WAITING TIME DISTRIBUTIONS IN MULTI-QUEUE DELAY SYSTEMS WITH GRADINGS

Paul Kühn
University of Stuttgart
Stuttgart, Federal Republic of Germany

ABSTRACT

This paper deals with the problem of waiting times in delay systems having n graded servers with accessibility $k < n$ and $g > 1$ parallel input queues. Various types of gradings are considered with respect to wiring mode and mean interconnecting number M. The servers within a grading group are hunted sequentially. Various operation modes are considered for the service of a certain queue (interqueue discipline) and for the service of a certain waiting call within a queue (queue discipline) when a server becomes idle. For both interarrival times and service times negative exponential distributions are assumed.

The exact calculation is carried out by multidimensional state descriptions and application of the Kolmogorov-forward and backward-equations. Examples of exact evaluations are given to demonstrate the influences of accessibility, mean interconnecting number, and wiring mode on the grade of service. The approximate calculation is based on the "Interconnection Delay Formula", which has been adapted to gradings of various type with arbitrary mean interconnecting number and two different interqueue disciplines. For the approximate calculation of the distribution function of waiting times, two models are suggested based on approaches by exponential and gamma distribution functions, respectively. All approximate calculation methods were checked by extensive event-by-event simulations.

Besides the development of analytical tools for the exact and approximate calculation of multi-queue delay systems, the question of optimum gradings for delay systems is answered by means of exact calculations as well as simulations.

Beitrag des Instituts für Nachrichtenvermittlung und Datenverarbeitung der Universität Stuttgart zum 7th International Teletraffic Congress Stockholm vom 13.-20. Juni 1973

1. INTRODUCTION

Delay mechanisms can be found in automatic telephone or data switching systems for reasons of an economic use of centralized devices (servers) as trunks, registers, markers, and processors for common control. The connection with these centralized servers is done either by single stage connecting networks with full or limited accessibility or by multi-stage connecting networks (link systems), respectively. Single stage systems with limited access and link systems are applied for reasons of a more economic number of crosspoints.

In single stage connecting systems with limited accessibility, the outgoing servers of several selector multiples are partially interconnected according to an interconnection scheme (grading). Compared with full accessibility (full interconnection), this method results in a smaller number of crosspoints at the expense of the grade of service.

During the past, gradings have extensively been studied for loss systems, cf. the survey of A. LOTZE [1]. Delay systems with fully accessible servers were investigated first by A.K. ERLANG [2]. For graded delay systems, an interpolation method was proposed by E. GAMBE [3] using results of full accessibility. M. THIERER [4,5] derived expressions for the probability of waiting and the mean waiting time on the basis of a two-dimensional state description using the combinatorial blocking probability for ideal gradings ("Interconnection Delay Formula IDF"). Combined delay and loss systems with gradings have been investigated by the author [6,7,8]. Multi-stage connecting networks with waiting were studied by E. GAMBE [9] and L. HIEBER [10].

The studies made in [3,4] were focused mainly on mean values as the probability of delay, the mean queue length, and the mean waiting time. In this paper, these investigations are extended to exact and approximate calculations of the probability distribution function (pdf) of waiting time which is a more detailed criterion with respect to the grade of service than the mean waiting time. This criterion must be taken into consideration in those cases where impatient customers cause defections when the waiting time exceeds a certain amount, or where the time conditions in successive processing phases become critical.

Another problem occurs by the fact that the efficiency of gradings can be quite different depending on the wiring mode, mean interconnecting number, and operation mode (hunting and interqueue disciplines). For this reason, the results of the "Interconnection Delay Formula" were extended to gradings of various types with arbitrary mean interconnecting number and two types of interqueue disciplines. Moreover, the results should also be applicable to delay systems with link arrangements which behave similar to a grading.

After a more detailed statement of the problems in Chapter 2, the exact calculation of multi-queue delay systems with gradings is outlined in Chapter 3. Approximate calculation methods are described in Chapter 4. Examples of exact and approximate evaluations are given for demonstration of the influences of

accessibility, mean interconnecting number, wiring and operation mode on the grade of service. Besides the development of analytical tools for the exact and approximate calculation of multi-queue delay systems, the question of optimum gradings for delay systems is answered by means of exact calculations and simulations in this paper.

2. STATEMENT OF THE PROBLEM

A queuing problem can generally be defined by the system structure, the operation modes, the input process, and the service time characteristic. In this chapter, the basic assumptions and problems will be discussed more in detail.

2.1 System structure of multi-queue delay systems

2.1.1 General structure

The multi-queue delay system consists of g input queues (grading group queues), each of them is assigned to an input process of calls. The calls are served by n servers which are fully or partially interconnected (commoned). For partially interconnected servers, calls of each group can only hunt k out of n servers (k accessibility). The interconnection scheme is also called as grading. In Fig. 1 an example is given having n = 8 servers, accessibility k = 4, and g = 4 grading groups.

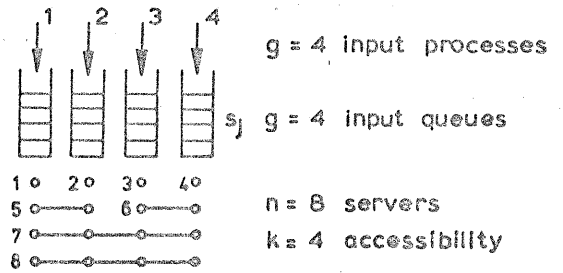

Fig. 1. Example of a multi-queue delay system with limited accessibility

In a pure delay system, the maximum number of storage places s_j in queue j must be sufficiently large such that no loss occurs (j = 1,2,...,g). A combined delay and loss system is generally obtained by limitation of the queues.

In case of full accessibility, all servers are fully interconnected. Full accessibility can therefore be considered as a limiting case (k = n) of a grading.

2.1.2 Types of gradings

A special interconnection scheme (wiring) has an important influence on the efficiency of a grading and has been intensively studied for loss systems, cf. [11-21]. Generally, three main wiring methods are applied for the construction of gradings:

- Commoning
- Skipping
- Slipping.

Applying these wiring methods on the above example (n = 8, k = 4, g = 4) leads to following gradings, cf. Fig. 2.

Besides these structural criteria, the properties of a grading are furthermore reflected (for given n and g) by the following criteria [15]:

- Accessibility k
- Mean interconnecting number M = gk/n
- Matrix for the distribution of busies (b_{ij}), where b_{ij} the number of interconnections between grading groups i and j.

In general, the efficiency of a grading increases with k and M. The criteria, which were found for efficient gradings in loss systems with respect to the probability of loss and traffic balance, cf. [15], can be extended to delay systems in principle. However, additional criteria have to be taken into consideration also, e.g., the influence of the service disciplines on

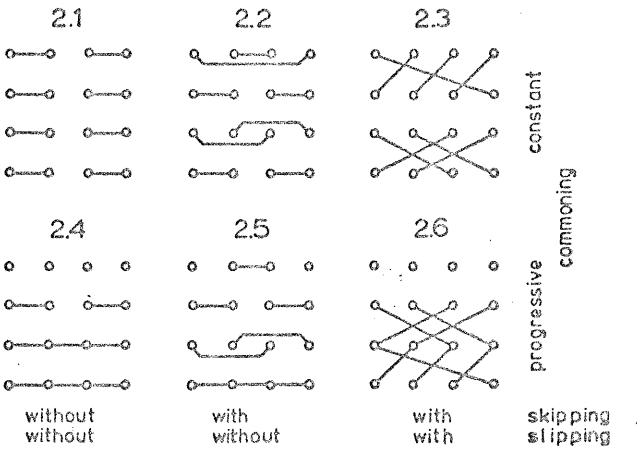

Fig. 2. Types of gradings with various wiring methods
Example: n = 8, k = 4, g = 4, (M = 2).
2.1 Straight homogeneous "grading"
2.2 Straight homogeneous grading with skipping
2.3 Homogeneous grading with skipping and slipping
2.4 Straight inhomogeneous grading ("O'Dell-Grading")
2.5 Straight inhomogeneous grading with skipping
2.6 Inhomogeneous grading with skipping and slipping

mean values and the influence of the mean interconnecting number M on the higher moments of the pdf of waiting time, cf. Chapter 4.

The effects of the various grading types of Fig. 2 on the mean values will be investigated in Chapter 3. Another study made in Chapter 3 shows the dependence of the grade of service from the accessibility k for gradings having n = 6 and g = 3, as shown in Fig. 3.

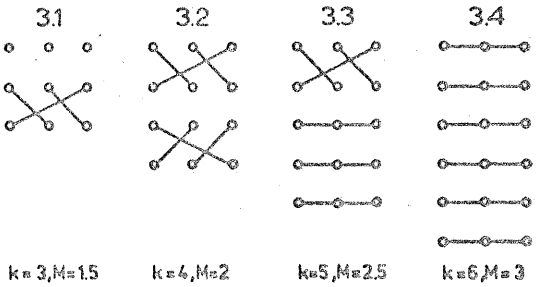

Fig. 3. Types of gradings with various accessibilities
Example: n = 6, g = 3, k = 3 up to 6.

In the examples Figs. 2.1 and 3.4 fully accessible service systems were obtained as limiting cases. Another limiting case, the "Ideal Erlang-Grading", is obtained for g = $\binom{n}{k}$k!. For practical systems g assumes too large values. This case, however, is of great theoretical interest as the blocking probability of this grading is explicitly known [2].

In practical switching systems, only few types of gradings are applied which have economic advantages as simple manufacturing and low costs for installation and extension. Two examples are given in Fig. 4, a straight inhomogeneous grading ("O'Dell-Grading") of the BPO and a standard grading of the German GPO. For comparison, another example of a "high-efficiency grading" with skipping and good traffic balance is given, too (high-efficiency gradings are individually constructed while gradings of the O'Dell-type, standard gradings and simplified standard gradings are regularly constructed).

2.2 Operation modes

2.2.1 Hunting disciplines

Servers can be hunted sequentially (with or without homing) or in random order. For gradings with progressive commoning the sequential hunting method (with homing) is optimal and has, therefore, been assumed for exact calculations and simulations throughout this paper.

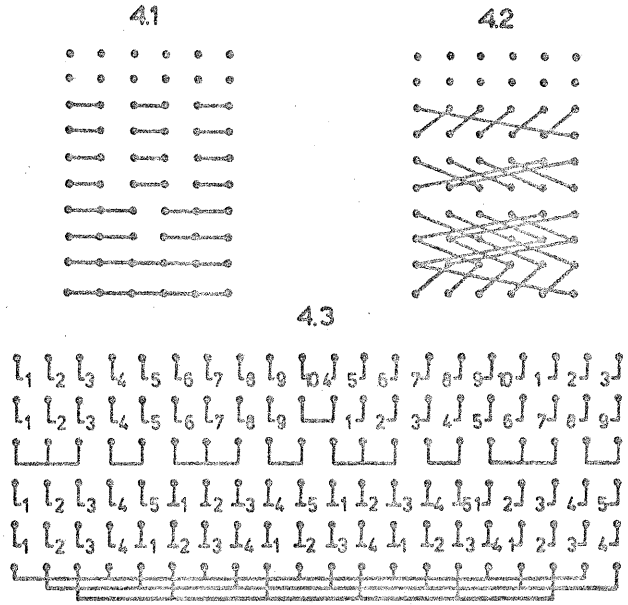

Fig. 4. Examples of practical gradings

4.1 O'Dell-grading, n=30, k=10, g=6, M=2
4.2 Standard grading, n=30, k=10, g=6, M=2
4.3 High-efficiency grading, n=40, k=6, g=20, M=3

2.2.2 Interqueue disciplines

The interqueue discipline controls the service of (non-empty) queues in a multi-queue system with full or partial common servers. For the exact calculation a general probabilistic discipline has been introduced [6,7,8] which yields some special cases as, e.g.

- priority service of queues
- random service of queues
- queue lengths-dependent service of queues.

The influence of the interqueue discipline on the grade of service has been studied by means of exact calculations, cf. [6,7,8]. For graded multi-queue delay systems, the RANDOM service of queues seems to be the most realistic interqueue discipline; this interqueue discipline will be compared with the idealistic interqueue discipline FIFO in Chapter 4.

2.2.3 Queue disciplines

The queue discipline controls the service of waiting calls within a certain queue. Except for displacing priorities, the queue discipline has no effect on the mean values but on the pdf of waiting time. These effects have been studied by exact calculations for FIFO, RANDOM, and LIFO queue disciplines in connection with various classes of interqueue disciplines [7]. In this paper, only FIFO is assumed for the study of waiting time distributions.

2.3 Input process and service times

The input processes are assumed to be Poissonian with mean arrival rate λ_j for group j, $j = 1,2,\ldots,g$. The service times are distributed according to negative exponential pdf's with mean termination rate ε_i for server i, $i = 1,2,\ldots,n$. For the numerical examples, only symmetrical conditions are assumed, i.e. $\lambda_j = \lambda/g$, $j = 1,2,\ldots,g$, $\varepsilon_i = \varepsilon$, $i = 1,2,\ldots,n$. Furthermore, the service system is considered in the stationary state.

2.4 Grade of Service

The grade of service of a multi-queue delay system may be given by the main characteristic values as

- probability of waiting
- probability of loss (finite queues)
- carried traffic
- mean queue length
- mean waiting time
- pdf of waiting time
- higher moments of the pdf of waiting time.

In the following, these more detailed values will be regarded for the service quality.

Another important quality parameter of a grading can be derived by consideration of unbalanced load as it was suggested for loss systems by A. LOTZE [13]. Studies of the influence of unbalanced load as well as extremely uneven interqueue disciplines (priority service) in delay systems have been reported in [6,7,8] and will not be dealt with in the following.

3. EXACT CALCULATION OF MULTI-QUEUE DELAY SYSTEMS WITH GRADINGS

The exact calculation of multi-queue delay systems with gradings is based on methods of state equations. In principle, the analysis can be performed in two steps [6,7,8]:

1. Solution of a system of linear equations for the stationary probabilities of state (Kolmogorov-forward- or equilibrium-state-equations)
2. Solution of a system of linear differential equations for the conditional pdf's (cpdf) of waiting time (Kolmogorov-backward-equations).

The main traffic values which characterize the grade of service can be derived from these values subsequently. In the following two sections, only the fundamental way of solution will be outlined; for a more detailed discussion it is referred to [7].

In pure delay systems, the number of states tends to infinity for an infinite number of sources. Therefore, numerical examples will be confined to combined delay and loss systems with a finite number of states.

3.1 The stationary state

3.1.1 Probabilities of state

A system state ξ may be defined by a (n+g)-dimensional vector

$$\xi = (\ldots,x_i,\ldots;\ldots,z_j,\ldots),\ \xi \in \Xi, \qquad (3.1)$$

where $x_i = 0(1)$ if server i is idle(busy), $i=1,2,\ldots,n$, and $z_j = 0,1,\ldots,s_j$ the number of occupied storage places within queue j, $j=1,2,\ldots,g$. The set Ξ of system states includes only those states which are physically possible (a queue j can only be built up if at least all accessible servers within grading group j are busy).

The stationary probabilities of state, $p(\xi)$, can be determined from the Kolmogorov-forward-equations considering the service system in equilibrium state

$$q_\xi p(\xi) - \sum_{\pi \neq \xi} q_{\pi\xi}\, p(\pi) = 0, \quad \xi \in \Xi, \qquad (3.2a)$$

completed by the normalizing relation

$$\sum_{\xi \in \Xi} p(\xi) = 1. \qquad (3.2b)$$

In Eq.(3.2a), $q_{\pi\xi}$ means the coefficient for the transition from state $\pi \neq \xi$ to state ξ, and q_ξ the coefficient for leaving state ξ, where $q_\xi = \sum_{\pi \neq \xi} q_{\xi\pi}$.

The general state representation of the process of system states is illustrated in Fig.5. Examples for special multidimensional state representations were given in [6,7,8] and will not further be discussed in this paper. Eq.(3.2a) can be formulated in detail regarding

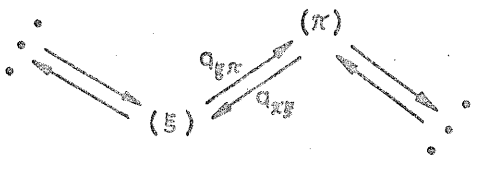

Fig. 5. Process of system states: General state representation with transitions

formal descriptions for gradings and operation modes. The general formulation was reported in [8] for sequential hunting. For the practical evaluation of Eqs. (3.2a,b), a computer program has been developed which generates the equations for arbitrary grading structures, queue lengths, arrival and termination rates, and various types of interqueue disciplines; the numerical calculations are carried out by the method of successive overrelaxation. By this method, service systems can be calculated with a total number of states up to the available computer storage capacity.

3.1.2 Characteristic mean values

The most important mean values can be obtained from the probabilities of state by following definitions:

a) Probability of waiting for group j

$$W_j = \sum_{\xi \in \Xi} (1-\delta_{z_j,s_j}) \cdot \prod_{h=1}^{k_j} (x_{g_{hj}}) \cdot p(\xi) \qquad (3.3)$$

b) Probability of loss for group j

$$B_j = \sum_{\xi \in \Xi} \delta_{z_j,s_j} \cdot p(\xi) \qquad (3.4)$$

c) Carried traffic on server i

$$Y_i = \sum_{\xi \in \Xi} x_i p(\xi) \qquad (3.5)$$

d) Mean queue length of queue j

$$\Omega_j = \sum_{\xi \in \Xi} z_j p(\xi) \qquad (3.6)$$

e) Mean waiting time referred to all waiting j-calls

$$t_{Wj} = \Omega_j / (\lambda_j W_j) \ , \qquad (3.7)$$

where δ_{ij} the Kronecker symbol, k_j the number of accessible servers from group j, j and g_{hj} the number of that server which is hunted at step h in group j, $h = 1,2,\ldots,k_j$; $j = 1,2,\ldots,g$.

3.2 Distribution of waiting time

3.2.1 Conditional pdf's of waiting time

For the exact calculation of waiting time distribution, the waiting process of a test call is considered within the j-th queue. A j-call enters the queue j and starts a waiting process; this process is being "alive" as long as the j-call is waiting and "dies" at that ent the j-call is selected for service. This waiting process can be constructed from the process of system states by neglecting all those transitions which do not influence the "life-time" of the j-call under consideration.

For the formal description of the waiting process in queue j, a waiting state ζ_j is introduced which considers all those calls in the system which may have an influence on the waiting time of the considered j-call. Therefore, ζ_j is built up by the states x_i of all those servers which have no access to group j, and the states z_ν of all queues (The waiting process for j-calls can only exist if at least all servers are busy which are accessible from group j). The states z_ν must be defined dependent on the special queue and interqueue disciplines.

Examples:

In the simplest case, the interqueue discipline does not depend on the actual lenghts of the various queues (e.g., random selection of queues), and the queue discipline is FIFO. Then, the waiting time of the j-test call is not influenced by subsequent arriving j-calls. In this case, the waiting state ζ_j can be defined by a $(n-k_j+g)$-dimensional vector

$$\zeta_j = (\ldots,x_i,\ldots;\ldots,z_\nu,\ldots), \ \zeta_j \in Z_j, \qquad (3.8)$$

where $x_i = 0(1)$ if server i is idle(busy), $i \neq g_{hj}$, $h = 1,2,\ldots,k_j$, z_ν = number of waiting calls in queue ν, $\nu \neq j$, and z_j = number of calls waiting in front of the j-test call; set Z_j includes all possible waiting states for a j-test call.

If the queue discipline were RANDOM, the same waiting state acc. to Eq.(3.8) could be used when z_j is defined as the number of competitors to the j-test call in queue j. For LIFO (with displacing priority in case of finite queues), z_j represents the number of calls in queue j which have arrived subsequently to the j-test call.

For other interqueue disciplines (e.g., selection of queues according to their lengths), the state z_j must be composed by two components, the numbers of predecessors and successors of the j-test call, and so on.

The distribution of waiting time for j-calls which met an arbitrary state ζ_j at their arrival is defined by a conditional (complementary) pdf (cpdf)

$$w_j(t|\zeta_j) = P\left\{T_{Wj} > t \,\middle|\, \zeta_j\right\}, \quad \zeta_j \in Z_j \ . \qquad (3.9)$$

The cpdf's of waiting time, $w_j(t|\zeta_j)$, can be determined from the Kolmogorov-backward- equations:

$$\frac{d}{dt} w_j(t|\zeta_j) = -q_{\zeta_j} w_j(t|\zeta_j) + \sum_{\eta_j \neq \zeta_j} q_{\zeta_j,\eta_j} w_j(t|\eta_j), \qquad (3.10)$$

$$\zeta_j, \eta_j \in Z_j,$$

where q_{ζ_j,η_j} the coefficient for the transition from waiting state ζ_j to waiting state η_j, and q_{ζ_j} the coefficient for leaving the state ζ_j (including "death" of the waiting process with coefficient ε_{ζ_j}), according to $q_{\zeta_j} = \sum_{\eta_j \neq \zeta_j} q_{\zeta_j,\eta_j} + \varepsilon_{\zeta_j}$, $w_j(0|\zeta_j)=1$.

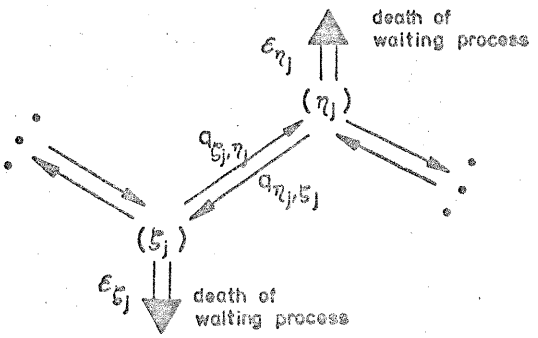

Fig. 6. Waiting process: General random walk diagram representation with transitions

The random process of waiting can be represented illustratively by multidimensional random walk diagrams, as generally shown in Fig. 6, [7,8]. Eq.(3.10) can be formulated in detail using a formal description for gradings and regarding the operation mode . For practical evaluations, the systems of linear differential equations can be suitably solved by methods of successive power series expansions up to orders of one third of the available computer storage capacity with a prescribed accuracy [7] .

3.2.2 Total pdf of waiting time

The total pdf of waiting time can be obtained by averaging over all cpdf's regarding the probabilities of states met at arrival (initial states) $p(\zeta_j)$:

$$W_j(>t) = P\left\{T_{Wj} > t\right\} = \sum_{\zeta_j \in Z_j} p(\zeta_j) w_j(t|\zeta_j) \ . \qquad (3.11)$$

The probabilities of initial states, $p(\zeta_j)$, are identical with the corresponding probabilities of state $p(\xi)$. In the limit case t=0, $W_j(>0)$ agrees with W_j of Eq.(3.3).

3.2.3 Mean waiting times and higher moments

From Eq.(3.10), corresponding systems of linear equations for the r-th ordinary moments of the cpdf's

$$m_{jr}(\xi_j) = -\int_{t=0}^{\infty} t^r dw_j(t|\xi_j) \qquad (3.12)$$

can be derived, including the mean waiting times $t_{Wj}(\xi_j) = m_{j1}(\xi_j)$ for waiting from initial state ξ_j. The total moments of waiting time referred to all waiting j-calls, m_{jr}, are obtained from the corresponding total pdf of m_{jr} waiting time referred to all waiting j-calls:

$$m_{jr} = -\int_{t=0}^{\infty} t^r d\frac{W_j(>t)}{W_j} = \frac{1}{W_j}\sum_{\xi_j \in Z_j} p(\xi_j) m_{jr}(\xi_j). \qquad (3.13)$$

The first moment m_{j1} agrees with the mean waiting time t_{Wj} according to Eq.(3.7).

3.3 Studies of the influence of grading parameters on the grade of service

In this section, two examples of exact calculations are given to study the influences of wiring methods and accessibility on the service characteristics.

3.3.1 Influence of wiring methods

As an example, the 6 main wiring methods shown in Fig.2 will be studied for a service system with n = 8 servers, accessibility k = 4, g = 4 grading groups, and s_j = 1 storage place for each grading group, j=1,2,3,4. The servers are hunted sequentially, the interqueue discipline is RANDOM.

The efficiency of the wiring method can be shown by comparison of the total probability of loss B versus the occupancy A/n ($A = \lambda/\varepsilon$ offered traffic, occupancy A/n = offered traffic per server), cf. Fig. 7 (Similar results hold also for the mean total queue length and the total probability of waiting W).

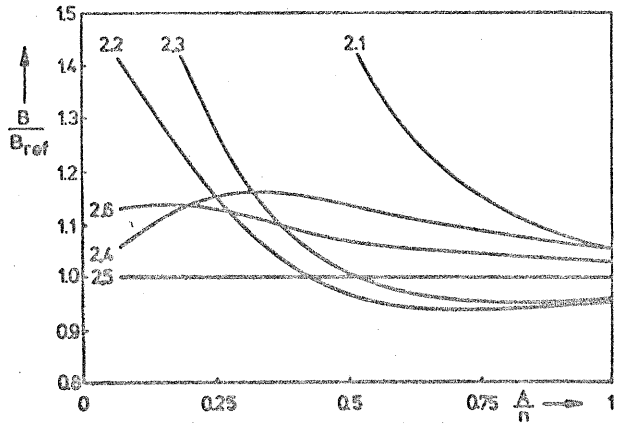

Fig. 7. Efficiency of wiring methods acc. to Fig. 2 with respect to loss probability B ($B_{ref} \triangleq$ Fig. 2.5)

As shown by Fig. 7, for small occupancies ($\frac{A}{n} < 0.4$) the straight inhomogeneous grading with progressive commoning and skipping is best, whereas for higher occupancies ($\frac{A}{n} > 0.4$) the straight homogeneous grading with skipping is best. Similar effects are already known from loss systems [11-21]. For delay systems with small occupancies, the optimum grading for a loss system will be the best, too. For higher occupancies, calls queue up and the termination process of all servers determines more and more the service quality: in this case, a grading with the best traffic balance is optimal; for given M, the optimum grading is a homogeneous one with a best possible traffic balance. Furthermore, the comparison of Figs. 2.2 and 2.3 shows that for sequential hunting slipping is worse than skipping. The optimum grading for a delay system, irrespective of the offered traffic, should therefore be a grading with certain progression and a considerable homogeneous part with skipping. Grading 2.5 forms a good compromise (cf. also simulation results Section 4.1.3).

3.3.2 Influence of accessibility k

The accessibility has the most important influence on the characteristic traffic values. This effect will be demonstrated for the various service systems shown in Fig. 3 having n = 6 servers, g = 3 grading groups, accessibilities k = 3 up to 6, and s_j = 4, j=1,2,3, with respect to the probability of waiting W and the mean waiting time of waiting calls referred to the mean service time $\tau_W = t_W/h$, cf. Fig. 8.

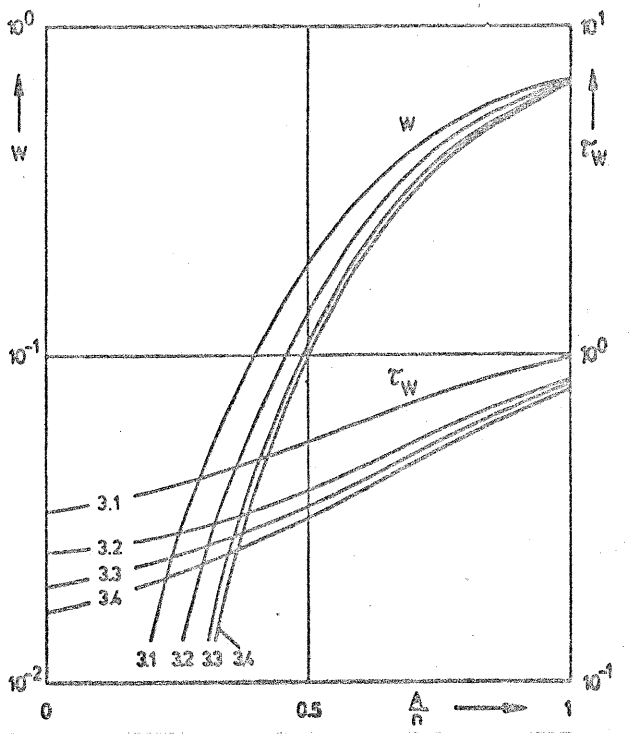

Fig. 8. Influence of the accessibility k
W and τ_W versus A/n
Service systems acc. to Fig. 3

In Fig. 9, the pdf's of waiting time for waiting calls, W(>t)/W, are shown for the accessibilities k = 4 and k = 6 for $\frac{A}{n}$ = 0.5, 1, and 1.5. The interqueue discipline is RANDOM, the queue discipline FIFO.

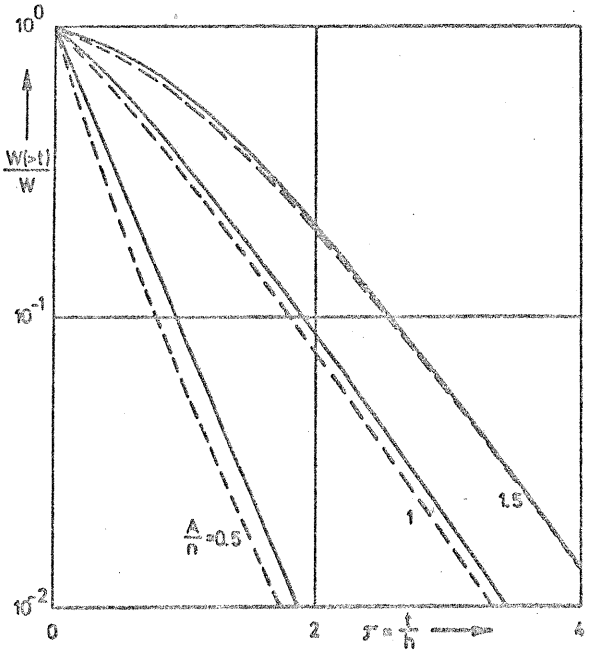

Fig. 9. Influence of the accessibility k on the pdf of waiting time
——— k = 4 (Fig. 3.2)
– – – – – k = 6 (Fig. 3.4)
Parameter: A/n

4. APPROXIMATE CALCULATION OF MULTI-QUEUE DELAY SYSTEMS WITH GRADINGS

For practical gradings, the number of unknowns is too large for exact calculations so that efficient approximation procedures are necessary. In this chapter, approximation methods are reported for mean values as well as for the pdf of waiting time.

4.1 The stationary state

4.1.1 The "Interconnection Delay Formula" (IDF)

For graded delay systems, M.THIERER[4] suggested a calculation method based on a two-dimensional state description (x,z), where x the number of busy servers, and z the total number of waiting calls. A part of the two-dimensional state space is shown in Fig.10.

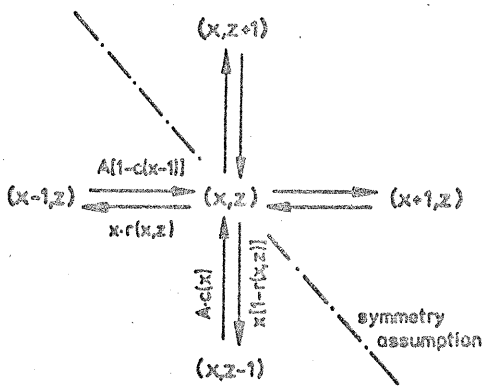

Fig. 10. Two-dimensional state space (x,z) and transitions

In Fig.10, c(x) means the blocking probability of the grading in state x, and r(x,z) a conditional probability indicating that no waiting call is served when a server becomes idle. By application of a special symmetry assumption for the statistical equilibrium, cf. Fig. 10, recursive equations were derived which led to explicit expressions for p(x), W, Ω, and τ_W:

$$p(x) = p(0)A^x \prod_{i=0}^{x-1}[1-c(i)] \Big/ \prod_{i=1}^{x}[i-Ac(i)], \qquad (4.1)$$

$$W = \sum_{x=k}^{n} p(x)c(x) , \qquad (4.2)$$

$$\Omega = \sum_{x=k}^{n} p(x) \sum_{i=k}^{x} \frac{Ac(i)}{i-Ac(i)} , \qquad (4.3)$$

$$\tau_W = \Omega/(AW) , \qquad (4.4)$$

with blocking probability

$$c(x) = \binom{x}{k}/\binom{n}{k} , \quad k \leq x \leq n, \qquad (4.5)$$

according to Ideal Erlang-Gradings [2]. As proved by extensive event-by-event simulations[4,23], above formulas yield good results for ideal gradings, ideally homogeneous gradings, and nonideal gradings with a relatively high mean interconnecting number M. Moreover, in these cases the simulation results turned out to be nearly independent of the interqueue disciplines FIFO and RANDOM, respectively.

4.1.2 Adaptation of the IDF to gradings of various type

Simulation results have shown that the results obtained by the IDF are too optimistic for real gradings with small M ($M \approx 2$). Additionally, the more realistic RANDOM-interqueue discipline even increases the results for W, Ω, and τ_W compared with the idealistic FIFO-interqueue discipline in case of gradings with progressive commoning. Both effects led to adaptation formulas for practical grading types with progressive commoning, cf. Figs. 4.1, 4.2, and 4.3.

The adaptation principle is based on a modification of the blocking probability c(x) in a graded delay system by means of a reduced accessibility $k^* \leq k$. In delay systems with gradings, the storage effect increases the probabilities of blocking occupation patterns compared to loss systems. Because of the progressive commoning, the servers at the last hunting steps with a large interconnecting number are less available for calls of unblocked grading groups which can be described approximately by a reduced mean accessibility k^*. Insertion of k^* into Eq.(4.5) leads to the modified blocking probability for delay systems:

$$c(x) = \binom{x}{k^*}/\binom{n}{k^*} . \qquad (4.6)$$

The principle of reduced mean accessibility has already been applied successfully to loss systems [22].

Intensive simulation runs for a large number of gradings with progressive commoning have shown that k^* is closely correlated with k, M, A, type of grading, and the interqueue discipline. By approximation of the mean queue length Ω, the following approximation formulas have been found:

$$k^* \cong k - \frac{n^2-k^2}{n^2}\left[\frac{k}{5}\left(\frac{A}{n}\right) + \frac{k-3}{2}\left(\frac{A}{n}\right)^4 + a\frac{k}{5}\left(\frac{A}{n}\right)^4\right]\frac{1}{M-1} \qquad (4.7a)$$

for standard gradings, simplified standard gradings, and high-efficiency gradings,

$$k^* \cong k - \frac{n^2-k^2}{n^2}\left[\frac{3}{4}k\left(\frac{A}{n}\right)^{\frac{3}{2}} + a\frac{k}{10}\left(\frac{A}{n}\right)^2\right]\frac{1}{M-1} \qquad (4.7b)$$

for O'Dell-gradings, where a =0 for the FIFO-, and a = 1 for the RANDOM-interqueue discipline. Both formulas hold for $M \geq 2$.

Remark:

The efficiency of the method by the reduced mean accessibility will be demonstrated by the practically interesting probability of waiting, W, and the mean waiting time referred to waiting calls, τ_W. Differences between calculations and simulations originate mainly from that fact, as k^* has been adapted to the mean queue length Ω (or the mean waiting time referred to all calls). The influences of k^* on W, Ω, and τ_W are indeed slightly different. The adaptation to Ω, however, yields the best results for the triple W, Ω, and τ_W. Another adaptation has been carried out to τ_W, by which the probability of waiting W got too much overestimated [23].

4.1.3 Studies of the influence of grading parameters and interqueue disciplines on the grade of service

In this section, the efficiency of gradings and the adaptation methods will be demonstrated for some examples.

Example 1: Influence of mean interconnecting number M, cf. Fig. 11

In the first example, two standard gradings having n = 30, k = 10, M = 2 and M = 3.33 are compared with each other, respectively. The interqueue discipline is FIFO.

The dotted lines indicate the results obtained for a c(x) for ideal gradings. The simulation results (with 95% confidence intervals) show the accuracy of the method.

Example 2: Influence of the grading type, cf. Fig. 12

This example shows the efficiency of a simplified standard grading compared with an O'Dell-grading, both having n = 60, k = 10, and M = 2. The interqueue discipline is FIFO.

Example 3: Influence of the interqueue discipline, cf. Fig. 13

This example demonstrates the influence of the two interqueue disciplines, FIFO and RANDOM, on W and τ_W for a simplified standard grading with n = 120, k = 10, and M = 2. The RANDOM-interqueue discipline yields, as usual for gradings with progressive commoning, values worse than FIFO for higher occupancies.

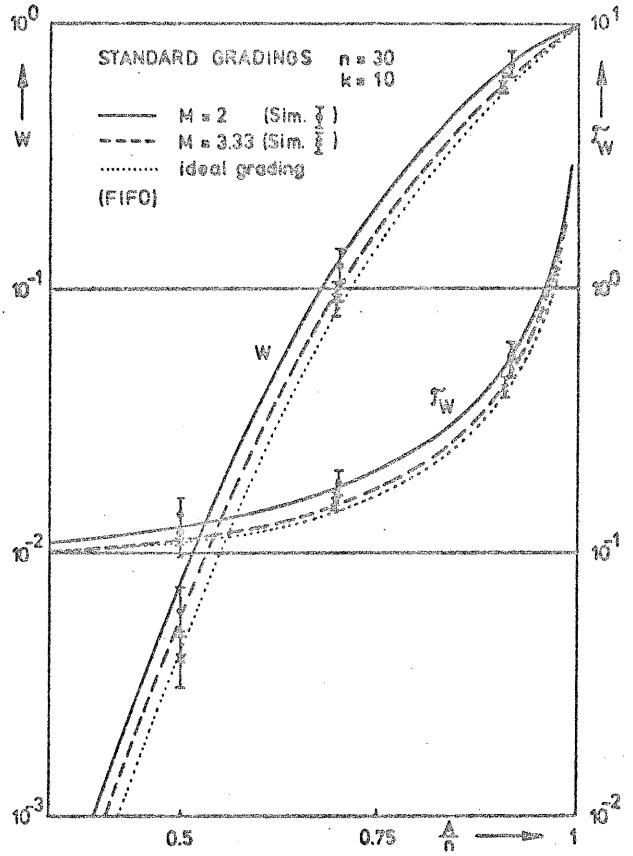

Fig. 11. Influence of the mean interconnecting number M
W and τ_W versus A/n

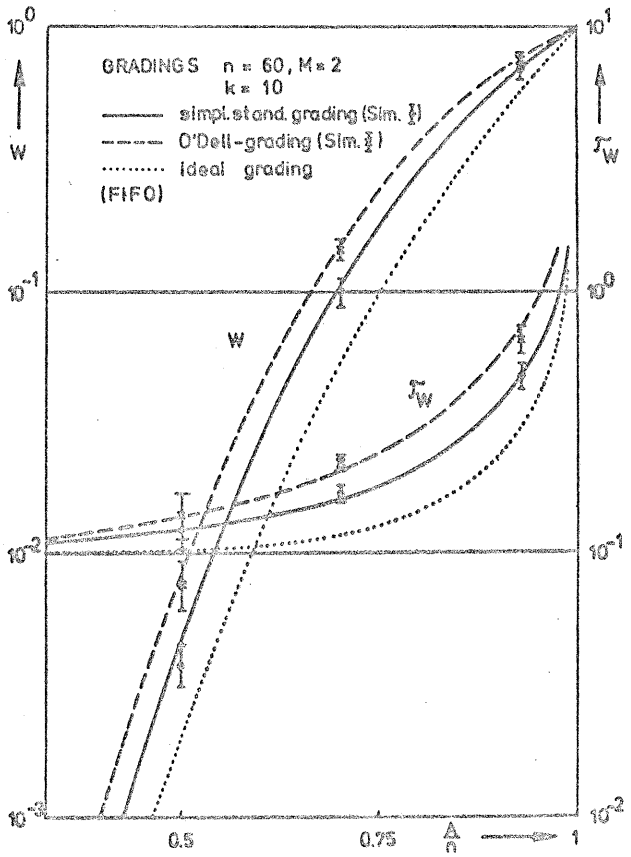

Fig. 12. Influence of the type of grading
W and τ_W versus A/n

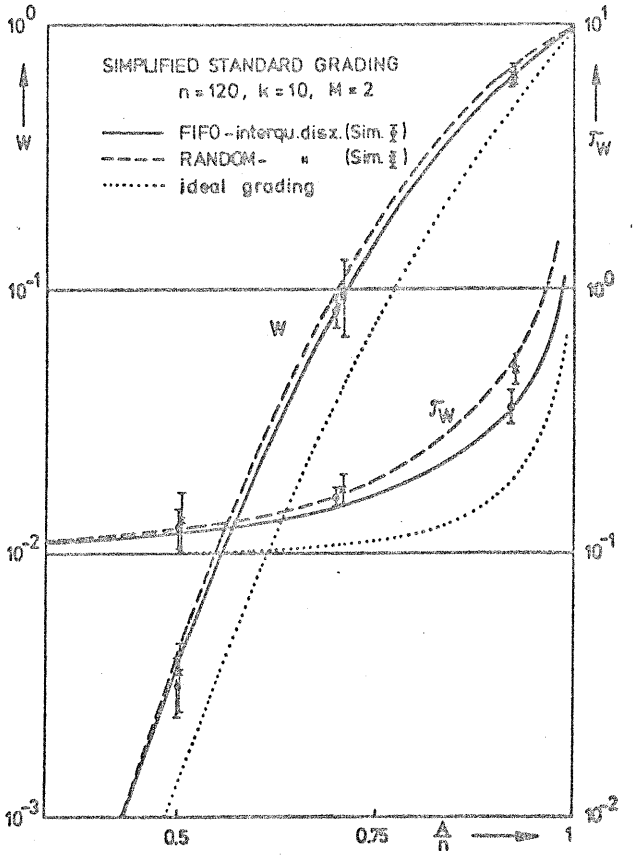

Fig. 13. Influence of the interqueue disciplines
FIFO and RANDOM
W and τ_W versus A/n

Example 4: Study on optimum gradings for delay systems

For the study of the influence of the wiring method on the efficiency of a grading, various grading types having n = 30, k = 10, g = 6 (M = 2) were investigated by simulations, namely

- O'Dell-grading acc. to Fig. 4.1
- Standard grading acc. to Fig. 4.2
- Homogeneous grading with slipping acc. to Fig. 14.1
- Homogeneous grading with skipping acc. to Fig. 14.2
- "Optimum grading for delay systems" acc. to Fig. 14.3

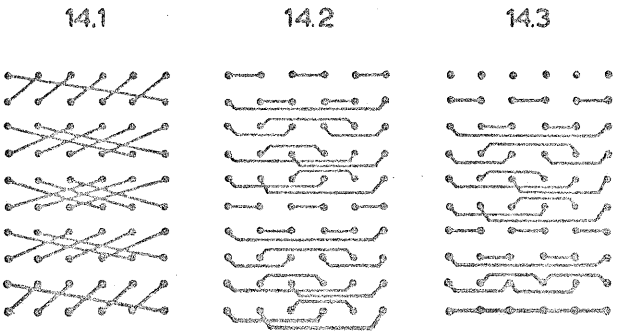

Fig. 14. Further grading types with
n = 30, k = 10, g = 6 (M = 2)

14.1 Homogeneous grading with slipping
14.2 Homogeneous grading with skipping
14.3 "Optimum grading for delay systems"

The grading type Fig. 14.3 was constructed according to the results of exact calculations, cf. Section 3.3.1, with certain progression and a large homogeneous part. This grading forms a compromise between the other gradings and yields the best efficiency over the whole range of occupancies ("optimum grading for delay system"), cf. Fig. 15.

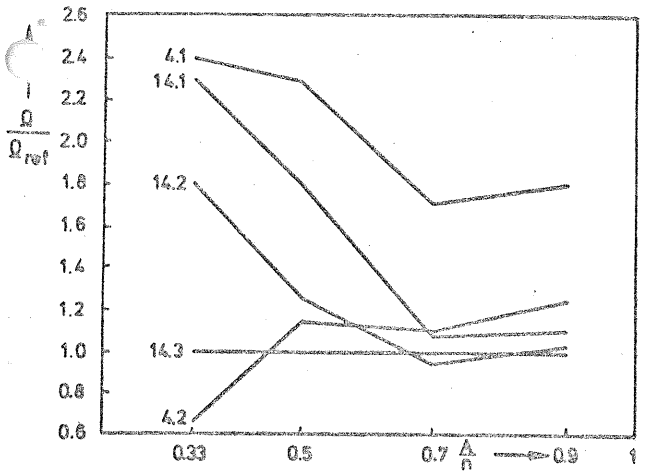

Fig. 15. Efficiency of various grading types with respect to Ω (Simulation)
n = 30, k = 10, g = 6 (M = 2)
Interqueue discipline: FIFO
($\Omega_{ref} \hat{=}$ Fig. 14.3)

Finally, it may be remarked that the interqueue disciplines FIFO and RANDOM yield practically the same results in case of homogeneous gradings (including ideal gradings!) while RANDOM is worse than FIFO in case of gradings with progressive commoning.

4.2 Distribution of waiting time

For the study of waiting time distributions in multiqueue delay systems, two different interqueue disciplines are considered: FIFO and RANDOM. Since the queue discipline is FIFO, there are two operation modes: FIFO/FIFO (F/F) and RANDOM/FIFO (R/F); the first one is identical with "FIFO with respect to all accessible waiting calls".

4.2.1 Approximation by exponential functions

4.2.1.1 Operation mode F/F

In delay systems with full accessibility, the waiting times referred to all waiting calls are negative-exponentially distributed according to

$$\frac{W(>\tau)}{W} = \exp\left(-\frac{\tau}{\tau_W}\right) . \qquad (4.8)$$

In a first approximation [4], this law can also be assumed for graded delay systems inserting the value for τ_W obtained from Eq.(4.4) for the considered grading type. Simulations have shown, however, that the actual pdf of waiting time is slightly hypoexponential, cf. Sections 4.2.2 and 4.2.3.

4.2.1.2 Operation mode R/F

This operation mode yields generally pdf's of waiting time with a hyperexponential character. To achieve larger values of the higher moments, it is assumed that not the total but the conditional waiting times for waiting from initial state x are negative-exponentially distributed each so that

$$\frac{W(>\tau)}{W} = \frac{1}{W}\cdot\sum_{x=k}^{n} p(x)c(x)\exp\left(-\frac{\tau}{\tau_x}\right) , \qquad (4.9)$$

where

$$\tau_x = \frac{1}{c(x)}\cdot\sum_{i=k}^{x}\frac{c(i)}{i-Ac(i)} \qquad (4.10)$$

the conditional mean waiting time for waiting from initial state x. Compared with simulation results, this solution fits well for practical purposes; the higher moments, however, are still underestimated, cf. Sections 4.2.2 and 4.2.3.

As a result, both approximations can be used for practical purposes. For a more detailed insight, however, the higher moments of the pdf of waiting time have been investigated and used for another approximation described in the following section.

4.2.2 Approximation by moments

Besides the mean waiting time τ_W (or first moment m_1), the pdf of waiting time is essentially characterized by the higher moments m_2 and m_3. These moments have been studied for various grading types and resulted in the following general characteristics.

4.2.2.1 Higher moments for operation mode F/F

For full accessibility, the pdf of waiting time and its moments are explicitly known, cf. Eq.(4.8). Within gradings, the following effects were observed by simulations:

- The pdf of waiting time behaves hypoexponential with increasing occupancy A/n and increasing ratio n/k.
- The higher moments m_2 and m_3 show only little (if at all) dependence on the mean interconnecting number M.
- The higher moments m_2 and m_3 are closely correlated.

When the first moment m_1 is known, the second moment m_2 can be well approximated for all grading types by

$$\frac{m_2}{m_1^2} \cong \frac{8}{4 + \left[1-\frac{k}{n}\right]\cdot\left(\frac{A}{n}\right)^3} , \qquad (4.11)$$

which yields the exact limit value 2 for $\frac{A}{n}\to 0$ or k=n.

For full accessibility, the exponential distribution of waiting times originates from an exact balance between the hypoexponential characteristic, given by the Erlang-distributions for waiting from an initial state, and the hyperexponential characteristic which is represented by the various possibilities for meeting initial states. In a graded delay system, it may happen that a queue exists even if there are idle servers; this effect increases the mean waiting time and the hypoexponential influence, as well.

4.2.2.2 Higher moments for operation mode R/F

For full accessibility, only two limiting cases are explicitly known for the pdf of waiting time: g = 1, which yields the normal FIFO-queue, and g (or M)$\to\infty$, which is equivalent to a normal RANDOM-queue. Simulations for full and limited accessible servers yielded the following effects:

- The pdf of waiting time behaves hyperexponential with increasing occupancy A/n.
- The hyperexponential behavior increases with increasing mean interconnecting number M.
- The higher moments m_2 and m_3 are closely correlated.

For all grading types, the following formula approximates the second moment m_2 by

$$\frac{m_2}{m_1^2} \cong \frac{4}{2 - \left(\frac{A}{n}\right)^{\frac{n}{k}}\left[1-M^{-\frac{2}{3}}\right]} , \qquad (4.12)$$

which yields the exact limits for $\frac{A}{n}\to 0$ or k=n and M=1 or M$\to\infty$, respectively.

By the RANDOM-selection of queues, the waiting times clearly vary more than in case of FIFO-selection; this effect is further enforced for increasing values of M.

4.2.2.3 Approximate pdf of waiting time

Knowing the first and second moments, the pdf of waiting time can be approximated for both operation modes by a gamma pdf:

$$\frac{W(>\tau)}{W} = 1 - \frac{\gamma(p, b\tau)}{\Gamma(p)} , \qquad (4.13)$$

where $\Gamma(p)$ the complete, and $\gamma(p,z)$ the incomplete gamma functions with $p = 1/(m_2/m_1^2 - 1)$ and $b = p/m_1$.

Eq.(4.13) yields the first and second moments, m_1 and m_2, exactly as they were approximated. Investigations of the third moment of Eq.(4.13), $m_3=p(p+1)(p+2)/b^3$, have shown, that the ratios m_3/m_3' obtained from Eq.(4.13) and simulations, differ less than 10% in most cases, respectively. Therefore, the most critical point for the accuracy of the approximated pdf of waiting time is the mean waiting time $\tau_W = m_1$.

4.2.3 Numerical results

The influence of the grading on the pdf of waiting time and the accuracy of the various approximation methods will be demonstrated for two examples. First, in Fig. 16 the pdf's of waiting time are shown for a simplified standard grading with n = 30, k = 6, and M = 2. In the second example of a standard grading having n = 30, k = 10, and M = 5, the first three moments of the pdf of waiting time are given in Table 1 together with the simulation results. Both examples show sufficient accuracy for the purpose of applications.

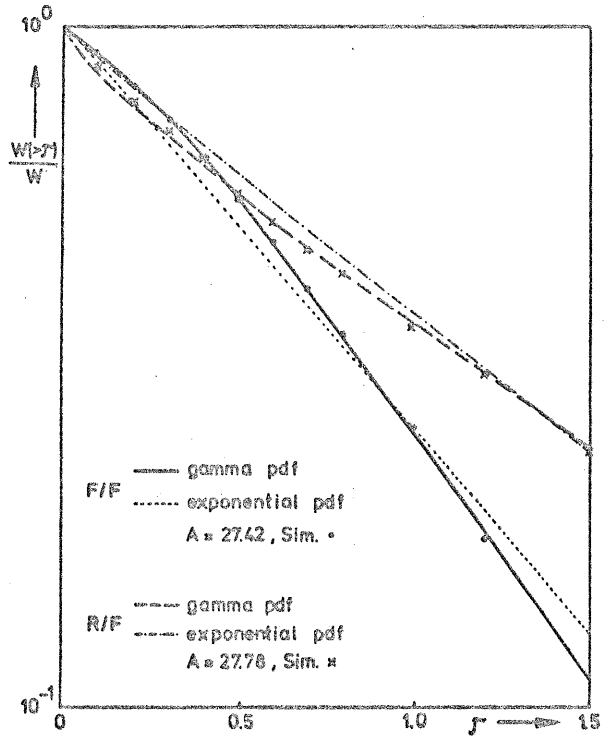

Fig. 16. Pdf of waiting time
Simplified standard grading n=30, k=6, M=2
Operation modes F/F and R/F

	F/F				R/F			
	A	m_1	m_2/m_1^2	m_3/m_1^3	A	m_1	m_2/m_1^2	m_3/m_1^3
C	14.91	0.1105	1.96	5.72	15.05	0.1112	2.08	6.62
S		0.1011	1.96	5.80		0.1138	2.12	7.00
C	20.95	0.1487	1.90	5.26	20.79	0.1479	2.24	7.84
S		0.1551	1.90	5.25		0.1628	2.19	7.44
C	27.45	0.4322	1.77	4.51	27.48	0.4413	2.69	11.70
S		0.4410	1.86	4.85		0.4529	2.81	14.10

Table 1. Moments of the pdf of waiting time
Standard grading, n = 30, k = 10, M = 5
Appr. Calculation (C) acc. to Section 4.2.2
Simulation (S)

CONCLUSION

For multi-queue delay systems with gradings, methods have been reported for exact and approximate calculation of mean values and the pdf of waiting time, as well. It has been shown that delay systems with gradings of various type can be calculated sufficiently accurate by the "Interconnection Delay Formula" introducing a modified blocking probability. The pdf of waiting time can be well approximated by a gamma pdf by the aid of its second moment. A number of examples has been given to study the influences of grading parameters and operation modes on the grade of service. Finally, for graded delay systems with a much varying load an optimum grading type has been suggested.

ACKNOWLEDGEMENTS

The author wishes to express his thanks to Prof. Dr.-Ing. A. Lotze and Dipl.-Ing. G. Kampe for supporting this work and many valuable discussions.

REFERENCES

[1] Lotze,A.: History and development of grading theory. AEÜ 25(1971), 402 - 410.

[2] Brockmeyer,E.,Halstrøm,H.L. and Jensen,A.: The life and works of A.K.Erlang. Transact.Danish Acad. Techn. Sci. No.2, Copenhagen, 1948.

[3] Gambe,E.: A study on the efficiency of graded multiple delay systems through artificial traffic trials. 3.ITC Paris 1961, Doc.16.

[4] Thierer,M.: Delay-tables for limited and full availability according to the Interconnection Delay Formula (IDF). 7th Report on Studies in Congestion Theory. Inst.for Switching and Data Technics, Univ. of Stuttgart, 1968.

[5] Thierer,M.: Delay systems with limited availability and constant holding time. 6.ITC München 1970, Congressbook, 322/1-6.

[6] Kühn,P.: Combined delay and loss systems with several input queues, full and limited accessibility. 6.ITC München 1970, Congressbook, 323/1-7.

[7] Kühn,P.: On the calculation of waiting times in switching and computer systems. 15th Report on Studies in Congestion Theory. Inst. for Switching and Data Technics, Univ. of Stuttgart, 1972.

[8] Herzog,U. and Kühn,P.: Comparison of some multiqueue models with overflow and load-sharing strategies for data transmission and computer systems. Symp. on Computer-Communications Networks and Teletraffic. Polytechnic Press of the Polytechnic Inst. of Brooklyn, New York, 1972.

[9] Gambe,E.,Suzuki,T. and Itoh,M.: Artificial traffic studies in a two-stage link system with waiting. 5.ITC New York 1967, Prebook, 351 - 359.

[10] Hieber,L.: About multi-stage link systems with queuing. 6.ITC München 1970, Congressbook,233/1-7.

[11] Longley,H.A.: The efficiency of gradings, Part I. POEEJ 41(1948), 45 - 49.

[12] Elldin,A.: On the congestion in gradings with random hunting. Ericsson Technics No.1 (1955), 35-94.

[13] Lotze,A.: Verluste und Gütemerkmale einstufiger Mischungen. NTZ 14(1961), 449 - 453.

[14] Helms,R. and Kuntze,W.: Erhöhung der Leistungsfähigkeit unvollkommener Bündel durch homogene Mischungen. Nachrichtentechnik 12(1962), 314 - 320.

[15] Bretschneider,G.: Die exakte Bestimmung der Verkehrsleistung kleiner unvollkommener Fernsprechbündel. NTZ 16(1963), 199 - 205.

[16] Hofstetter,H. and Trautmann,K.: Der Einfluß der Mischung auf die Verkehrsleistung der Abnehmerschaltglieder hinter einstufigen Vermittlungsanordnungen. NTZ 16(1963), 635 - 642.

[17] Rubas,J.: Survey of gradings and interconnecting schemes. The Telecommunication Journal of Australia (1965), 120 - 124.

[18] Stell,F.K.: Zweckmäßige Gestaltung von Mischungen. Wiss. Zeitschr. der Hochschule für Verkehrswesen "Friedrich List", Dresden. Part 1: 12(1965), 271 - 277, Part 2: 12(1965), 459 - 464.

[19] Vanek,N.: Ist die homogene Mischung wirklich am besten ? Nachrichtentechnik 15(1965), 213 - 218.

[20] Cappetti,I.: Simulation methods on telephone gradings: analysis of macrostructures and microstructures. 5.ITC New York 1967.

[21] Herzog,U.,Lotze,A. and Schehrer,R.: Calculation of trunk groups for simplified gradings. NTZ 22(1969), 684 - 689.

[22] Herzog,U.: Calculation of fully available groups and gradings for mixed pure chance traffic. NTZ 24(1971), 627 - 629.

[23] Kampe,G.,Kühn,P. and Ventouris,P.: Mean waiting times and distributions of waiting time in delay systems with real gradings. Inst. for Switching and Data Technics, Univ. of Stuttgart, Monograph No. 371, 1972.