:

$$\begin{array}{c|c|c|c}
 \begin{array}{c} W_1(s_1,0) \quad W_1(s_1,1) \quad \dots \quad W_1(s_1,s_2) \\ \hline (s+\lambda_2)\mu \quad -\lambda_2 \\ \mu_2 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \vdots \\ \mu_2 \quad (s+\mu) \end{array} & \begin{array}{c} W_1(s_1,0) \quad W_1(s_1,1) \quad \dots \quad W_1(s_1,s_2) \\ \hline -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} W_1(s_1-1,0) \quad \dots \quad W_1(s_1-1,s_2) \\ \hline -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \\
 \hline
 \begin{array}{c} \mu_1 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \mu_1 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \vdots \\ \mu_1 \quad (s+\mu) \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \\
 \hline
 \begin{array}{c} \mu_1 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \mu_1 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \vdots \\ \mu_1 \quad (s+\mu) \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \\
 \hline
 \begin{array}{c} \mu_1 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \mu_1 \quad (s+\lambda_2)\mu \quad -\lambda_2 \\ \vdots \\ \mu_1 \quad (s+\mu) \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array}
 \end{array} \quad (20)$$

From the special structure of eq.(20) one can see that the system need not be solved as a whole; the  $W_1(s|z_1, z_2)$  can be determined sectionwise, as shown by the dashed lines. In any case, it is only necessary to solve  $s_1$  systems of  $(s_2+1)$  st order instead of one system of  $s_1(s_2+1)$ st order.

For priority type discipline eq.(20) takes on an even somewhat simpler form, because  $\mu_2 = 0$ . Then all the unknowns can be calculated recursively starting at  $W_1(s|0, s_2)$ , which can be determined directly, up to  $W_1(s|0, 0)$ , then starting at  $W_1(s|1, s_2)$  up to  $W_1(s|1, 0)$ , and so on.

The total d.f.w.t. for 1-calls according to eq. (16a,b) is

$$W_1(>t) = \sum_{z_1=0}^{s_1-1} \sum_{z_2=0}^{s_2} p(n, z_1, z_2) w_1(t|z_1, z_2). \quad (21)$$

b) Full accessibility, queue discipline D3

For the above system let us now consider last-come, first-served service for 1-calls which wait successfully. Let  $w_1^*(t|i_1, i_2)$  be the corresponding conditional d.f.w.t.. Clearly,  $w_1^*(t|0, i_2)$  is the conditional d.f.w.t. for the whole waiting time of a 1-test call, because new 1-calls are allowed to occupy the first waiting place in their queue. The quantities  $w_1^*(t|i_1, i_2)$ ,  $i_1 > 0$ , are the conditional d.f. of partial waiting times for waiting for a server from the  $(i_1+1)$ st waiting place in the first queue. The condition  $i_2$  denotes the number of waiting 2-calls at the arrival of the 1-test call.

The system for calculation of the d.f.w.t. is given by eq.(22). At first, the initial values  $w_1^*(0|i_1, i_2)$  must be calculated from a linear equation system. The matrix eq.(22) shows that the system must be solved as a whole.

The total d.f.w.t. for 1-calls waiting successfully is given by

$$W_1^*(>t) = \sum_{i_2=0}^{s_2} P_1(0, i_2) w_1^*(t|0, i_2), \quad (23a)$$

where

$$P_1(i_1, i_2) = \begin{cases} \sum_{z_1=0}^{s_1} p(n, z_1, z_2) & \text{for } i_1=0, i_2=z_2 \\ 0 & \text{for } i_1>0. \end{cases} \quad (23b)$$

$$\begin{array}{c|c|c|c}
 \begin{array}{c} W_1^*(s_1,0) \quad W_1^*(s_1,1) \quad \dots \quad W_1^*(s_1,s_2) \\ \hline (s+\lambda_2+\mu) \quad -\lambda_1 \\ \mu_2 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \vdots \\ \mu_2 \quad (s+\lambda+\mu) \end{array} & \begin{array}{c} W_1^*(s_1,0) \quad W_1^*(s_1,1) \quad \dots \quad W_1^*(s_1,s_2) \\ \hline -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} W_1^*(s_1-1,0) \quad \dots \quad W_1^*(s_1-1,s_2) \\ \hline -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \\
 \hline
 \begin{array}{c} \mu_1 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \mu_1 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \vdots \\ \mu_1 \quad (s+\lambda+\mu) \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \\
 \hline
 \begin{array}{c} \mu_1 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \mu_1 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \vdots \\ \mu_1 \quad (s+\lambda+\mu) \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \\
 \hline
 \begin{array}{c} \mu_1 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \mu_1 \quad (s+\lambda_2+\mu) \quad -\lambda_1 \\ \vdots \\ \mu_1 \quad (s+\lambda+\mu) \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} -\lambda_2 \\ -\lambda_2 \\ \vdots \\ -\lambda_2 \end{array} & \begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array}
 \end{array} \quad (22)$$

c) Limited accessibility, queue discipline D2

The d.f.w.t. for limited accessible server systems can be calculated in a similar way as for fully accessible servers when the different servers are taken into account. For demonstration, let us consider the system according to fig.1b. For 1-calls a conditional d.f.w.t.  $w_1(t|x_1, x_2, x_3; z_1, z_2)$  will be defined, where  $(x_1, x_2, x_3; z_1, z_2)$  is that state the 1-test call met on its arrival. A waiting process starts only when at least  $x_1 = x_3 = 1$ .

For demonstration, eq.(17) will be applied only to two states the 1-test call met on its arrival.

- 1) The 1-test call met  $(1, 1, 1; 2, 0)$  on its arrival:
 
$$\begin{aligned}
 & (s+\lambda_1+\lambda_2+\epsilon_1+\epsilon_2+\epsilon_3)W_1(s|1, 1, 1; 2, 0) \\
 & - \lambda_1 W_1(s|1, 1, 1; 3, 0) \\
 & - \lambda_2 W_1(s|1, 1, 1; 2, 1) \\
 & - \frac{2}{3}(\epsilon_1+\epsilon_3 p_{31})W_1(s|1, 1, 1; 1, 0) \\
 & - (\epsilon_2+\epsilon_3 p_{32})W_1(s|1, 0, 1; 2, 0) = 1.
 \end{aligned}$$
- 2) The 1-test call met  $(1, 1, 1; 2, 3)$  on its arrival:
 
$$\begin{aligned}
 & (s+\lambda_1+\lambda_2+\epsilon_1+\epsilon_2+\epsilon_3)W_1(s|1, 1, 1; 2, 3) \\
 & - \lambda_1 W_1(s|1, 1, 1; 3, 3) \\
 & - \lambda_2 W_1(s|1, 1, 1; 2, 4) \\
 & - \frac{2}{3}(\epsilon_1+\epsilon_3 p_{31})W_1(s|1, 1, 1; 1, 3) \\
 & - (\epsilon_2+\epsilon_3 p_{32})W_1(s|1, 1, 1; 2, 2) = 1.
 \end{aligned}$$

The solution of such systems is obtained in a similar way as in case of full accessibility.

3.4 Application to Interqueue Discipline C

The interqueue discipline C serves a certain queue with a probability depending on the actual queue lengths of the different queues. For demonstration, let us consider a fully accessible server system with 2 queues according to fig.1a. The queue discipline is first-come, first-served service (D1). For the interqueue discipline eq.(5) is assumed.

A 1-test call arrives and starts waiting in the first queue. All those 1-calls which arrive after the 1-test call and are accepted (i.e. there is still a waiting place available) cannot be served prior to the 1-test call, because of D1. In contrast to a single-queue system, however, those 1-calls also influence the waiting time of the 1-test call because of the interqueue discipline: the longer queue is served with a greater probability (acceleration effect). According to the definition of the  $\mathcal{F}_1(t)$ -process above, those calls also have to be considered for the calculation of the d.f.w.t..

Computer programmes have been developed for exact calculation of the d.f.w.t. for systems with full and limited accessibility. Because of the high order of those differential equation systems only relatively small systems can be calculated exactly. Further investigations are being carried out on approximate calculation of d.f.w.t..

#### 4. SYMMETRICAL SYSTEMS AND THEIR APPROXIMATE CALCULATION

##### 4.1 Definition of Symmetrical Systems

A symmetrical system with  $g$  input queues will be defined by

1. Each input has access to the same number  $k$  out of  $n$  servers.
2. Each input queue has a maximum of  $s$  waiting places.
3. Each input has the same intensity  $\lambda/g$  of offered calls.
4. Each server has the same terminating intensity  $\varepsilon$ .

No special assumptions are made about the inter-queue disciplines and the disciplines within the queues.

##### 4.2 Number of Blocked Grading Groups

The grading shall be described by the number of grading groups  $g$ , the number of servers  $n$ , and the accessibility  $k$ . For ideal gradings according to Erlang [4,7] the number of grading groups is given by

$$g = \binom{n}{k}. \quad (24)$$

When  $x$  servers are busy then the blocking probability  $\sigma(x)$  can be calculated by

$$\sigma(x) = \frac{\binom{x}{k}}{\binom{n}{k}}, \quad x \geq k. \quad (25)$$

The number of blocked grading groups for ideal gradings is

$$b(x) = \binom{x}{k} = g \cdot \sigma(x). \quad (26)$$

For standardized gradings in general  $g \ll \binom{n}{k}$ . The blocking probability  $\sigma(x)$  according to eq.(25) holds approximately also for standardized gradings [7]. A mean value for the number of blocked grading groups is given by

$$b(x) = g \cdot \sigma(x). \quad (27)$$

For ideal gradings  $b(x)$  is an integer and holds exactly. For standardized gradings  $b(x)$  is a real number.

The number of blocked grading groups is identical with the maximum number of possible input queues. For the present only ideal gradings will be considered.

##### 4.3 Definition of States

There are two possibilities to describe symmetrical systems; first, the description by "grading group input queues", and second, by "waiting place rows" [8].

###### a) Description by "grading group input queues"

A state  $(x; z_1, z_2, \dots, z_g)$  is defined by " $x$  servers are busy and  $z_j$  waiting places are occupied within the  $j$ -th input queue,  $j=1, 2, \dots, g$ ". The total number of states is given by

$$Z_1 = k + \sum_{x=k}^n (s+1)^g \sigma(x). \quad (28)$$

###### b) Description by "waiting place rows"

The waiting calls which are waiting at the  $\nu$ th waiting place within their queue form the  $\nu$ th "waiting place row",  $\nu=1, 2, \dots, s$ . A state  $(x; u_1, \dots, u_s)$  is defined by " $x$  servers are busy and  $u_\nu$  waiting places are occupied within the  $\nu$ th waiting place row,  $\nu=1, 2, \dots, s$ ".

Clearly, in the  $(\nu+1)$ st row there cannot wait more calls than in the  $\nu$ th row. This leads to a restriction condition:

$$g \sigma(x) \geq u_1 \geq u_2 \geq \dots \geq u_s. \quad (29)$$

The description by waiting place rows needs fewer states than the description by grading group input queues. This is a typical example of a description by macrostates.

The total number of states is

$$Z_2 = k + \sum_{x=k}^n \left\{ \prod_{\nu=1}^s \frac{g\sigma(x) + \nu}{\nu} \right\}. \quad (30)$$

The proof of this formula is given in [8]. For large systems, especially  $Z_2$  is far smaller than  $Z_1$ . Thus, the approximate calculation of symmetrical systems starts with the description by waiting place rows.

##### 4.4 Equations of State

In fig.5 part of the  $(s+1)$ -dimensional state space is shown.

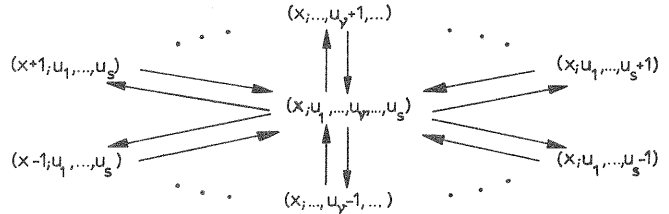


Fig.5.  $(s+1)$ -dimensional state space for symmetrical systems

The transition coefficients for the following transitions are:

$$(x; u_1, \dots, u_s) \longrightarrow (x-1; u_1, \dots, u_s) : \rho_0(x; u_1, \dots, u_s) \cdot \varepsilon \quad (31a)$$

$$(x; \dots, u_\nu, \dots) \longrightarrow (x; \dots, u_{\nu-1}, \dots) : \rho_\nu(x; \dots, u_\nu, \dots) \cdot \varepsilon, \nu=2, \dots, s \quad (31b)$$

$$(x-1; u_1, \dots, u_s) \longrightarrow (x; u_1, \dots, u_s) : [1 - \sigma(x-1)] \cdot \lambda \quad (31c)$$

$$(x; u_1-1, \dots, u_s) \longrightarrow (x; u_1, \dots, u_s) : [g\sigma(x) - (u_1-1)] \cdot \frac{\lambda}{g} \quad (31d)$$

$$(x; \dots, u_\nu-1, \dots) \longrightarrow (x; \dots, u_\nu, \dots) : [u_{\nu-1} - (u_\nu-1)] \cdot \frac{\lambda}{g}, \nu=2, 3, \dots, s. \quad (31e)$$

In eq.(31a, b)  $\rho_\nu(x; u_1, \dots, u_s)$ ,  $\nu=0, 1, \dots, s$ , denotes a conditional transition probability which contains the information about the interqueue disciplines. In any case, the relation

$$\sum_{\nu=0}^s \rho_\nu(x; u_1, \dots, u_s) = 1 \quad (32)$$

holds for all possible states.

The statistical equilibrium for the state  $(x; u_1, \dots, u_s)$  with regard to eq.(32) leads to

$$\begin{aligned} & \{x\varepsilon + \lambda[1 - \sigma(x)] + \frac{\lambda}{g}[g\sigma(x) - u_s]\} \cdot p(x; u_1, \dots, u_s) \\ &= \lambda[1 - \sigma(x-1)] \cdot p(x-1; u_1, \dots, u_s) \\ &+ \frac{\lambda}{g}[g\sigma(x) - (u_1-1)] \cdot p(x; u_1-1, \dots, u_s) \\ &+ \frac{\lambda}{g} \sum_{\nu=2}^s [u_{\nu-1} - (u_\nu-1)] \cdot p(x; \dots, u_{\nu-1}, \dots) \\ &+ (x+1)\varepsilon \cdot \rho_0(x+1; u_1, \dots, u_s) \cdot p(x+1; u_1, \dots, u_s) \\ &+ x\varepsilon \cdot \sum_{\nu=1}^s \rho_\nu(x; \dots, u_\nu, \dots) \cdot p(x; \dots, u_\nu-1, \dots). \end{aligned} \quad (33)$$

The equation system (33) must be solved as a whole when the conditional transition probabilities are known and if exact calculation is wanted. The aim of an approximate calculation, however, is to find a recursion algorithm.

To realize this, an assumption of "half-symmetry" within the multi-dimensional state space is made, i.e. it is presumed that the state  $(x; u_1, \dots, u_s)$  is already in statistical equilibrium with the "lower" neighbour states. An assumption of this kind was made first for the calculation of the pure waiting system with limited accessibility [9]. By this assumption the equations of state take on a simpler form

$$\begin{aligned} p(x; u_1, \dots, u_s) &= \frac{A}{x} [1 - \sigma(x-1)] \cdot p(x-1; u_1, \dots, u_s) \\ &+ \frac{A}{gx} [g\sigma(x) - (u_1-1)] \cdot p(x; u_1-1, \dots, u_s) \\ &+ \frac{A}{gx} \sum_{\nu=2}^s [u_{\nu-1} - (u_\nu-1)] \cdot p(x; \dots, u_{\nu-1}, \dots), \end{aligned} \quad (34)$$

where  $A = \lambda/\varepsilon$  denotes the "traffic offered".

From the equation system (34) two main conclusions can be drawn:

1. The system does not contain information about the interqueue disciplines, i.e. the quantities  $\beta_j$  are eliminated.
2. The probabilities of state can be solved altogether recursively.

Because of restriction (29) there do not exist all states within the multi-dimensional space. A programme has been developed which allows the transformation of the multi-dimensional states onto a one-dimensional space without any gap [10]. By this, the whole working storage of a digital computer can be used for the unknowns.

From the probabilities of state the characteristic values can be calculated in a quite similar way as shown above. Distinctions need not be made between calls of different inputs because of the symmetrical system. Below, results will be given for some examples.

#### 4.5 Remarks on Systems with Full Accessibility and Systems with Nonideal (Standardized) Gradings

##### a) Full accessibility

Full accessibility is a limiting case of a grading. Because of  $k = n$  from eq.(25) follows:

$$G(x) = \begin{cases} 0 & \text{for } x < n \\ 1 & \text{for } x = n \end{cases} \quad (35)$$

It can be shown that the results obtained by recursion from eq.(34) with regard to eq.(35) hold exactly for an interqueue discipline according to

$$p_j = \frac{z_j}{z_1 + z_2 + \dots + z_g}, \quad j=1, 2, \dots, g, \quad (36)$$

i.e. for A and C.

##### b) Nonideal (Standardized) gradings

As mentioned above, for standardized gradings the number of blocked grading groups is in general a real number if there are  $k \leq x < n$  servers busy. The above theory can be applied also to standardized gradings if the real number  $b(x)$  is properly rounded up or rounded off [11].

#### 4.6 Results

The following figures 6 and 7 show results of the approximate calculation of symmetrical systems with ideal gradings and nonideal gradings. The results of traffic tests which were made on a digital computer are given for comparison (confidence interval 95%).

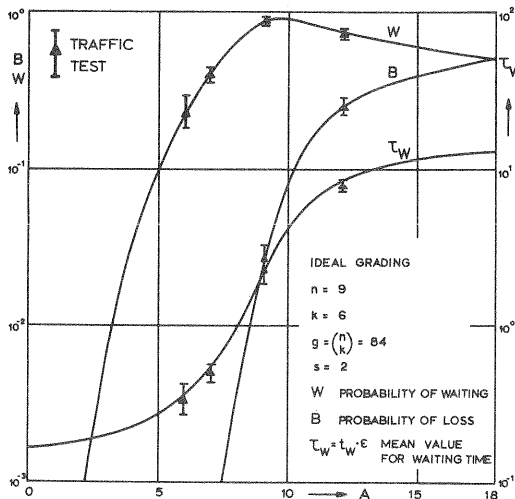


Fig.6. Characteristic values of a symmetrical service system with ideal grading

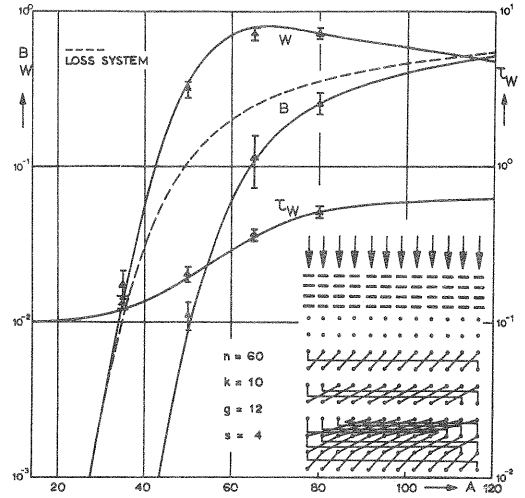


Fig.7. Characteristic values of a symmetrical service system with standardized grading

#### CONCLUSION

For combined delay and loss systems exact methods were investigated for the calculation of the stationary probabilities of state and the d.f.w.t. in case of different disciplines within the queues and different interqueue disciplines. An approximate calculation method was given for symmetrical systems with ideal and nonideal gradings.

#### REFERENCES

- [1] Syski, R. Markovian Queues  
Symposium on Cong. Theory  
Univ. North Carolina, 1965.
- [2] Kühn, P. Parallel Waiting Queues in Real-Time Computer Systems  
Proc. Int. Comp. Symp. 1970  
ACM Conf., Bonn, 1970.
- [3] Kühn, P. Über ein kombiniertes Warte-Verlustsystem mit verschiedenen Abfertigungsdisziplinen  
To be published in A.E.Ü. 1970.
- [4] Brockmeyer, E. The Life and Works of A.K. Erlang  
Halstrøm, H.L. Trans. Danish Acad. Tech. Sci.  
Jensen, A. No. 2., Copenhagen, 1948.
- [5] Störmer, H. Wartezeitlenkung in handbedienten Vermittlungsanlagen  
A.E.Ü. 10, 1956, H. 2, p. 58.
- [6] Riordan, J. Stochastic Service Systems  
J. Wiley & Sons, Inc.,  
New York/London, 1962.
- [7] Lotze, A. Verluste und Güteerkmale einstufiger Mischungen  
N.T.Z. 14, 1961, H. 9, p. 241.
- [8] Kühn, P. Kombinierte Warte-Verlustsysteme mit einstufiger Koppelanordnung und vollkommener bzw. unvollkommener Erreichbarkeit  
Inst. f. Switch. a. Data Techn.  
No. 209, Univ. Stuttg., 1967.
- [9] Thierer, M. Delay Systems with Limited Accessibility  
5th ITC, New York, 1967.
- [10] Kiefer, H. Über die Berechnung des kombinierten Warte- und Verlustsystems mit unvollkommener Erreichbarkeit  
Inst. f. Switch. a. Data Techn.  
No. 253, Univ. Stuttg., 1968.
- [11] Baur, W. Über die Berechnung des kombinierten Warte- und Verlustsystems mit vereinfachten Normmischungen  
Inst. f. Switch. a. Data Techn.  
No. 286, Univ. Stuttg., 1969.