COMPARISON OF SOME MULTIQUEUE MODELS WITH OVERFLOW AND LOAD-SHARING STRATEGIES FOR DATA TRANSMISSION AND COMPUTER SYSTEMS

Ulrich Herzog and Paul Kühn

Institute for Switching and Data Technics, University of Stuttgart, Stuttgart, GERMANY

This paper deals with different routing strategies for data transmission as well as different load-sharing and reliability configurations for computer systems.

Starting from two separate service systems without mutual overflow, a review of different overflow systems is given considering three types of overflow strategies applicable to data transmission. Furthermore, systems of two and more computers are inspected with various configurations and operating strategies. For reasons of simplicity, all multiqueue models are demonstrated by example of systems with two (limited) queues.

A comparison between the models of both application areas shows a close similarity with respect to system structure and operating strategies. Therefore, the models of both areas can be treated by the same mathematical methods.

The analysis of the different systems with respect to the service quality is carried out on the basis of the state equations under Markovian assumptions. Finally, the most important models will be compared with each other with respect to various traffic criteria.

I. INTRODUCTION

During the past few years, many investigations have been published about single-queue models for data transmission and computer systems under various operating strategies and traffic properties.

In modern communication networks, special routing strategies are used, such as alternative or adaptive routing [1-3]. Real-time computers are duplicated or operate in a load-sharing mode [4]. For such systems single-queue models are often not applicable in order to investigate their traffic behavior.

Data communication networks with different routing strategies as well as different reliability configurations are rather described by multiqueue models.

In Section II various configurations of many-server systems with two queues are discussed under three modes of overflow strategies:

(1) Overflow from primary to a secondary server group
(2) Overflow from a storage in front of a primary server group to a secondary server group

---

Presented at the Symposium on Computer-Communications Networks and Teletraffic
Polytechnic Institute of Brooklyn, April 4–6, 1972.

(3) Overflow from primary storage to a secondary storage *or* directly to a secondary server group.

All three overflow strategies can be applied to data transmission networks as well as to load-shared or breakdown-reliable computer systems.

Section III gives an outline of the mathematical analysis of such overflow systems to investigate the capability of the different configurations and strategies. For the analysis Markovian assumptions are assumed. Solutions are given either by exact evaluation of systems of equations or by approaches using two moments of the overflow traffic.

In Section IV comparisons are made for different system configurations and strategies. Numerical results are presented for probabilities of waiting and loss, mean waiting times, carried traffic (throughput) as well as the distribution of waiting times.

II. MULTIQUEUE MODELS FOR DATA TRANSMISSION AND COMPUTER SYSTEMS

In Figs. 1 and 2 some configurations of multiserver systems with two queues are reviewed which are able to operate under various overflow and load-sharing strategies. The models in Fig. 1 are suited for data transmission systems; the models in Fig. 2 are applicable to systems of computers.

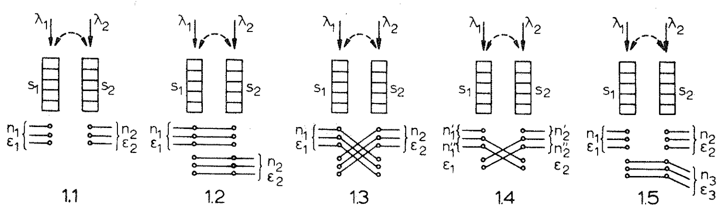

Fig. 1. Configurations of multiserver systems with two queues and overflow capability for data transmission systems.

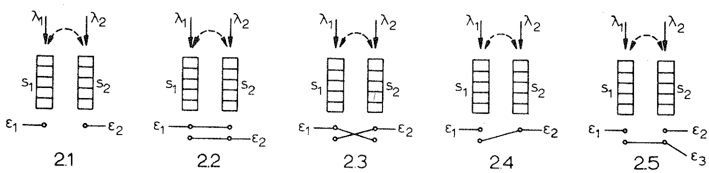

Fig. 2. Configurations of multiserver systems with two queues and load-sharing capability for computer systems.

2.1 Multiqueue Server Configurations

The model in Fig. 1.1 shows two (separate) queuing systems for the traffic to direct routes 1 and 2 with n_r servers (trunks, lines) for route r ($r = 1, 2$). A call of an incoming group j is allowed to wait in a storage with s_j waiting places ($j = 1, 2$) when the servers of the corresponding direct route are blocked and if there is at least one waiting place available. In Figs. 1.2 and 1.3 two different configurations of multiserver systems are shown with two queues and "fully accessible" servers. Assuming a sequential hunting mode, route number 1 of Fig. 1.2 represents a high usage route which carries most of the traffic while route number 2 takes only the "overflow" traffic. In Fig. 1.3 both routes number 1 and 2 are working as high usage routes for their own offered traffic; in addition, the overflow traffic of the other route is offered to servers of each route. In the model of Fig. 1.4, a part of each high usage route is reserved for its own offered traffic, the residual servers are allowed to also carry overflow traffic. Finally, the model in Fig. 1.5 shows a configuration where the traffic for two direct routes can overflow to a third ("final") route number 3. Both models in Figs. 1.4 and 1.5 incorporate "limited accessible" servers.

The configurations of Figs. 2.1 to 2.5 are very similar to the models of Figure 1. The servers may be considered as computers serving calls (real-time requests or batch jobs) stored within the queues, respectively. All the different configurations with mutual aid allow better utilization, load-sharing, and reliability with respect to breakdown.

2.2 Overflow Strategies

The multiqueue server configurations discussed in Section 2.1 (Figs. 1 and 2) may operate under different overflow strategies. Three main modes will be considered:

- Overflow Strategy 1 (S1):
  Overflow from primary (direct) to a secondary (final) server group
  Waiting is only allowed if all accessible servers of the primary *and* secondary server group are occupied
- Overflow Strategy 2 (S2):
  Overflow from a storage in front of a primary server group to a secondary server group
  The accessible servers of the secondary server group are only hunted if both the primary server group and the primary storage are fully occupied
- Overflow Strategy 3 (S3):
  Overflow from primary storage to a secondary storage or directly to a secondary server group
  The call, which finds the own server group and all accessible servers of other server groups occupied, queues in the storage of its incoming group. However, if there is no free waiting place, the call is diverted to the other incoming group and is treated there like an original call of this group.

These different overflow strategies have a significant influence on the traffic criteria, in particular the probabilities of loss and waiting, as well as the mean waiting times. In the following sections, the analysis of these queuing systems is outlined and the most important configurations are compared with each other with respect to the overflow strategies in question.

III. MATHEMATICAL ANALYSIS

This section gives an outline of the mathematical analysis of multiqueue models under different overflow strategies. In Section 3.1 the exact calculation of multiqueue models is shown for Overflow Strategy 1. Sections 3.2 and 3.3 handle multiqueue models for Overflow Strategies 2 and 3 based on a method regarding two moments of the overflow traffic, respectively. For the mathematical analysis, the Markovian properties are assumed throughout the paper.

3.1 Analysis of Multiqueue Models for Overflow Strategy 1

3.1.1 Model

The general multiqueue model with overflow from primary servers (e.g., trunks of high usage routes, highly used computers, etc.) to secondary servers (e.g., trunks of final routes, remote computers, etc.) can generally be considered as a service system with full or limited accessibility. The special assignment of servers to incoming groups and outgoing routes is determined by a "grading matrix." In the most general case, each server represents a different route; by combination of various servers to outgoing groups all possible configurations can be obtained from the general case.

The structure of the multiqueue overflow model is laid down by the following parameters:

- g : Number of incoming groups or input queues
- n : Number of servers
- k_j : Accessibility within incoming group j $(j = 1, 2, 3, \cdots, g)$
- s_j : Number of waiting places available for calls of incoming group j $(j = 1, 2, 3, \cdots, g)$
- $\|g_{hj}\|$: Grading matrix, where g_{hj} is the number of that server which is hunted at step (order) h within incoming group j $(h = 1, 2, 3, \cdots, k_j; j = 1, 2, 3, \cdots, g)$
- $\|a_{rj}\|$: Accessibility matrix, where $a_{rj} = 1(0)$ if incoming group j has (has no) access to server number r $(r = 1, 2, 3, \cdots, n; j = 1, 2, 3, \cdots, g)$.

The operation mode is characterized by the following criteria:

- Sequential hunting of all accessible servers within each incoming group (Overflow Strategy $S1$)
- First-in; first-out (FIFO) service within each queue (queue discipline)
- Arbitrary probability law for service between the queues (interqueue discipline).

The interqueue discipline is fully determined by

- $\|p_{rj}\|$: Interqueue discipline matrix, where p_{rj} is the probability that queue j will be served when server r finishes service $(r = 1, 2, 3, \cdots, n; j = 1, 2, 3, \cdots, g)$.

The interqueue discipline matrix is in general state-dependent. By choice of special matrix elements, however, special cases are obtained; for instance, the case of nonpreemptive priorities between the queues.

Input and termination processes are Markovian, i.e., the interarrival times a and service times b are exponential

$$A_j(t) = \Pr[a_j \leq t] = 1 - \exp(-\lambda_j t) \quad (j = 1, 2, 3, \cdots, g) , \tag{1.1}$$

$$B_r(t) = \Pr[b_r \leq t] = 1 - \exp(-\epsilon_r t) \quad (r = 1, 2, 3, \cdots, n) . \tag{1.2}$$

For illustration, Fig. 3 shows a simple 3-server system with two queues.

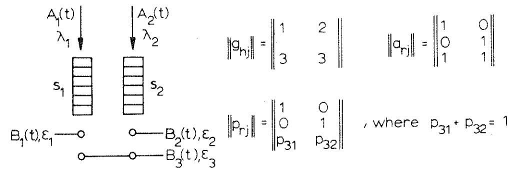

Fig. 3. Example of a 3-server overflow system with two queues.

3.1.2 Principle of Solution

At first, the occupations of servers and waiting places will be described by a multidimensional state. For the probabilities of state, the linear system of equations (Kolmogorov forward equation) is derived in Section 3.1.3. By means of the state probabilities, characteristic traffic values (mean values) are defined in Section 3.1.4. For the treatment of waiting time distributions, a waiting process is considered which is constructed from the process of system states. In Section 3.1.5 the linear system of differential equations (Kolmogorov backward equation) for the conditional distribution functions (cdfs) of waiting time is discussed. The total probability distribution function (pdf) of waiting time is found by regarding the probabilities of initial states (where the waiting process starts) combined with the corresponding cdf of waiting time, cf Section 3.16.

3.1.3 Stationary Probabilities of State

The system state ξ may be defined as

$$\xi = (x_1, x_2, x_3, \cdots, x_r, \cdots, x_n; z_1, z_2, z_3, \cdots, z_j, \cdots, z_g) \qquad (\xi \in \Xi) \;, \qquad (1.3)$$

where

$$x_r = \begin{cases} 0, & \text{server } r \text{ is idle} \;, \\ 1, & \text{server } r \text{ is occupied} \qquad (r = 1, 2, 3, \cdots, n) \;, \end{cases}$$

z_j : Number of occupied waiting places within queue j,
$z_j \in [0, s_j] \qquad (j = 1, 2, 3, \cdots, g)$.

In Eq. (1.3) not all possible patterns are physically realizable: a queue j can only be built up if all accessible k_j servers are blocked $(j = 1, 2, 3, \cdots, g)$.

The stationary probabilities of states $p(\xi)$ can be determined by the Kolmogorov forward equations

$$q_\xi \, p(\xi) - \sum_{\pi \neq \xi} q_{\pi\xi} \, p(\pi) = 0 \qquad (\xi \in \Xi) \;, \qquad (1.4)$$

which are found by considering all states ξ in statistical equilibrium with their neighbor states [5, 6]. In Eq. (1.4) $q_{\xi\pi}$ means the transition coefficient for transition from state ξ to state π, and $q_\xi = \sum_{\pi \neq \xi} q_{\xi\pi}$.

Application of the statistical equilibrium to the general state Eq. (1.3) results in the following equation

$$\left[\sum_{j=1}^{g} (1-\delta_{z_j,s_j})\lambda_j + \sum_{r=1}^{n} x_r\,\epsilon_r\right] p(\cdots, x_r, \cdots; \cdots, z_j, \cdots)$$

$$= \sum_{r=1}^{n} (1-x_r)\,\epsilon_r\, p(\cdots, x_r+1, \cdots; \cdots, z_j, \cdots) +$$

$$+ \sum_{j=1}^{g} (1-\delta_{z_j,s_j})\,\delta_{\kappa_j,k_j}\left(\sum_{r=1}^{n} \epsilon_r\, p_{rj}\right) p(\cdots, x_r, \cdots; \cdots, z_j+1, \cdots) + \qquad (1.5)$$

$$+ \sum_{j=1}^{g} \delta_{z_j,0}\,\lambda_j \left[\sum_{h=1}^{kj} \prod_{\kappa=1}^{h} (x_{g_{\kappa j}})\; p(\cdots, x_{g_{hj}}-1, \cdots; \cdots, z_j, \cdots)\right] +$$

$$+ \sum_{j=1}^{g} (1-\delta_{z_j,0})\,\lambda_j\; p(\cdots, x_r, \cdots; \cdots, z_j-1, \cdots) \qquad (\xi \in \Xi)\;,$$

where: $\delta_{i,j}$ is the Kronecker delta; and $\kappa_j = \sum_{r=1}^{n} a_{rj}x_r$: the number of occupied servers in group j. In Eq. (1.5) all probabilities of physically not possible states are zero.

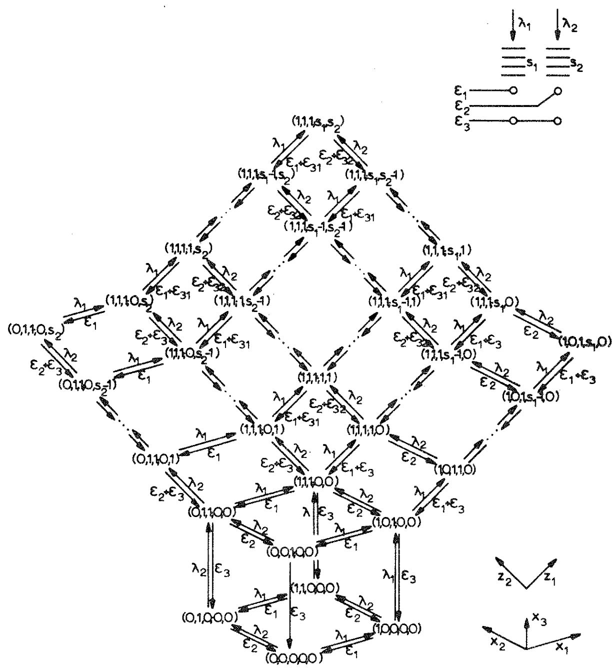

Fig. 4. State space and transitions for a 3-server overflow system with two queues according to Figure 3. System state $(x_1, x_2, x_3; z_1, z_2)$.

To illustrate Eq. (1.5), in Fig. 4 the 5-dimensional state space with transition coefficients is shown for the 3-server system of Figure 3. In Fig. 4, $\epsilon_{rj} = \epsilon_r p_{rj}$ is used as an abbreviation for the service rate of queue j with respect to server r.

The linear system of equations (Eq. (1.5)) is generally solved by an iterative method (successive overrelaxation) and normalized by the condition that the sum of all state probabilities equals unity. In some special cases of fully accessible servers, an explicit solution or recursion algorithms can be derived [7].

3.1.4 Characteristic Traffic Values

The most important mean values are the probabilities of waiting and loss, the carried traffics, the mean queue lengths, and mean waiting times. All these values can be derived from the probabilities of state.

(a) Probability of waiting W_j

$$W_j = \sum_{x_1=0}^{1} \cdots \sum_{x_n=0}^{1} \sum_{z_1=0}^{s_1} \cdots \sum_{z_j=0}^{s_j-1} \cdots \sum_{z_g=0}^{s_g} \prod_{h=1}^{k_j} (x_{g_{hj}})\, p\,(\xi) \ . \tag{1.6}$$

(b) Probability of loss B_j

$$B_j = \sum_{x_1=0}^{1} \cdots \sum_{x_n=0}^{1} \sum_{z_1=0}^{s_1} \cdots \sum_{z_g=0}^{s_g} \delta_{z_j,\, s_j}\, p\,(\xi) \ . \tag{1.7}$$

(c) Carried traffic Y_r

$$Y_r = \sum_{x_1=0}^{1} \cdots \sum_{x_n=0}^{1} \sum_{z_1=0}^{s_1} \cdots \sum_{z_g=0}^{s_g} x_r\, p\,(\xi) \ . \tag{1.8}$$

(d) Mean queue length Ω_j

$$\Omega_j = \sum_{x_1=0}^{1} \cdots \sum_{x_n=0}^{1} \sum_{z_1=0}^{s_1} \cdots \sum_{z_g=0}^{s_g} z_j\, p\,(\xi) \ . \tag{1.9}$$

(e) Mean waiting time t_{wj} referred to all waiting j-calls

$$t_{wj} = \frac{\Omega_j}{\lambda_j W_j} \ . \tag{1.10}$$

3.1.5 Conditional Distribution Functions of Waiting Time

For the exact calculation of waiting time distributions, the waiting process for a test call will be considered within the j^{th} queue. A j-call enters the j-queue and starts a waiting process; this process is being "alive" as long as the j-call is waiting and "dies" at the moment the j-call is selected for service. The 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.

To describe the waiting process of j-calls, a waiting state ζ_j is introduced. The waiting state ζ_j is built from the occupation states x_r of all those servers, which have no access to group j, and the states i_ν of all queues. When the interqueue discipline does not depend on the actual length of the queues, the waiting time of the j-test call is not influenced by subsequent arriving j-calls (FIFO). In such cases, the waiting state ζ_j can be defined by a $(n - k_j + g)$-dimensional variable

$$\zeta_j = (\cdots, x_r, \cdots; \cdots, i_\nu, \cdots) \qquad (r \neq g_{hj}; h = 1, 2, 3, \cdots, k_j; \; \nu = 1, 2, 3, \cdots, g, \zeta_j \in Z_j) \;, \qquad (1.11)$$

where

$$x_r = \begin{cases} 0, & \text{server } r \text{ is idle} \;, \\ 1, & \text{server } r \text{ is occupied} \;, \end{cases}$$

i_ν : Number of waiting calls within queue ν $(\nu \neq j)$

i_j : Number of calls waiting in front of the j-test call.

For the waiting time w_j of a j-test call, which met an arbitrary state ζ_j at arrival, a conditional *complementary* distribution function is defined

$$P_j^c(t|\zeta_j) = \Pr\{w_j > t|\zeta_j\} \qquad (\zeta_j \in Z_j) \;. \qquad (1.12)$$

For the cdf of waiting time a linear system of differential equations (Kolmogorov backward equation) holds [7–9]

$$\frac{d P_j^c(t|\zeta_j)}{dt} = -q_{\zeta_j} P_j^c(t|\zeta_j) + \sum_{\eta_j \neq \zeta_j} q_{\zeta_j, \eta_j} P_j^c(t|\eta_j) \qquad (\zeta_j, \eta_j \in Z_j), \qquad (1.13)$$

with initial conditions $P_j^c(0|\zeta_j) = 1$ $(\zeta_j \in Z_j)$. In Eq. (1.13) q_{ζ_j, η_j} means the transition coefficient for transition of waiting state ζ_j to waiting state η_j; q_{ζ_j} is the coefficient for leaving the state ζ_j, including the "death" of the waiting process.

To illustrate Eq. (1.13), the state space of the waiting process is given in Fig. 5 for the 3-server system of Fig. 3 with respect to 1-calls.

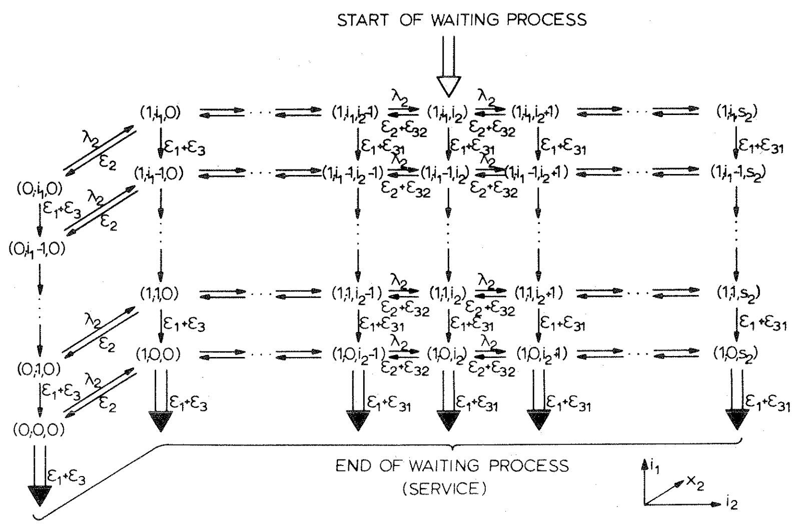

Fig. 5. State space and transitions for the waiting process of 1-calls for a 3-server system with two queues according to Figure 3. Waiting state $\zeta_1 = (x_2; i_1, i_2)$. Initial state (example) $\zeta_1 = (1; i_1, i_2)$

For the example of Fig. 5, the differential equation for the initial state $\zeta_1 = (x_2; i_1, i_2) \quad (x_2 = 1; i_1 > 0; 0 < i_2 < s_2)$, reads as follows

$$\frac{d P_1^c(t|1; i_1, i_2)}{dt} = -(\lambda_2 + \epsilon_1 + \epsilon_2 + \epsilon_3) P_1^c(t|1; i_1, i_2) + \lambda_2 P_1^c(t|1; i_1, i_2 + 1)$$
$$+ (\epsilon_1 + \epsilon_{31}) P_1^c(t|1; i_1 - 1, i_2)$$
$$+ (\epsilon_2 + \epsilon_{32}) P_1^c(t|1; i_1, i_2 - 1) . \tag{1.14}$$

At the bottom of Fig. 5, the coefficients for termination ("death") of the waiting process are also shown.

A detailed discussion of waiting processes for various queue and interqueue disciplines has been reported by Kühn [7]. The linear systems of differential equations can be suitably solved by a method of successive power series expansions even for a high order of the differential equation system and prescribed accuracy [7]. By integration of Eq. (1.13), the corresponding linear equations for the conditional mean waiting times and higher moments can be obtained as well [7, 8].

3.1.6 Total Distribution Function of Waiting Time

The pdf of waiting time can be obtained by averaging over all cdfs combined with the probabilities of initial states (conditions)

$$P_j^c(>t) = \Pr[w_j > t] = \sum_{\zeta_j \in Z_j} p(\zeta_j) P_j^c(t|\zeta_j) . \tag{1.15}$$

The probabilities of initial states $p(\zeta_j)$ are identical with the corresponding probabilities of state $p(\xi)$; the difference between the states ξ and ζ_j originates from the k_j (occupied) servers accessible from incoming group j.

3.2 Analysis of Multiqueue Models for Overflow Strategy 2

3.2.1 Model

Given a structure as shown in Fig. 6: at first, the traffic $A_j \quad (j = 1, 2)$ is offered to a primary arrangement consisting of n_j servers and s_j waiting places. If there is blocking, i.e., all primary servers and all waiting places are occupied, calls will be diverted to a secondary group (overflow group) with n_3 servers. Calls waiting for service in a queue

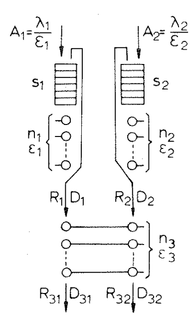

Fig. 6. The model.

are served in the order of their arrival (FIFO). The service rates in the different server groups may be different.

3.2.2 Principle of Solution

In principle, an exact solution for the characteristic traffic values is possible: introducing three-dimensional state probabilities, the equations of state can be found relatively easily and the evaluation may be done by a relaxation method [10]. However, for arrangements with a large number of trunks and waiting places as well as for structures with more than two primary arrangements, this evaluation is not possible, even on the largest digital computers. Therefore, a handy approximate method is suggested, using the fundamental idea of the so-called "substitute primary arrangements" [11–15]:

(a) All overflow traffics are characterized by their first and second moment (mean value R and variance V or variance coefficient $D = V - R$, respectively). Moments of third and higher order are neglected. The exact calculation of overflow traffic moments is outlined later on in Section 3.2.3

(b) Because all traffics overflowing from different primary arrangements are independent of each other, the total overflow traffic, offered to the common secondary group, can be described by the sum of all mean values R_j and the sum of all variance coefficients D_j. Therefore, one gets in case of two primary arrangements as shown in Fig. 6

$$\begin{aligned} \overline{R} &= R_1 + R_2 , \\ \overline{D} &= D_1 + D_2 \end{aligned} \tag{2.1}$$

(c) In order to calculate the traffic characteristics of the secondary group, a "substitute primary arrangement" and a "generating traffic" A^* are determined such that an overflow traffic is generated with mean value $\overline{R}$ and variance coefficient $\overline{D}$ (cf. Figure 7). In other words: all actual traffics overflowing from the

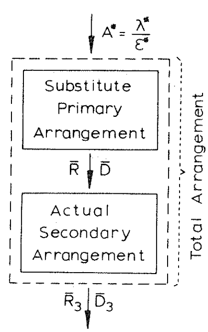

Fig. 7. Replacement of the real traffic by an equivalent traffic.

various primary arrangements are described approximately by one substitute overflow traffic with the same mean $\overline{R}$ and variance $\overline{V}$. Then all characteristic values of the actual secondary group are calculated, taking into account the structure of the substitute primary arrangement as well as the generating traffic A^*.

Artificial traffic trials performed on a digital computer have shown that for practical applications it is unimportant whether the substitute arrangement is a fully available group with or without waiting places. Therefore, by reason of simple evaluations, a fully available group with n^* servers, offered traffic A^*, and termination rate $\epsilon^* = \epsilon_3$ is chosen as substitute primary arrangement (n^* and A^* have to be determined by iteration such that the overflow traffic $(\overline{R}, \overline{D})$ is generated).

Now, obviously, the traffic $(\overline{R}_3, \overline{D}_3)$ rejected also by the secondary group is given by [11-13, 15]

$$\left.\begin{aligned} \overline{R}_3 &= A^* E_{1,n^*+n_3}(A^*) , \\ \overline{D}_3 &= \overline{R}_3^2 \left[\frac{1}{E_{1,n^*+n_3}(A^*) \{n^* + n_3 + 1 - A^* + R_3\}} - 1 \right] , \end{aligned}\right\} \tag{2.2}$$

where

$$E_{1,n}(A) = \frac{\dfrac{A^n}{n!}}{\displaystyle\sum_{i=0}^{n} \frac{A^i}{i!}} .$$

3.2.3 Calculation of the moments of overflow traffic

Given a fully available primary arrangement with n_j servers and s_j waiting places ($j = 1, 2$) and an infinite secondary arrangement ($n_3 \longrightarrow \infty$). If there is offered random traffic (Poisson input and negative exponential service times, cf. Eqs. (1.1) and (1.2)) to the primary arrangement, the mean value R_j of the traffic overflowing to the secondary group with termination rate ϵ_3 is given by the well known formula [16]

$$R_j = \frac{\lambda_j}{\epsilon_3} \frac{\dfrac{A_j^{n_j}}{n_j!}\left(\dfrac{A_j}{n_j}\right)^{s_j}}{\displaystyle\sum_{x=0}^{n_j-1} \frac{A_j^x}{x!} + \frac{A_j^{n_j}}{n_j!} \frac{1 - \left(\dfrac{A_j}{n_j}\right)^{s_j+1}}{1 - \dfrac{A_j}{n_j}}} , \tag{2.3}$$

where

$$A_j = \lambda_j / \epsilon_j .$$

Basharin [17] seems to be the first dealing with the variance of overflow traffic in delay-loss systems: recurrent formulae are presented for the computation of the moments of overflow traffic if there are $s_j \geq 1$ waiting places and uniform termination rate $\epsilon_j = \epsilon_3 = \epsilon$ for both primary and (infinite) secondary arrangement.

In the following it is concisely outlined how to calculate recursively all overflow traffic moments even if there are different termination rates ϵ_j and ϵ_3. For uniform termination rates also an *explicit* solution is presented for the variance V_j or variance coefficient D_j, respectively.

The calculation has been performed following a way of solution which is similar to that one successfully applied for overflow systems without waiting places [18, 19]:

(a) Defining with $\{u_j, u_3\}$ the state that there are simultaneously u_j demands being served or waiting for service in the primary arrangement and u_3 demands in the secondary system, it is possible to find the following equations of state for the state probabilities $p\,(u_j, u_3)$

$u_j < n_j$:

$$\lambda_j\, p\,(u_j - 1, u_3) + (u_3 + 1)\,\epsilon_3\, p\,(u_j, u_3 + 1) + (u_j + 1)\,\epsilon_j\, p\,(u_j + 1, u_3) = (u_3\,\epsilon_3 + u_j\,\epsilon_j + \lambda_j)\, p\,(u_j, u_3)\ , \tag{2.4a}$$

$n_j \leq u_j < n_j + s_j$:

$$\lambda_j\, p\,(u_j - 1, u_3) + (u_3 + 1)\,\epsilon_3\, p\,(u_j, u_3 + 1) + n_j\,\epsilon_j\, p\,(u_j + 1, u_3) = (u_3\,\epsilon_3 + n_j\,\epsilon_j + \lambda_j)\, p\,(u_j, u_3)\ , \tag{2.4b}$$

$u_j = n_j + s_j$:

$$\lambda_j\, p\,(n_j + s_j - 1, u_3) + \lambda_j\, p\,(n_j + s_j, u_3 - 1) + (n_3 + 1)\,\epsilon_3\, p\,(n_j + s_j, u_3 + 1) = (u_3\,\epsilon_3 + n_j\,\epsilon_j + \lambda_j)\, p\,(n_j + s_j, u_3) \tag{2.4c}$$

(b) Solving this system of linear equations, the basic idea is to introduce factorial moments $M_k\,(u_3)$, conditional factorial moments $M_k\,(u_3 \,|\, u_j)$, and conditional moment generating functions $F\,(u_3 \,|\, u_j, t)$ for the overflow traffic

$$M_k\,(u_3) = \sum_{u_j=0}^{n_j+s_j} M_k\,(u_3 \,|\, u_j) = \sum_{u_j=0}^{n_j+s_j} \sum_{u_3=0}^{\infty} k!\binom{u_3}{k}\, p\,(u_j, u_3)\ , \tag{2.5}$$

$$F\,(u_3 \,|\, u_j, t) = \sum_{k=0}^{\infty} M_k\,(u_3 \,|\, u_j)\,\frac{t^k}{k!} = \sum_{u_3=0}^{\infty} (1 + t)^{u_3}\, p\,(u_j, u_3)\ . \tag{2.6}$$

All normal moments $m_k\,(u_3)$ of the overflow traffic can be expressed by factorial moments, especially

$$\begin{aligned} k = 0 &: m_0\,(u_3) = M_0\,(u_3) = 1\ , \\ k = 1 &: m_1\,(u_3) = M_1\,(u_3) = R_j\ , \\ k = 2 &: m_2\,(u_3) = M_2\,(u_3) + m_1\,(u_3)\ , \end{aligned} \tag{2.7}$$

and the variance coefficient

$$D_j = M_2\,(u_3) - R_j^2\ . \tag{2.8}$$

Therefore, in order to determine the variance coefficient D_j we have to calculate the factorial moment of second order

(c) By analogy to the Laplace transform [20] it is possible to transform Eqs. (2.4a, b, c) into corresponding equations for the generating function. Then using Eq. (2.6) it is possible to find, by comparing coefficients, the following equations for the conditional factorial moments

$u_j < n_j$:

$$-\lambda_j\, M_k\,(u_3 \,|\, u_j - 1) + (u_j\,\epsilon_j + k\,\epsilon_3 + \lambda_j)\, M_k\,(u_3 \,|\, u_j) - (u_j + 1)\,\epsilon_j\, M_k\,(u_3 \,|\, u_j + 1) = 0\ , \tag{2.9a}$$

$$n_j \leq u_j < n_j + s_j :$$

$$- \lambda_j M_k (u_3 | u_j - 1) + (n_j \epsilon_j + k \epsilon_3 + \lambda_j) M_k (u_3 | u_j) - n_j \epsilon_j M_k (u_3 | u_j + 1) = 0 , \quad (2.9b)$$

$$u_j = n_j + s_j :$$

$$- \lambda_j M_k (u_3 | n_j + s_j - 1) + (n_j \epsilon_j + k \epsilon_3) M_k (u_3 | n_j + s_j) = k \lambda_j M_{k-1} (u_3 | n_j + s_j) . \quad (2.9c)$$

Summing up Eqs. (2.9a, b, c) over all possible values of u_j we find directly the following equation

$$\epsilon_3 \sum_{u_j=0}^{n_j+s_j} M_k (u_3 | u_j) = \epsilon_3 M_k (u_3) = \lambda_j M_{k-1} (u_3 | n_j + s_j) . \quad (2.10)$$

Equation (2.10) shows that factorial moments of arbitrary order k may be obtained if only the conditional factorial moment of order $(k - 1)$ is known.

Therefore, all factorial and conditional factorial moments of arbitrary order can be determined by the following procedure:

- By means of Eqs. (2.9a, b) with $k = 1$ and the relation

$$\sum_{u_j=0}^{n_j+s_j} M_1 (u_3 | u_j) = M_1 (u_3) = R_j , \quad (2.11)$$

  it is possible to determine all conditional factorial moments $M_1 (u_3 | u_j)$ (iteration of the values of the moments such that Eq. (2.11) is fulfilled). Then, Eq. (2.10) allows (with $k = 2$) us to determine the factorial moment $M_2 (u_3)$ and hence Eq. (2.8), the variance coefficient D_j
- For $k = 2$ a second iteration allows to determine all conditional factorial moments $M_2 (u_3 | u_j)$ and hence the calculation of the third factorial moment $M_3 (u_3)$ by means of Eq. (2.10) with $k = 3$
- Adequate procedures allow to determine factorial and normal moments of arbitrary order.

It should be mentioned at this point that, if there exists the same termination rate for both primary and secondary arrangement, an explicit solution is also available for all conditional factorial moments $M_1 (u_3 | u_j)$ and therefore an explicit solution for the variance coefficient D_j. This explicit solution can be found when determining all moments $M_1 (u_3 | u_j)$ as a function of $M_1 (u_3 | 0)$ by means of Eqs. (2.9a, b) (linear homogeneous difference equations of second order). Then, taking into account Eq. (2.11), the explicit solution for the variance coefficient D_j is given by

$$D_j = R_j^2 \left[\frac{c_1 A_j}{c_2 R_j} - 1 \right] , \quad (2.12)$$

where

$$c_1 = \frac{1}{2k_2}\left\{\left[1+\frac{A_j}{n_j}c_3-(k_1-k_2)\right](k_1+k_2)^{s_j+1} - \left[1+\frac{A_j}{n_j}c_3-(k_1+k_2)\right](k_1-k_2)^{s_j+1}\right\},$$

$$c_2 = n_j - A_j + A_j c_3 + \frac{1}{2\cdot k_2}\left\{\left[1+\frac{A_j}{n_j}c_3-(k_1-k_2)\right]\sum_{z_j=0}^{s_j}(k_1+k_2)^{z_j+1} - \left[1+\frac{A_j}{n_j}c_3-(k_1+k_2)\right]\sum_{z_j=0}^{s_j}(k_1-k_2)^{z_j+1}\right\},$$

$$c_3 = \frac{\dfrac{A_j^{n_j-1}}{(n_j-1)!}}{\sum_{i=0}^{n_j-1}\dfrac{A_j^i}{i!}} = E_{1,n_j-1}(A_j)\ ,$$

$$k_1 = \frac{n_j+1+A_j}{2n_j}\ ,$$

$$k_2 = \sqrt{\left(\frac{n_j+1+A_j}{2n_j}\right)^2 - \frac{A_j}{n_j}}\ .$$

3.2.4 Probability of Waiting, Mean Waiting Time, and Waiting Time Distribution

Remember that waiting is allowed only in the primary arrangements (cf. Section 3.2.1) and not for calls overflowing to the secondary route. Therefore, the probability of waiting, the mean waiting time, and the waiting time distribution can be obtained immediately when using Störmers [16] well-known formulaes for delay-loss systems.

3.2.5 Comparison with Simulation Results

As shown before, all moments of the overflow traffic distribution (section 3.2.3) and all waiting characteristics (Section 3.2.4) can be determined exactly.

Using the method of "substitute primary arrangements" (see Section 3.2.2) also the moments of its lost traffic ($\overline{R}_3, \overline{D}_3$) and, therefore, the probability of loss $B_3 = \overline{R}_3/\overline{R}$ can be determined in close approximation. This approximate method has been checked by a large number of simulation runs using a digital computer [21]. Figure 8 presents a typical example. Comparison shows the very good accordance between simulated and calculated values.

3.3 Analysis of Multiqueue Models for Overflow Strategy 3

3.3.1 Model

Given two delay-loss systems as shown in Fig. 9. New calls of class one and two are offered at first to the corresponding arrangement number one (primary) or two (secondary). If the primary arrangement is blocked, i.e., the server and all waiting places s_1 are occupied, new arriving demands of class one are overflowing to the secondary

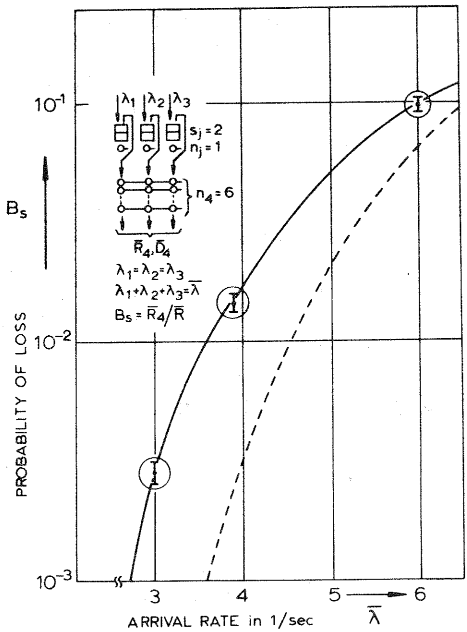

Fig. 8. Probability of loss for the secondary arrangement. Comparison between calculation and simulation (simulation results with 95% confidence intervals within the circles; assuming Poisson behavior for the overflow traffic the dashed line is obtained).

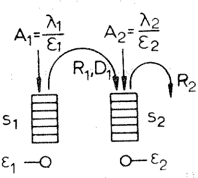

Fig. 9. The model.

arrangement (one server and s_2 waiting places). This secondary arrangement accepts all arriving demands of class one or two if there is at least one free waiting place. Calls of both class one and two are served without priorities in the order of their arrival (FIFO). If there is blocking, arriving calls are rejected.

3.3.2 Principle of Solution

Calculating all characteristic traffic values, the basic idea is the same as shown in Section 3.2. However, the investigation of the secondary arrangement with offered random plus overflow traffic is more complicated:

(a) Traffic, overflowing from the primary to the secondary arrangement, is described by the first and second overflow traffic moment (mean value R_1 and variance V_1 or variance coefficient $D_1 = V_1 - R_1$, respectively). Exact formulas for these moments are also included in the solutions presented in Section 3.2.3

(b) The total traffic offered to the secondary group is described by the sum of both overflow traffic (R_1, D_1) and direct traffic $(A_2, 0)$

$$\overline{R} = R_1 + A_2 ,$$
$$\overline{D} = D_1 + 0$$

(c) In order to investigate the traffic characteristics of the secondary arrangement, a "substitute" primary arrangement (one server and s^* waiting places) and a "generating" random traffic A^* are determined such that an overflow traffic is generated with the exact mean value $\overline{R}$ and the exact variance coefficient $\overline{D}$ (cf. Fig. 10. Contrary to Section 3.2.2, the substitute primary arrangement has also waiting places; this generalization allows us to investigate special structures and service strategies to be published later on).

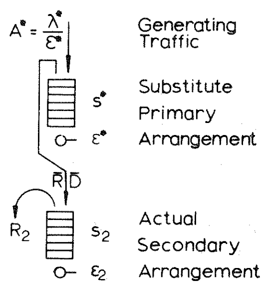

Fig. 10. Replacement of the real traffic by an equivalent traffic.

Describing the traffic flow in the substitute primary and the actual secondary arrangement by a two-dimensional Markovian process, characteristic traffic values for the secondary arrangement are determined in the following sections.

3.3.3 Calculation of the State Probabilities $p(u_1, u_2)$

Defining with $\{u_1, u_2\}$ the state that there are at the same time u_1 calls in the primary arrangement and u_2 in the secondary system, the following equations of state are found for the state probabilities:

$u_1 < 1 + s_1, u_2 = 1 + s_2$:

$$\left.\begin{aligned} \epsilon_1 p(1, 1+s_2) - (\epsilon_2 + \lambda_1) p(0, 1+s_2) &= 0, \\ \epsilon_1 p(u_1+2, 1+s_2) - (\epsilon_1+\epsilon_2+\lambda_1) p(u_1+1, 1+s_2) + \lambda_1 p(u_1, 1+s_2) &= 0. \end{aligned}\right\} \quad (3.1a)$$

This equation is a system of $(n_1 + 1)$ linear homogeneous difference equations of second order with constant coefficients. Using the Z-Transformation [20] it is possible to express all probabilities $p(u_1, 1 + s_2)$ as a function of $p(0, 1 + s_2)$

$$p(u_1, 1+s_2) = p(0, 1+s_2) \{(f + A_1)\beta_{u_1} - A_1 \beta_{u_1 - 1}\} , \quad (3.2a)$$

with

$$\beta_{u_1} = \frac{\alpha_1^{u_1} - \alpha_2^{u_1}}{\alpha_1 - \alpha_2} ; f = \frac{\epsilon_2}{\epsilon_1} ; A_1 = \frac{\lambda_1}{\epsilon_1}$$

$$\alpha_1, \alpha_2 = \frac{(1+f+A_1) \pm \sqrt{(1+f+A_1)^2 - 4A_1}}{2} .$$

By analogy, we also find

$u_1 \leq s_1, 0 < u_2 \leq s_2$:

$$p(u_1, u_2) = p(0, u_2) \{(f+A_1)\beta_{u_1} - A_1 \beta_{u_1-1}\} - \sum_{\xi=0}^{u_1-1} p(\xi, u_2+1)\beta_{u_1-\xi} , \qquad (3.2b)$$

$u_1 \leq s_1, u_2 = 0$:

$$p(u_1, 0) = p(0,0) \{A_1 \beta^*_{u_1} - A_1 \beta^*_{u_1-1}\} - \sum_{\xi=0}^{u_1-1} p(\xi, 1)\beta^*_{u_1-\xi} , \qquad (3.2c)$$

with

$$\beta^*_{u_1} = \frac{(\alpha_1^*)^{u_1} - (\alpha_2^*)^{u_1}}{\alpha_1^* - \alpha_2^*} ; \alpha_1^*, \alpha_2^* = \frac{(f+A_1) \pm \sqrt{(f+A_1)^2 - 4A_1}}{2} ,$$

$u_1 = 1 + s_1, 0 \leq u_2 \leq s_2$:

$$fp(1+s_1, u_2+2) - (1+f+A_1)p(1+s_1, u_2+1) + A_1 p(1+s_1, u_2) = -A_1 p(s_1, u_2+1) . \qquad (3.2d)$$

Equations (3.2a, b, c) allow us to describe all probabiliites $p(u_1, u_2)$ as a function of the "basic state" $p(0, u_2)$ and "higher" values $p(\xi, u_2+1)$ $(\xi = 0, 1, 2, \cdots, u_1 - 1)$. Therefore, when starting with the highest possible value $u_2 = s_2 + 1$ and systematic replacement of unknown values it is possible to express all state probabilities $p(u_1, u_2)$ by the marginal values $p(0, u_2), \cdots, p(0, s_2+1)$, i.e., the two-dimensional state relations can be reduced to a one-dimensional system [10]. Finally, Eq. (3.2d) and the normalizing condition allow us to find an explicit solution for all state probabilities.

$$p(u_1, u_2) = \frac{A_1^{u_1}}{1 + A_1 \dfrac{1 - A_1^{s_1+1}}{1 - A_1}} \frac{\sum_{\rho=u_2}^{1+s_2} c_{u_1, u_2, \rho}\, b_\rho}{\sum_{\nu=0}^{1+s_2} \sum_{\rho=\nu}^{1+s_2} c_{u_1, \nu, \rho}\, b_\rho} ,$$

$$c_{u_1, u_2, \rho} = -\sum_{\xi=0}^{u_1-1} c_{\xi, u_2+1, \rho}\, \beta_{u_1-\xi} \qquad (u_2 \neq \rho; u_2 \neq 0; \rho \geq u_2) ,$$

$$c_{u_1, 0, \rho} = -\sum_{\xi=0}^{u_1-1} c_{\xi, 1, \rho}\, \beta^*_{u_1-\xi} \qquad (u_2 \neq \rho; u_2 = 0) ,$$

$$c_{u_1,u_2,\rho} = (f+A_1)\,\beta_{u_1} - A_1\,\beta_{u_1-1} \qquad (u_2=\rho;u_2\neq 0)\ ,$$

$$c_{u_1,u_2,\rho} = A_1\,\beta^*_{u_1} - A_1\,\beta^*_{u_1-1} \qquad (u_2=\rho;u_2=0)\ ,$$

$$c_{u_1,u_2,\rho} = 0 \qquad (u_2>\rho)\ ,$$

$$b_{u_2} = \sum_{z_1=u_2+1}^{1+s_2} a_{u_2,z_1} \sum_{z_2=z_1+1}^{1+s_2} a_{z_1,z_2} \sum_{z_3=z_2+1}^{1+s_2} a_{z_2,z_3} \cdots \sum_{z_w=z_{w-1}+1}^{1+s_2} a_{z_{w-1},z_w};$$

$$(w = 1+s_2-u_2)\ ,$$

$$a_{u_2,\rho} = \frac{\left(-A_1 c_{1+s_1,u_2,\rho} + (1+f+v\cdot A_1)\,c_{1+s_1,u_2+1,\rho} - v c_{1+s_1,u_2+1,\rho} - A_1 c_{s_1,u_2+1,\rho}\right)}{A_1 c_{1+s_1,u_2,u_2}}\ ,$$

$$v = \begin{cases} 1\ , & 0 \le u_2 < s_2\ ,\\ 0\ , & \text{otherwise}\ .\end{cases}$$

3.3.4 Characteristic Traffic Values

As already mentioned, all characteristic traffic values for the primary arrangement are independent of the secondary arrangement and given by well-known formulas [16]. In addition, all overflow traffic moments are determined in Section 3.2.3.

All state probabilities $p\,(u_1, u_2)$ for a substitute primary and the actual secondary arrangement are determined explicitly in Section 3.3.3. Therefore, it is easy to find the following characteristic values for the secondary arrangement:

Carried traffic

$$Y_2 = \sum_{u_1=0}^{1+s_1} \sum_{u_2=1}^{1+s_2} p\,(u_1,u_2) = \frac{\lambda_1}{\epsilon_2}\left[p\,(u_1=s_1+1) - p\,(s_1+1, s_2+1)\right]\ . \qquad (3.4)$$

Lost traffic

$$R_2 = \frac{\lambda_1}{\epsilon_2}\ p\,(1+s_1, 1+s_2)\ . \qquad (3.5)$$

Probability of waiting (probability that an arriving call has to wait under the condition that it is offered to the secondary arrangement)

$$W_2 = \frac{\sum_{u_2=1}^{s_2} p\,(1+s_1,u_2)}{\sum_{u_2=0}^{s_2+1} p\,(1+s_1,u_2)}\ . \qquad (3.6)$$

Mean number of calls waiting for service

$$\Omega_2 = \sum_{u_1=0}^{1+s_1} \sum_{u_2=1}^{1+s_2} (u_2 - 1)\, p(u_1, u_2) \,. \tag{3.7}$$

Mean waiting time for waiting calls

$$t_{w2} = \frac{\Omega_2}{\lambda_1 \sum_{u_2=1}^{s_2} p(1+s_1, u_2)} \,. \tag{3.8}$$

In addition, it is possible to determine explicitly the waiting time distribution for the total system as well as for the secondary arrangement only. These results will be published at some future time.

3.3.5 Comparison with Simulation Results

As shown before, all characteristic traffic values referring to the primary arrangement (such as waiting times and overflow traffic) can be determined exactly.

All characteristic traffic values referring to the secondary arrangement are approximate values because the actually offered overflow traffics are replaced by a ficticious traffic generating exactly only the first two moments of the actual traffics.

Therefore, a large number of simulation runs using a digital computer have been performed [22]. Comparison of both simulated and calculated values shows the accuracy of this approach (cf Figure 11).

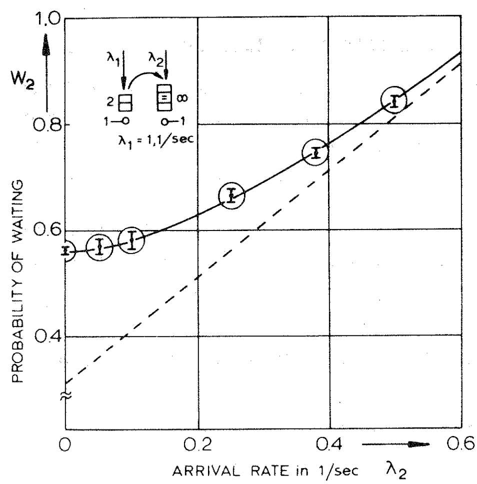

Fig. 11. Probability of waiting for (directly offered and overflowing) calls waiting in the secondary arrangement for service. Comparison between calculation and simulation (simulation results with 95% confidence intervals within the circles; assuming Poisson behavior for the overflow traffic the dashed line is obtained).

IV. COMPARISON OF DIFFERENT MULTIQUEUE MODELS WITH RESPECT TO UNBALANCED LOAD

In this section, three numerical examples are given for demonstration of the effects of unbalanced load, different server configurations, as well as different overflow strategies on the characteristic traffic values (grade of service). By means of such numerical evaluations, suitable server configurations and operation strategies can be found, which meet the requirements of overflow, load-sharing, and reliability with respect to breakdowns in data transmission networks and computer systems.

4.1 Comparison of Single-server Systems with Overflow Strategies S1 and S3

In the upper left of Fig. 12(a), three different models are shown each having two servers and two queues. Model (a) represents two separated single-server, single-queue systems without any mutual aid; Model (b) works under overflow strategy S1, and Model (c) under overflow strategy S3. The arrival rate to the second input remains constant $\lambda_2 = 0.6 \text{ sec}^{-1}$, while λ_1 varies. The service rates of the servers are identical $\epsilon_1 = \epsilon_2 = 1 \text{ sec}^{-1}$. The second server of Model (b) serves the second queue according to a nonpreemptive priority mode.

Figure 12(a) shows the probabilities of loss B_1 and B_2, Fig. 12(b) shows the mean waiting times t_{w1} and t_{w2} referred to waiting calls, dependent on the arrival rate λ_1. The comparison of Model (a), Model (b), and Model (c) results in lower loss probabilities B_1 at the expense of B_2. Model (b) reduces the mean waiting times t_{w1} drastically, while t_{w2} is not influenced. Model (c) guarantees a maximum throughput of 1-calls at the expense of increasing mean waiting times t_{w2}. In Figure 12(c) the pdf of waiting times are shown for the case of $\lambda_1 = 1.0 \text{ sec}^{-1}$, $\lambda_2 = 0.6 \text{ sec}^{-1}$.

4.2 Comparison of Multiserver Systems with Overflow Strategies S1 and S2

In the lower right of Fig. 13(a), two overflow models are shown each having two primary routes and a common secondary route. Model (a) works under overflow strategy S1, Model (b) operates according to overflow strategy S2. Again, λ_2 remains constant, $\lambda_2 = 2 \text{ sec}^{-1}$, while λ_1 varies. The service rates of the different servers are identical $\epsilon = 1 \text{ sec}^{-1}$.

Figure 13(a) shows the probabilities of loss B_1 and B_2. Model (b) results in greater loss probabilities compared with Model (a). However, Model (b) yields a greater utilization for the primary routes by decreasing the load of the common secondary route simultaneously, cf Figure 13(b).

4.3 Comparison of Different Overflow Server Configurations with Overflow Strategy S1

In Fig. 14 four different models are shown with two routes each having different service rates (transmission speeds) $\epsilon_1 = 2 \text{ sec}^{-1}$, $\epsilon_2 = 1 \text{ sec}^{-1}$, as previously discussed in Section 2.1. As before, the arrival rate λ_2 remains constant $\lambda_2 = 2 \text{ sec}^{-1}$, while λ_1 assumes different values (cf. Table I). Common servers serve their own queue according to a nonpreemptive priority rule.

The characteristic traffic values B_j, W_j, t_{wj}, and Y_j $(j = 1, 2)$ are given in Table I for comparison of the efficiency of the different models (Figures 14.1 to 14.4).

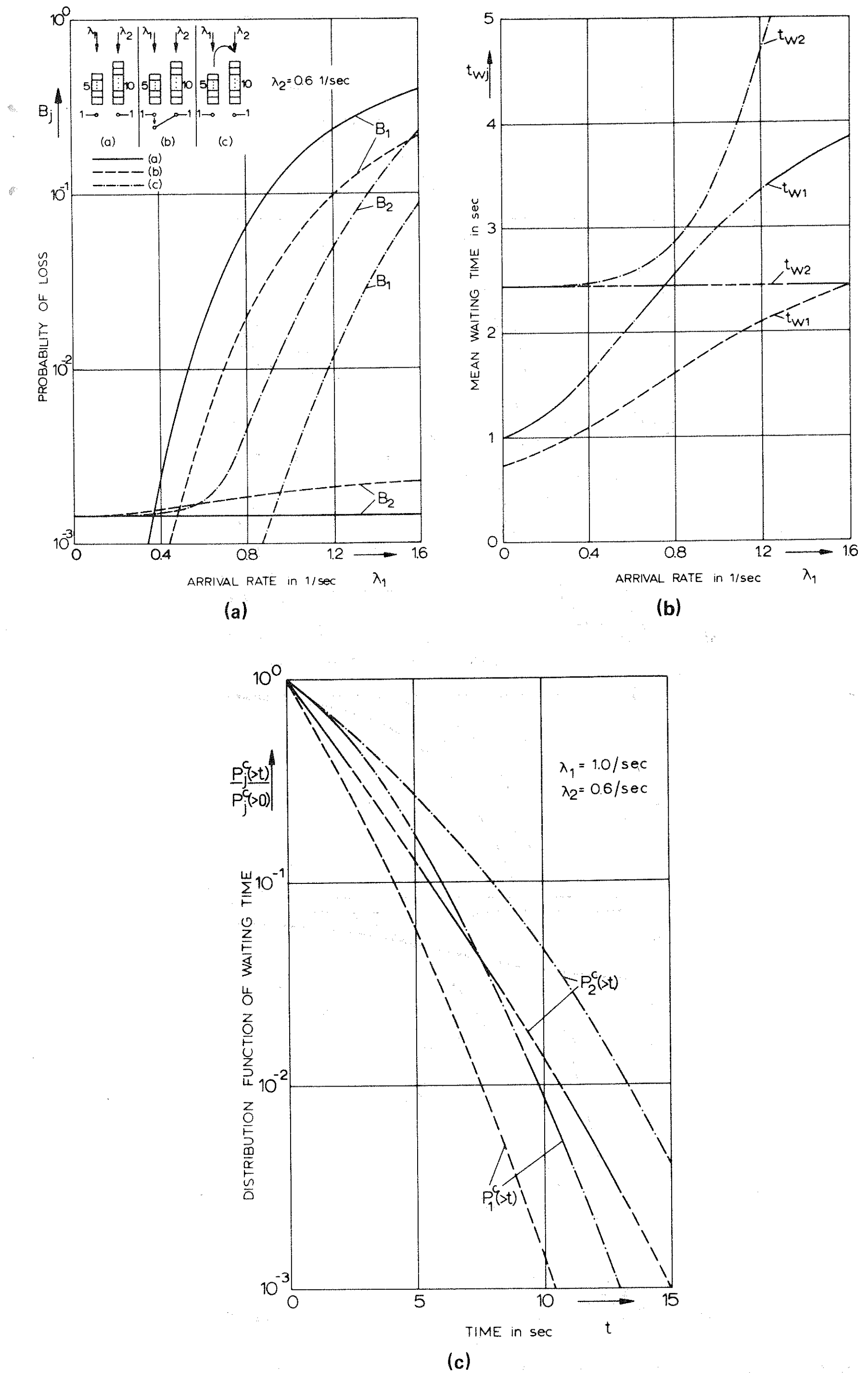

Fig. 12. Comparison of single-server systems, overflow strategies S1 and S3. (a) Probabilities of loss vs arrival rate λ_1. (b) Mean waiting times vs arrival rate λ_1. (c) Distribution functions of waiting time.

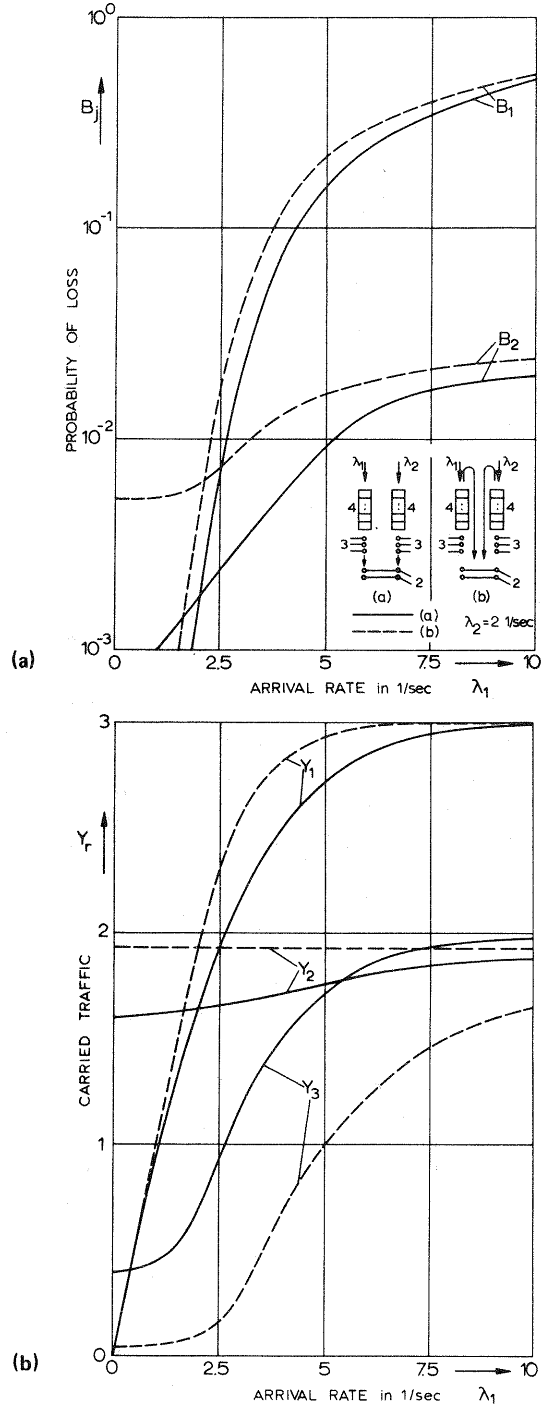

Fig. 13. Comparison of multiserver systems, overflow strategies S1 and S2. (a) Probabilities of loss vs arrival rate λ_1. (b) Carried traffics vs arrival rate λ_1.

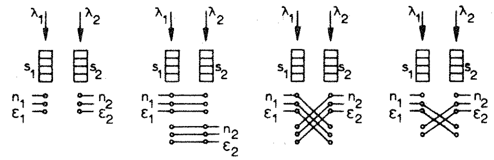

Fig. 14. Comparison of different overflow server configurations, overflow strategy S1, cf. Table I.

Table I. Comparison of different overflow server configurations, overflow strategy S1 (server configurations, cf. Fig. 14). Interqueue discipline: nonpreemptive priority. Parameters: $n_1 = n_2 = 3$, $s_1 = s_2 = 4$, $\epsilon_1 = 2\ sec^{-1}$, $\epsilon_2 = 1\ sec^{-1}$.

CHARACTERISTIC VALUES	ARRIVAL RATES $[sec^{-1}]$ λ_1	λ_2	Fig. 14.1	Fig. 14.2	Fig. 14.3	Fig. 14.4
B_1	2	2	0.000748	0.000196	0.000281	0.000337
	6	2	0.145161	0.074883	0.076463	0.086445
	10	2	0.413106	0.315564	0.315921	0.331694
B_2	2	2	0.031083	0.000373	0.000535	0.000877
	6	2	0.031083	0.020303	0.020731	0.020871
	10	2	0.031083	0.053768	0.053828	0.047899
W_1	2	2	0.089820	0.049255	0.070659	0.073504
	6	2	0.580645	0.487385	0.497666	0.519068
	10	2	0.539351	0.587893	0.588558	0.584370
W_2	2	2	0.378825	0.049078	0.070405	0.090133
	6	2	0.378825	0.541965	0.553398	0.503555
	10	2	0.378825	0.849689	0.850651	0.720658
t_{w1} [sec]	2	2	0.241667	0.175295	0.175295	0.188719
	6	2	0.416667	0.307951	0.307951	0.325327
	10	2	0.515931	0.401491	0.401491	0.418531
t_{w2} [sec]	2	2	0.671795	0.213105	0.213105	0.246009
	6	2	0.671795	0.430525	0.430525	0.460312
	10	2	0.671795	0.577314	0.577314	0.596066
Y_1	2	2	0.999251	1.610001	1.156481	1.154515
	6	2	2.564516	2.554727	2.483971	2.498991
	10	2	2.934469	2.914542	2.907159	2.911325
Y_2	2	2	1.937834	0.778860	1.685407	1.688537
	6	2	1.937834	2.400645	2.531818	2.441604
	10	2	1.937834	2.907733	2.918808	2.764607

V. CONCLUSION

Multiqueue models were investigated working under three types of overflow strategies. The overflow strategies are applicable to the control of data networks for reasons of message routing, and various configurations of computers for reasons of reliability and loadsharing, as well. The analysis was carried out by exact and approximate calculations using methods of state equations and equivalent random traffic. The influence of the overflow strategies on the main traffic criteria was demonstrated by various examples showing the interdependency between computer and line utilization, waiting time, and loss probabilities. The results can also be used for the optimal design of data networks and computer systems taking into account the costs for waiting times and service equipment.

REFERENCES

[1] A. Lotze, "Problems of Traffic Theory in the Design of International Direct Distance Dialing Networks," *NTZ Comm. J.*, 7, No. 2/3, 41–46 (November 1968).

[2] B. W. Boehm, and R. L. Mobley, "Adaptive Routing Techniques for Distributed Communications Systems," *IEEE Trans. Comm. Tech.*, **COM-17**, 340–349 (June 1969).

[3] A. W. Butrimenko, "Adaptive Routing Technique and Simulation of Communication Networks," *Proc. 6th Int. Teletraffic Congress*, Munich (1970).

[4] J. Martin, *Design of Real-Time Computer Systems* (New Jersey: Prentice Hall, 1967).

[5] J. Riordan, *Stochastic Service Systems* (New York: John Wiley, 1962).

[6] R. Syski, *Introduction to Congestion Theory in Telephone Systems* (London: Oliver and Boyd, 1960).

[7] P. Kühn, Über die Berechnung der Wartezeiten in Vermittlungs- und Rechnersystemen" PhD. Thesis, University of Stuttgart (1972).

[8] ——, "Combined Delay and Loss Systems with Several Input Queues, Full and Limited Accessibility," *Arch. Elektr. Übertr. (AEÜ)*, **25**, 449–454 (September–October 1971).

[9] R. Syski, "Markovian Queues," *Proc. Symp. on Congestion Theory*, 170–227 (North Carolina: The University of North Carolina Press, 1964).

[10] R. Schehrer, "Über die exakte Berechnung von Überlaufsystemen der Wählvermittlungstechnik," Ph.D. Thesis, University of Stuttgart (1969).

[11] R. I. Wilkinson, "Theories for Toll Traffic Engineering in the USA," *Bell Sys. Tech. J.*, **35**, 421–514 (1956).

[12] G. Bretschneider, "Die Berechnung von Leitungsgruppen für überfließenden Verkehr in Fernsprechwählanlagen," *Nachrichtentechn Z. (NTZ)* **9**, 533-540 (November 1956).

[13] A. Lotze, "A Traffic Variance Method for Gradings of Arbitrary Type," *Post Office Elec. Eng. J. (POEEJ)*, Special Issue (1966) (papers of the 4th International Teletraffic Congress, London, 1964).

[14] U. Herzog, and A. Lotze, "The RDA-Method, a Method Regarding the Variance Coefficient for Limited Access Trunk Groups," *NTZ Comm. J.*, **7**, No. 2/3, 47–52 (November 1968).

[15] A. Lotze, and R. Schehrer, "The Design of Alternate Routing Systems with Regard to the Variance Coefficient," *NTZ Comm. J.*, **7**, No. 2/3, 52–56 (November 1968).

[16] H. Störmer, "Wartezeitlenkung in Handbedienten Vermittlungsanlagen," *Arch. Elektr. Übertr. (AEÜ)*, **10**, 58–64 (February 1956).

[17] G. P. Basharin, "On Analytical and Numerical Methods of Switching System Investigation," *Proc. 6th Int. Teletraffic Congress*, Munich (1970).

[18] U. Herzog, "Die Exakte Berechnung des Streuwertes von Überlaufverkehr hinter Koppelanordnungen beliebiger Stufenzahl mit Vollkommener bzw. Unvollkommener Erreichbarkeit," *Arch. Elektr. Übertr. (AEÜ)*, **20**, 180–184 (March 1966).

[19] ——, "Die Bemessung ein- und Mehrstufiger Koppelanordnungen der Vermittlungstechnik für Angebotenen Überlaufverkehr, Ph.D. Thesis, University of Stuttgart (1968).

[20] G. Doetsch, *Anleitung zum praktischen Gebrauch der Laplace-Transformation und der Z-Transformation* (München-Wien: R. Oldenbourg Verlag, 1967).

[21] B. Werres, "Wartesysteme mit Überlauf," Diploma Thesis, University of Stuttgart (1971).

[22] R. Ott, "Wartesysteme mit Angebotenem Überlaufverkehr," Diploma Thesis, University of Stuttgart (1972).