Special reprint
from
AEÜ ARCHIV FÜR ELEKTRONIK UND ÜBERTRAGUNGSTECHNIK
ELECTRONICS AND COMMUNICATION
Band 25 (1971), pp. 435–441

The Multi-server Queuing System with Preemptive Priority

by G. JOACHIM BRANDT*

The solution of a waiting system is presented, in which the calls are served according to the preemptive priority rule:
A call of higher priority has absolute precedence over a call of lower priority, not only in the queue, but also in the servers.
The general case of n servers and s waiting places is dealt with.
In the author's thesis the loss-delay system has been solved for an infinite and a finite number of Poisson sources. In this paper the infinite case is dealt with.

Ein Wartesystem mit mehreren Bedienungseinheiten und unterbrechenden Prioritäten

Ein Wartesystem wird behandelt, in dem die ankommenden Rufe in Klassen verschiedener Priorität eingeteilt sind:
Ein Ruf höherer Priorität hat nicht nur im Wartespeicher absoluten Vorrang über einen Ruf niederer Priorität, sondern auch in den Bedienungseinheiten.
Das allgemeine Warteverlustsystem mit n Bedienungseinheiten und s Warteplätzen wird gelöst.
In der Dissertation des Verfassers wird das preemptive Warteverlustsystem sowohl für eine endliche als auch für unendliche Zahl von Poisson-Quellen gelöst. Hier wird der Fall unendlicher Quellenzahl behandelt.

* Dr. G. J. BRANDT, c/o Institute for Switching and Data Technics, University of Stuttgart, D-7 Stuttgart, Breitscheidstrasse 2. (Now with SITA, Systems Planning Department, 112, Avenue Charles de Gaulle, F-92 Neuilly Paris, France.)

1. Introduction

The loss-delay system has n servers and s waiting places. The calls to be served belong to one of r priority classes so that a call of class i has absolute priority over all calls of classes $i+1, i+2, \dots$ up to class r. Combining the calls of classes 1, 2, ... up to class i, we speak of class $\leqq i$, accordingly of class $< i$ or $> i$. A call of class $\leqq r$ is any call without regard to its individual priority class. A call of class i is abbreviated as "i-call".

The traffic characteristics of a certain priority class are labeled with the class index preceding the abbreviation of the characteristic.

We investigate the preemptive priority assuming the first come — first served rule within each priority class. This service discipline prescribes an order of precedence, which can be symbolized in the following manner:

If a total of j calls is in the system (being in service or waiting), then we will say: The j calls occupy the places 1, 2, ... up to j. The call of highest importance occupies place 1, the next important one place 2 and so on up to place j. A place $\nu \leqq n$ represents a server and a place $\nu > n$ a waiting place.

This order of precedence has to be remembered for the following two situations:

1. There are j calls in the system and one of the calls being served terminates its service. Therefore, a place ν (with $\nu \leqq n$) becomes free. The calls, having until now occupied the places $\nu + 1$, $\nu + 2, \ldots, j$, advance one place so that they occupy the places ν, $\nu + 1, \ldots, j - 1$. The call of place $n + 1$ proceeds to place n, that means this call now gets a free server.
2. There are j calls in the system and a new call arrives, say an i-call. It is placed in front of all $> i$-calls, but behind all $\leqq i$-calls. If all $n + s$ places are occupied by $\leqq i$-calls, then the arriving call is lost. If the call occupies a place $\nu \leqq j$, then the calls in place ν, $\nu + 1, \ldots$ up to j are pushed back one place so that they now occupy the places $\nu + 1$, $\nu + 2, \ldots, j + 1$. A call which is pushed back from place n to place $n + 1$, interrupts its service and waits until its service can be continued. A call which is pushed back from place $n + s$ is displaced from the system and thus lost.

We have an infinite number of Poisson sources that means the interarrival times are distributed negative exponentially. The mean number of arriving calls per unit of time is denoted as the total arrival rate $_{\leqq r}\lambda$. The share of class i is denoted by $_ip$ so that

$$_{\leqq r}p = \sum_{i=1}^{r} {}_ip = 1 . \tag{1}$$

The arrival rate $_i\lambda$ of class i results in

$$_i\lambda = {}_ip \, {}_{\leqq r}\lambda . \tag{2}$$

Concerning the service process, we assume that the service times are distributed negative exponentially with the mean service time h. The Markov property of the negative exponential distribution has an important consequence for the preemptive system. Since the remaining service time of an interrupted call and the whole service time of the interrupting call are identically distributed, the interruption has no influence upon the termination process: With respect to the service time, two equivalent calls are exchanged.

If there are j calls in the system, the termination rate μ_j results in

$$\mu_j = \begin{cases} j/h & \text{for} \quad j \leqq n \\ n/h & \text{for} \quad j > n. \end{cases} \tag{3}$$

We conclude these introductory remarks by defining the parameter $_iA$, the traffic offered by class i

$$_iA = h \, {}_i\lambda . \tag{4}$$

$_iA$ is the expected number of arriving i-calls within the mean service time h.

2. The Probabilities of State and the Related Traffic Characteristics

The random variable $_{\leqq i}X(t)$ indicates the number j of $\leqq i$-calls which are in the system at time t.

We calculate the probabilities of state for class $\leqq i$:

$$_{\leqq i}P_j(t) = P\{_{\leqq i}X(t) = j\} . \tag{5}$$

Considering the system at time t and at time $t + \Delta t$, the state j at time $t + \Delta t$ can come about in the following mutually exclusive ways:

i) from state $j - 1$ with probability

$$_{\leqq i}\lambda \, \Delta t + o(\Delta t) ,$$

ii) from state $j + 1$ with probability

$$\mu_{j+1} \Delta t + o(\Delta t) ,$$

iii) from state j for $j < n + s$ with probability

$$1 - (_{\leqq i}\lambda + \mu_j) \Delta t + o(\Delta t) ,$$

and for $j = n + s$ with probability

$$1 - \mu_{n+s} \Delta t + o(\Delta t) ,$$

iv) from any other state with probability

$$o(\Delta t) .$$

$o(\Delta t)$ is a function of higher order in Δt.

Combining these events and proceeding to the limit $\Delta t \to 0$ we get a system of differential equations, where the derivatives with respect to the time are denoted by primes:

$$\begin{aligned} {}_{\leqq i}P'_0(t) &= - {}_{\leqq i}\lambda \, {}_{\leqq i}P_0(t) + \mu_1 \, {}_{\leqq i}P_1(t) , \\ {}_{\leqq i}P'_j(t) &= {}_{\leqq i}\lambda \, {}_{\leqq i}P_{j-1}(t) - ({}_{\leqq i}\lambda + \mu_j) \, {}_{\leqq i}P_j(t) + \\ &\quad + \mu_{j+1} \, {}_{\leqq i}P_{j+1}(t) , \\ {}_{\leqq i}P'_{n+s}(t) &= {}_{\leqq i}\lambda \, {}_{\leqq i}P_{n+s-1}(t) - \mu_{n+s} \, {}_{\leqq i}P_{n+s}(t) . \end{aligned} \tag{6}$$

We sum up these equations from $j = m$ up to $j = n + s$ and obtain

$$\sum_{j=m}^{n+s} {}_{\leqq i}P'_j(t) = {}_{\leqq i}\lambda \, {}_{\leqq i}P_{m-1}(t) - \mu_m \, {}_{\leqq i}P_m(t) . \tag{7}$$

Summing up these equations again, we get

$$\sum_{m=1}^{n+s} \sum_{j=m}^{n+s} {}_{\leqq i}P'_j(t) = {}_{\leqq i}\lambda \sum_{m=0}^{n+s-1} {}_{\leqq i}P_m(t) - \sum_{m=1}^{n+s} \mu_m \, {}_{\leqq i}P_m(t) . \tag{8}$$

A comparison of eq. (6) with the priority-less system proves the following statement: Class $\leqq i$ — the union of the priority classes 1, 2, ... up to i — behaves in the preemptive system as if:

i) the priority classes $> i$ do not exist,

ii) the sub-division of class $\leqq i$ in preemptive priority classes does not exist,

that means as if no interruption and replacement within class $\leqq i$ would occur.

The solution of the system (6) of differential equations leads to the determination of the roots $_i\xi_\nu$ of the characteristic equation $_iD(\xi) = 0$, where $_iD(\xi)$ is the determinant

$$_iD(\xi) = \tag{9}$$

$$= \begin{vmatrix} {}_{\leqq i}\lambda + \xi & -\mu_1 & 0 & \cdots & 0 \\ -{}_{\leqq i}\lambda & {}_{\leqq i}\lambda + \mu_1 + \xi & -\mu_2 & \cdots & 0 \\ 0 & -{}_{\leqq i}\lambda & & \cdot & 0 \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & -\mu_{n+s} \\ 0 & 0 & & -{}_{\leqq i}\lambda & {}_{\leqq i}\lambda + \mu_{n+s} + \xi \end{vmatrix} .$$

One root of this polynomial is $_i\xi = 0$, which corresponds to the stationary solution $_{\leqq i}P_j$. It can be shown [4] that the other roots ${}_i\xi$ are negative and distinct. Therefore, the time-dependent probabilities of state can be written in the form

$$_{\leq i}P_j(t) = {}_{\leq i}P_j + \sum_{\nu=1}^{n+s} {}_{\leq i}C_{j,\nu}\, e^{{}_i\xi_\nu t}. \tag{10}$$

The constants ${}_{\leq i}C_{j,\nu}$ depend on the initial probabilities ${}_{\leq i}P_j(t=0)$.

We are interested especially in the stationary solution ${}_{\leq i}P_j$, which can be determined by substituting the stationary conditions

$$_{\leq i}P_j(t) = {}_{\leq i}P_j \tag{11}$$

into eq. (7). Now, the probabilities of state are time-independent. This leads to the "statistical equilibrium"

$$_{\leq i}\lambda\, {}_{\leq i}P_{m-1} = \mu_m\, {}_{\leq i}P_m. \tag{12}$$

The probabilities of state satisfy the condition

$$\sum_{j=0}^{n+s} {}_{\leq i}P_j = 1. \tag{13}$$

Using the traffic parameter

$$_{\leq i}\beta_j = \mu_j / {}_{\leq i}\lambda \tag{14}$$

the following formula for the time-independent probabilities of state can easily be verified:

$$_{\leq i}P_j = \frac{\prod\limits_{m=j+1}^{n+s} {}_{\leq i}\beta_m}{\sum\limits_{j=0}^{n+s} \prod\limits_{m=j+1}^{n+s} {}_{\leq i}\beta_m}. \tag{15}$$

The denominator of this expression is identical to the basic function $\Phi_{0,n+s}({}_{\leq i}A)$. This basic function is introduced in the Appendix, where also some calculating rules are stated. Next, we will agree upon the abbreviations

$$\Phi_\varkappa = \Phi_{0,\varkappa}, \quad {}_{\leq i}\Phi_{\nu,\varkappa} = {}_{\leq i}\Phi_{\nu,\varkappa}({}_{\leq i}A). \tag{16), (17}$$

For $\nu > n$ the function $\Phi_{\nu,\varkappa}$ depends only on the difference $\varkappa - \nu$, so we define for $\nu > n$

$$\Psi_\varrho = \Phi_{\nu,\nu+\varrho}. \tag{18}$$

Using the basic fuction $\Phi_{\nu,\varkappa}$ the probability of state can be written in the form

$$_{\leq i}P_{n+s} = \frac{1}{{}_{\leq i}\Phi_{n+s}}, \quad {}_{\leq i}P_j = {}_{\leq i}P_{n+s} \prod_{m=j+1}^{n+s} {}_{\leq i}\beta_j. \tag{19}$$

The probability ${}_iF_j$ that an i-call occupies place j at its arrival is given by

$$_iF_j = {}_{\leq i}P_j. \tag{20}$$

In the following, those traffic characteristics calculated upon the probabilities of state shall be determined.

The probability ${}_iC$ that an i-call is lost at its arrival results in

$$_iC = {}_{\leq i}P_{n+s} = \frac{1}{{}_{\leq i}\Phi_{n+s}} \tag{21}$$

and satisfies — in accordance with the priority rule — the relation

$$_1C < {}_2C < \cdots < {}_rC \quad \text{with} \quad {}_rC = B. \tag{22}$$

B is the probability of loss in the loss-delay system without priorities.

The probability ${}_iR$ that an i-call occupies a waiting place at its arrival results in

$$_iR = \frac{n}{{}_{\leq i}A}\, \frac{{}_{\leq i}\Psi_{s-1}}{{}_{\leq i}\Phi_{n+s}}. \tag{23}$$

The probability ${}_{\leq i}X_j$ that place j is occupied by an $\leq i$-call is

$$_{\leq i}X_j = \sum_{\nu=j}^{n+s} {}_{\leq i}P_\nu = \frac{{}_{\leq i}\Phi_{j,n+s}}{{}_{\leq i}\Phi_{n+s}}. \tag{24}$$

The mean total number ${}_{\leq i}X$ of places occupied by $\leq i$-calls reads

$$_{\leq i}X = \sum_{j=1}^{n+s} j\, {}_{\leq i}P_j. \tag{25}$$

${}_{\leq i}X$ is the traffic of class $\leq i$ carried by the system. When we transform the sum of eq. (25) we get

$$_{\leq i}X = \sum_{j=1}^{n+s} {}_{\leq i}X_j. \tag{26}$$

Therefore, ${}_{\leq i}X_j$ can be interpreted as the traffic of class i carried by place j.

Analogically, we gain the traffic ${}_{\leq i}Y$ of class $\leq i$ carried by the servers:

$$_{\leq i}Y = \sum_{j=1}^{n} {}_{\leq i}X_j. \tag{27}$$

Using the definition that ${}_{\leq i}Y$ is the mean number of servers occupied by class $\leq i$, we find another formula for ${}_{\leq i}Y$:

$$_{\leq i}Y = h \sum_{j=1}^{n+s} \mu_j\, {}_{\leq i}P_j. \tag{28}$$

Introducing the stationary conditions (11) into eq. (8), we obtain

$$_{\leq i}Y = {}_{\leq i}A\,(1 - {}_{\leq i}P_{n+s}). \tag{29}$$

We will return to this formula in Section 4. The mean queue length ${}_{\leq i}Q$ of class $\leq i$ results in

$$\begin{aligned} {}_{\leq i}A = n:&\quad {}_{\leq i}Q = \frac{{}_{\leq i}A}{n - {}_{\leq i}A}\, \frac{{}_{\leq i}\Psi_s - (s+1)}{{}_{\leq i}\Phi_{n+s}}, \\ {}_{\leq i}A \neq n:&\quad {}_{\leq i}Q = \frac{1}{2}\, s(s+1)\, \frac{1}{{}_{\leq i}\Phi_{n+s}}. \end{aligned} \tag{30}$$

We have the relation

$$_{\leq i}X = {}_{\leq i}Y + {}_{\leq i}Q. \tag{31}$$

The corresponding terms for class i are given as differences of class $\leq i$ and of class $< i$, in particular the traffic of class i, carried by place j results in

$$_iX_j = {}_{\leq i}X_j - {}_{<i}X_j = \frac{{}_{\leq i}\Phi_{j,n+s}}{{}_{\leq i}\Phi_{n+s}} - \frac{{}_{<i}\Phi_{j,n+s}}{{}_{<i}\Phi_{n+s}}. \tag{32}$$

3. The Random Walk Principle

After having calculated the probabilities of state in the last section, we were able to determine some important traffic characteristics, but there are many others which cannot be found only by means of the probabilities of state. Not only the distribution of the waiting time or the distribution of the total time a call stays in the system belong to this class, but also e.g. such an elementary characteristic as the probability ${}_iU$ that an i-call is interrupted.

In order to calculate these other traffic characteristics we will follow the random walk principle, which has been introduced for these problems in [1].

Let us consider a call which starts in place j. We will observe its "life" in the system and we will describe this "life" as a random walk. The states of the random walk are the places which the call occupies one after another. Beyond these places we define two states, which are absorbing. That means the random walk of the observed call is terminated as soon as the call reaches an absorbing state. The two absorbing states shall be specified as those of {success} and {non-success}. In order to simplify the formulas we identify the state {success} with the figure 0 and the state {non-success} with the figure $n+s+1$. The call can occupy the places $1, 2, \ldots$ up to $n+s$ until it reaches one of the states 0 or $n+s+1$.

We are interested in the probability E_j that an observed call reaches the state {success} under the condition that it starts (or stays) in place j. Moreover the distribution of the time shall be calculated, which a call in place j needs until it reaches the state 0. In Section 4, we will see that we get the various traffic characteristics by suitably defining the state {success}.

The random variable $Y(t)$ shall indicate the places of a call during its random walk. The transition probability that the call changes from place k to place j ($j=0$ or $n+s+1$ included) within the time Δt is denoted by

$$P\{Y(t+\Delta t)=j\,|\,Y(t)=k\}=u_{k,j}\,\Delta t+o(\Delta t)\,. \quad (33)$$

The probability that the call remains in state k during the time Δt can be found by means of the condition

$$\sum_{j=0}^{n+s+1} P\{Y(t+\Delta t)=j\,|\,Y(t)=k\}=1\,. \quad (34)$$

The coefficients $u_{k,j}$ are denoted "jump rates". They result directly from the arrival and termination rates, as we will see in Section 4.

Summing up all jump rates $u_{k\,j}$ over j (the two absorbing states included), we obtain the jump rate $u_{k,k}$, which describes the event that the call leaves its place k within Δt:

$$u_{k,k}=\sum_{\substack{j=0\\ j\neq k}}^{n+s+1} u_{k,j}\,. \quad (35)$$

Using this abbreviation the probability that the call remains in its place k during Δt reads

$$P\{Y(t+\Delta t)=k\,|\,Y(t)=k\}=1-u_{k,k}\,\Delta x+o(\Delta x)\,. \quad (36)$$

The following results are derived explicitly in [1].

The probability E_j that a call reaches the absorbing state {success} under the condition that it starts in place j satisfies the system of equations

$$u_{j,j}\,E_j-\sum_{\substack{k=1\\ k\neq j}}^{n+s} u_{j,k}\,E_k=u_{j,0}\,. \quad (37)$$

The probability $Z_j(x)$ that a call reaches the state {success} and needs a time greater x under the condition that it starts in place j satisfies the following system of differential equations:

$$Z_j'(x)=\sum_{\substack{k=1\\ k\neq j}}^{n+s} u_{j,k}\,Z_k(x)-u_{j,j}\,Z_j(x)\,. \quad (38)$$

The initial values are the success probabilities E_j in place j:

$$Z_j(x=0)=E_j\,. \quad (39)$$

Finally, we have the following system (40) of equations for the mean times Z_j of a successful call in place j related to all calls in place j:

$$u_{j,j}\,Z_j-\sum_{\substack{k=1\\ k\neq j}}^{n+s} u_{j,k}\,Z_k=E_j\,. \quad (40)$$

This equation results by integration from the differential system (38).

For the derivation of the calculations (37), (38), and (40) we do not need to assume that the process is stationary. These conditioned characteristics are valid also for the time-dependent process.

4. Displacement Characteristics

In this chapter the random walk principle will be applied in order to calculate those traffic characteristics which are connected with the displacement of calls. The state {success} of our first random walk shall be designated as the departure from the system of the observed call (either that the call successfully terminates its service or that it is displaced from the system). Each i-call reaches this state with certainty; therefore, the success probability E_j equals unity.

The jump rates of this random walk are: An i-call in place j proceeds to place $j-1$, when one of the $j-1$ calls which precede it, terminates service:

$$u_{j-1,j}=\mu_{j-1}\,. \quad (41)$$

An i-call in place $j<n+s$ recedes to place $j+1$, when an $<i$-call arrives:

$$u_{j+1,j}={}_{<i}\lambda\,. \quad (42)$$

An i-call leaves its place j, when a $<i$-call arrives or when one of its $j-1$ preceeding calls or the observed call itself terminates its service:

$$u_{j,j}={}_{<i}\lambda+\mu_j\,. \quad (43)$$

Substituting these jump rates in eq. (38), we get the following system of differential equations for the probability: $G_j(x)$ that an i-call stays in the system for a time greater x:

$$\begin{aligned} {}_iG_1'(x)&=-({}_{<i}\lambda+\mu_1)\,{}_iG_1(x)+{}_{<i}\lambda\,{}_iG_2(x)\,,\\ {}_iG_j'(x)&=\mu_{j-1}\,{}_iG_{j-1}(x)+{}_{<i}\lambda\,{}_iG_{j+1}(x)-\\ &\quad-({}_{<i}\lambda+\mu_j)\,{}_iG_j(x)\,,\\ {}_iG_{n+s}'(x)&=\mu_{n+s-1}\,{}_iG_{n+s-1}(x)-\\ &\quad-({}_{<i}\lambda+\mu_{n+s})\,{}_iG_{n+s}(x)\,. \end{aligned} \quad (44)$$

Since the success probabilities of this random walk equal 1, we get as initial values

$$G_j(x=0)=1\,. \quad (45)$$

As ${}_iG_j$, we denote the total time, which an i-call, starting in place j, stays in the system. For this mean total time ${}_iG_j$ of an i-call in place j we obtain from eq. (40)

$$\begin{aligned} ({}_{<i}\lambda+\mu_1)\,{}_iG_1-{}_{<i}\lambda\,{}_iG_2&=1\,,\\ -\mu_{j-1}\,{}_iG_{j-1}+({}_{<i}\lambda+\mu_j)\,{}_iG_j-{}_{<i}\lambda\,{}_iG_{j+1}&=1\,,\\ -\mu_{n+s-1}\,{}_iG_{n+s-1}+({}_{<i}\lambda+\mu_{n+s})\,{}_iG_{n+s}&=1\,. \end{aligned} \quad (46)$$

This inhomogenous linear system of equations is basical for the following calculations. Therefore, its solution for an arbitrary "right hand side" C_j is given in the Appendix.

Beyond that the characteristics for calls in place j (e.g.: ${}_iG_j$) are weighted with the probabilities ${}_iF_j$, that an i-call starts in place j. Thus we obtain directly the traffic characteristics of an i-call independently of its starting place.

In order to get the mean total time ${}_iG$, which an arbitrary i-call (all lost i-calls included) stays in the system, we substitute the "right-hand side" $C_j=1$ into eq. (95). If we notice eq. (26) for the traffic ${}_iX$ carried by the system, we obtain the relation

$${}_iX={}_i\lambda\,{}_iG\,. \quad (47)$$

Eq. (47) can be interpreted easily: The mean number ${}_iX$ of i-calls in the system is equal to the mean number of i-calls which arrive during the mean total time. This theorem is valid in the priorityless case. Eq. (47) proves that it holds true also in the preemptive case where displacements and interruptions occur.

When we sum up the ${}_iX$ of eq. (47) over all i, we obtain the mean total time ${}_{\leq r}G$ of an arbitrary call averaged among all classes by means of

$$ {}_{\leqq r}X = \sum_{i=1}^{r} {}_i\lambda\, {}_iG = {}_{\leqq r}\lambda\, {}_{\leqq r}G\,. \tag{48}$$

In order to derive the probability ${}_iV$ that an i-call is displaced from the system by an $<i$-call, we define a new random walk. The state {success} shall be identified with the displacement of the observed i-call. This state can be reached only from place $n+s$. The corresponding jump rate is

$$u_{n+s,\,0} = {}_{<i}\lambda\,. \tag{49}$$

All other jump rates $u_{j,0}$ equal to 0.

The success probability of this random walk is the probability ${}_iV_j$, that an i-call is displaced, if it starts its random walk in place j. We get from eq. (93)

$$ {}_iV_j = \frac{{}_{<i}\Phi_{j-1}}{{}_{<i}\Phi_{n+s}}. \tag{50}$$

From eq. (95) we get the probability ${}_iV$ that an arbitrary i-call is displaced:

$$ {}_iV = \frac{{}_{\leqq i}\lambda}{{}_i\lambda}\left(\frac{1}{{}_{\leqq i}\Phi_{n+s}} - \frac{1}{{}_{<i}\Phi_{n+s}}\right). \tag{51}$$

Using eq. (19) we can write

$$ {}_i\lambda\, {}_iV = {}_{\leqq i}\lambda\,({}_{\leqq i}P_{n+s} - {}_{<i}P_{n+s})\,. \tag{52}$$

At the left-hand side we have the mean number of i-calls which are displaced per time unit. This number has to equal the mean number of $<i$-calls which displace i-calls. A displacement of an i-call can arise when all $(n+s)$ places are occupied by $\leqq i$-calls, but not all by $<i$-calls.

The probability ${}_iB$ that an i-call is lost (either at its arrival or by displacement), results in

$$ {}_iB = {}_iC + {}_iV\,. \tag{53}$$

The probability of loss for class $\leqq i$ is obtained by summation:

$$ {}_{\leqq i}B = \sum_{\nu=1}^{i} {}_\nu\lambda\, {}_\nu B = {}_iC \tag{54}$$

and in particular:

$$ {}_{\leqq r}B = {}_rC = B\,. \tag{55}$$

This overall probability of loss in the preemptive system equals the probability of loss B in the system without priorities.

Next, we substitute relation (54) into eq. (29). Thus, we get a new expression for the traffic carried by the servers:

$$ {}_{\leqq i}Y = {}_iA\,(1 - {}_{\leqq i}B)\,, \quad {}_iY = {}_iA\,(1 - {}_iB)\,. \tag{56), (57}$$

This relation (57) is not so obvious as the corresponding one in the system without priorities, because in the preemptive case also those calls which get lost by displacement contribute to the traffic carried by the servers. ${}_iY$ can be rewritten as

$$ {}_iY = {}_ip\, {}_{\leqq r}Y + {}_iA\,({}_{\leqq r}B - {}_iB)\,. \tag{58}$$

${}_ip\, {}_{\leqq r}Y$ is the traffic of class i which would be carried by the servers in the priorityless system. The traffic ${}_iY$ of class i is greater or smaller correspondingly as the probability of loss ${}_iB$ of class i is smaller or greater than the probability $B = {}_{\leqq r}B$ of the priorityless loss-delay system.

To clarify the meaning of the different total times which the calls spend in the system, an example may be given:

We observe z i-calls from which z_0 are lost at their arrival, z_1 are displaced and z_2 are successful (with or without interruption). We sum up the total times of the three different classes. Those calls which are lost at their arrival, have the total time 0 and thus the sum x_0 of their total times equals 0. Let us assume that the sum of all total times of displaced calls is x_1 and the sum of all "successfull" total times is x_2. The sum of the total times of all calls is denoted by $x = x_0 + x_1 + x_2$. Then the ratio x/z corresponds to the total time ${}_iG$. The following terms give analogous relations:

$$ {}_iC = z_0/z\,, \quad {}_iV = z_1/z\,, \quad {}_iB = z_0 + z_1\,,$$
$$ {}_iG = x/z\,, \quad {}_iJ = x_1/z\,, \quad {}_iL = x_1/z_1\,, \quad {}_iD = x_2/z_2\,.$$

The traffic characteristics ${}_iJ$, ${}_iL$, and ${}_iD$ will be derived in the following.

In order to get ${}_iD$ (the mean total time of an unsuccessful call) we investigate that random walk, where the state {success} is defined as the displacement of the observed call. ${}_iJ_j(x)$ is the probability that an i-call is displaced and that it stays in the system a time greater x until it is displaced under the condition that it starts in place j.

Then ${}_iJ_j(x)$ satisfies the same system (44) of differential equations as does ${}_iG(x)$, but with the initial values

$$C_j = {}_iV_j\,. \tag{59}$$

The mean total time ${}_iJ$ of an unsuccessful i-call related to all i-calls results from eq. (95) with the "right-hand side"

$$ {}_iJ_j(x=0) = {}_iV_j\,. \tag{60}$$

We get

$$ {}_i\lambda\, {}_iJ = \sum_{j=1}^{n+s} {}_iX_j\, {}_iV_j\,. \tag{61}$$

In eq. (61) we use the term ${}_iX_j$, which has been introduced in eq. (32) as the traffic of class i carried by place j. The term ${}_iX_j\,{}_iV_j$ is that share of the carried traffic ${}_iX_k$ of place j which comes from those i-calls which are displaced later on. Thus, we can interpret the "right-hand side" of eq. (61) as the unsuccessful share of traffic carried by the system. This unsuccessful traffic is not equal to ${}_iV\,{}_iX$. This is reasonable, since the displacement probabilities of the various places are not the same. The greater j, the greater is the displacement probability ${}_iV_j$ of place j.

We get the mean total time ${}_iD$ that a displaced call stays in the system, by dividing ${}_iJ$ by ${}_iV$:

$$ {}_iD = {}_iJ/{}_iV\,. \tag{62}$$

Correspondingly, the mean total time ${}_iL$ that a successful call stays in the system results in

$$ {}_iL = ({}_iG - {}_iJ)/{}_iS\,. \tag{63}$$

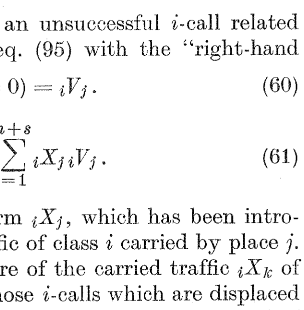

Fig. 1.
The total time L of a successfull call in the priorityless system, the total times ${}_1L$ and ${}_2L$ of a successful call of class 1 and 2 resp. and the total time ${}_2D$ of an unsuccessful 2-call as functions of the number s of waiting places.

The different total times of a preemptive loss-delay system are shown in Fig. 1 as a function of the number s of waiting places. We observe the obvious effect that the total times increase as s increases. For $s = 0$ we have the preemptive loss system, in which the total times are equal to the service times. For the priorityless loss system and for class 1 of the preemptive loss system the service time of a successful call equals the mean service time $h = 1$. However, the mean service time of a successful call of class 2 is smaller than h, since particularly those 2-calls, which have a long service time, are interrupted. For small values of s, only a few 2-calls are successful, but these have a small total time. As s increases more 2-calls are successful, but the total time increases too. The mean time ${}_2D$, during which an unsuccessful call stays in the system, increases analogously with increasing s.

5. Waiting Characteristics

This section deals with those traffic characteristics which are related to the waiting probability.

In general, the total waiting time in the preemptive loss-delay system is composed of several sections: The first waiting time, a second one following the first interruption and so on.

In order to investigate the first waiting time we consider the random walk of a waiting i-call within the queue. The state {success} shall be specified by the departure from the waiting room (either by beginning with its service or by being displaced). The distribution of the first waiting time can be solved in an explicit form, but this will not be treated here. We will confine ourselves to the mean first waiting time ${}_iT$ of an arbitrary i-call. Using a modified form of eq. (95) the following formulas can be derived:

$$ {}_i\lambda\,{}_iT = {}_{\leqq i}Q - \frac{{}_iE}{{}_{i-1}E}\,{}_{\leqq i}Q \tag{64}$$

where ${}_iE$ is the probability that an i-call cannot be served immediately at its arrival. Together with eqs. (21) and (23) we obtain

$$ {}_iE = {}_iC + {}_iR = \frac{{}_{\leqq i}\Psi_s}{{}_{\leqq i}\Phi_{n+s}}\,. \tag{65}$$

Please notice that we do not have an analogous result as in the priorityless system ($\lambda\,T = Q$). The reason being that the waiting line of class i is not only maintained by the incoming i-calls, but also by those which are interrupted. Only in the case $i = 1$, where no interruptions arise, do we get the corresponding formulas

$$ {}_1\lambda\,{}_1T = {}_1Q\,. \tag{66}$$

The mean waiting time ${}_iT_{n+1}$ of an i-call starting in place $n+1$ results in

$$ {}_iT_{n+1} = \frac{1}{{}_{<i}\lambda}\,\frac{{}_{<i}\Phi_{n+s}}{{}_{<i}\Psi_s}\,{}_{<i}Q\,. \tag{67}$$

${}_iT_{n+1}$ is also the (next) mean waiting time of an interrupted i-call, since each interrupted call begins its (next) waiting time in place $n+1$.

By means of another random walk we obtain the probability ${}_iK_{n+j}$ that an i-call starting in waiting place j is displaced without having reached a server:

$$ {}_iK_{n+j} = \frac{{}_{<i}\Psi_{k-1}}{{}_{<i}\Psi_s}\,. \tag{68}$$

The unconditioned probability that an arbitrary i-call is lost by displacement without having been in service results in

$$ {}_iK = \frac{{}_{\leqq i}p}{{}_ip}\left(\frac{1}{{}_{\leqq i}\Phi_{n+1}} - \frac{{}_iE}{{}_{i-1}E}\,\frac{1}{{}_{<i}\Phi_{n+s}}\right). \tag{69}$$

This probability of immediate displacement differs from the overall probability ${}_iV$ of displacement by the same factor which already appears in eq. (64).

6. Interruption Characteristics

In the last part of this paper the probability of interruption and related characteristics shall be determined. In the associated random walk an i-call is successful — in the sense of the random walk — if it is interrupted. Substituting the jump rates of this random walk in eq. (37), we get the following system of equations for the probability ${}_iU_j$ that an i-call is interrupted under the condition that it starts in place j:

$$\begin{aligned}
(\mu_1 + {}_{<i}\lambda)\,{}_iU_1 - {}_{<i}\lambda\,{}_iU_2 &= 0\,,\\
-\mu_{j-1}\,{}_iU_{j-1} + (\mu_j + {}_{<i}\lambda)\,{}_iU_j - {}_{<i}\lambda\,{}_iU_{j+1} &= 0\,, \quad j \neq n\,,\\
-\mu_{n-1}\,{}_iU_{n-1} + (\mu_n + {}_{<i}\lambda)\,{}_iU_n &= {}_{<i}\lambda\,,\\
-\mu_{n+s-1}\,{}_iU_{n+s-1} + (\mu_n + {}_{<i}\lambda)\,{}_iU_{n+s} &= 0\,;
\end{aligned} \tag{70}$$

$$ j = 1, 2, \ldots, n\,, \qquad {}_iU_j = \frac{{}_{<i}\Phi_{j-1}}{{}_{<i}\Phi_n}\,, \tag{71}$$

$$ j = 1, 2, \ldots, s\,, \qquad {}_iU_{n+j} = {}_iU_n(1 - {}_iK_{n+j})\,. \tag{72}$$

The last formulas can be interpreted such that an i-call, which starts in place $n+j$ and which will be interrupted later on, has at first to reach a server. This arises with probability $(1 - {}_iK_{n+j})$, known from eq. (68). Now, the i-call starts service in place n. From here the probability of interruption is ${}_iU_n$. It follows from eq. (72) that both these events are independent. Before using this fact, we calculate the unconditional probability ${}_iU$ that an i-call is interrupted:

$$ {}_iU = \sum_{j=1}^{n} {}_iF_j\,{}_iU_j + ({}_iR - {}_iK)\,{}_iU_n\,. \tag{73}$$

The sum can be evaluated using eq. (95) in the Appendix:

$$ \sum_{j=1}^{n} {}_iF_j\,{}_iU_j = \frac{{}_{<i}\lambda}{{}_i\lambda}\,\frac{n^s}{{}_{\leqq i}A^s}\,\frac{1}{{}_{\leqq i}\Phi_{n+s}}\left(1 - \frac{{}_{\leqq i}\Phi_n}{{}_{<i}\Phi_n}\right). \tag{74}$$

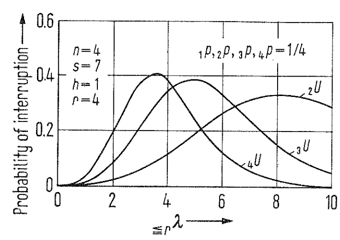

Fig. 2. Interruption probabilities ${}_iU$ as a function of the total arrival rate ${}_{\leqq r}\lambda$.

For the loss-delay system as in Fig. 2 the probability of interruption ${}_iU$ for classes $i = 2, 3, 4$ is shown as a function of the total arrival rate ${}_{\leqq r}\lambda$. For small ${}_{\leqq r}\lambda$ the probabilities ${}_iU$ increase with increasing ${}_{\leqq r}\lambda$. Then after having reached a maximum, they decrease and tend towards 0, since with increasing ${}_{\leqq i}\lambda$, more and more i-calls are lost at their arrival or are displaced without reaching a server and thus without interruptions.

A call has to wait, either if it starts at a waiting place or if it starts in a server and then is interrupted. Thus, the probability ${}_iW$ that an arbitrary i-call has to wait results in

$$ {}_iW = {}_iR + \sum_{j=1}^{n} {}_iF_j\,{}_iU_j\,. \tag{75}$$

The independence stated in eq. (72) and discussed above, enables us to directly determine the probability ${}_i^{\nu}U$ that an arbitrary i-call is interrupted ν times:

$$ {}_i^{\nu}U = {}_iU\,{}_iU_{n+1}^{\nu-1}(1 - {}_iU_{n+1})\,. \tag{76}$$

With eq. (76) we can calculate the mean number ${}_iN$ of interruptions of an i-call:

$$ {}_iN = \sum_{\nu=1}^{\infty} \nu\,{}_i^{\nu}U = \frac{{}_iU}{1 - {}_iU_{n+1}}\,. \tag{77}$$

The mean number of interruptions of an arbitrary call amounts to

$$ {}_{\leqq r}N = \sum_{i=1}^{r} {}_{i}p\, {}_{i}N \,. \tag{78}$$

In Section 5 we calculated in eq. (64) the mean first waiting time ${}_{i}T$ and in eq. (67) the mean waiting time in place $n+1$, which is valid for interrupted calls. Noticing that the mean number ${}_{i}N$ of interruptions for an arbitrary i-call is given by eq. (77), we are able to calculate the mean total waiting time ${}_{i}M$ of an i-call:

$$ {}_{i}M = {}_{i}T + {}_{i}N\, {}_{i}T_{n+1} \,. \tag{79}$$

Finally, we calculate the mean number Z of interruptions per time unit. This characteristic is important for the management of the system:

$$ Z = \frac{{}_{\leqq i}X}{{}_{\leqq i}G}\, {}_{\leqq r}N \,. \tag{80}$$

Substituting eqs. (78) and (48) into eq. (80) we get the following result for the mean number of interruptions per unit of time:

$$ Z = {}_{\leqq r}\lambda\, {}_{\leqq r}N \,. \tag{81}$$

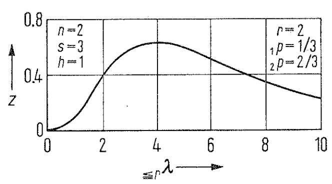

Fig. 3. The mean number of interruptions per time unit as a function of the total arrival rate ${}_{\leqq r}\lambda$.

Fig. 3 shows the mean number Z of interruptions per time unit as a function of the total arrival rate ${}_{\leqq r}\lambda$. For small ${}_{\leqq r}\lambda$ only few interruptions occur. As ${}_{\leqq r}\lambda$ increases, Z increases also, reaches a maximum and then decreases, since for high arrival rates the 2-calls are displaced more and more from the system and from the servers so that fewer interruptions take place.

Appendix

In the Appendix the solutions of two systems of equations are given, which are important in order to explicitly solve the displacement-, the waiting- and the interruption characteristics. The solution will be presented in terms of the basic function $\Phi_{\nu,\varkappa}$. For this basic function the definition and some calculating rules are stated. We will define the basic function $\Phi_{\nu,\varkappa}$ as a function of the traffic offered A and of the traffic parameter β_ϱ, respectively. According to eq. (14) we have

$$ \beta_\varrho = \mu_\varrho / \lambda \,. \tag{82}$$

The basic function $\Phi_{\nu,\varkappa}$ is given by the definition

$$ \Phi_{\nu,\varkappa} = \sum_{\sigma=\nu}^{\varkappa} \prod_{\varrho=\sigma+1}^{\varkappa} \beta_\varrho \,. \tag{83}$$

The basic function ${}_{\leqq i}\Phi_{\nu,\varkappa}$ results by substituting A by ${}_{\leqq i}A$ (and as a consequence β_ϱ by ${}_{\leqq i}\beta_\varrho$). If $\nu = 0$, then we use the abbreviation of eq. (16):

$$ \Phi_{0,\varkappa} = \Phi_\varkappa \,. \tag{84}$$

Since β_ϱ is equal to β_n for $\varrho \geqq n$, the basic function $\Phi_{\nu,\varkappa}$ depends only on the difference $\varkappa - \nu$ for $\nu \geqq n$. Therefore we abbreviate (cf. eq. (18)):

$$ \nu \geqq n: \quad \Psi_\varrho = \Phi_{\nu,\nu+\varrho} = \sum_{\sigma=0}^{\varrho} \beta_n^\sigma \,. \tag{85}$$

The basic function satisfies the relation

$$ \Phi_{\nu,\varkappa} = \Phi_{\sigma,\varkappa} + \Phi_{\nu,\sigma-1} \prod_{\varrho=\sigma}^{\varkappa} \beta_\varrho \tag{86}$$

and in particular for $\sigma = \nu + 1$ and $\sigma = \varkappa$, resp.

$$ \Phi_{\nu,\varkappa} = \Phi_{\nu+1,\varkappa} + \prod_{\varrho=\nu+1}^{\varkappa} \beta_\varrho \,, \tag{87}$$

$$ \Phi_{\nu,\varkappa} = 1 + \beta_\varkappa \Phi_{\nu,\varkappa-1} \,. \tag{88}$$

By means of the above recurrence formulas we can write $\Phi_{\nu,\varkappa}$ as an explicit function of the traffic offered A. We get different formulas according to the range of ν and $\varkappa$:

$$ \varkappa \leqq n: \quad \Phi_{\nu,\varkappa} = \frac{\varkappa!}{A^\varkappa} \sum_{\sigma=\nu}^{\varkappa} \frac{A^\sigma}{\sigma!} \,, \tag{89}$$

$$ \nu \geqq n: \tag{90}$$

$$ \Phi_{\nu,\varkappa} = \Psi_{\varkappa-\nu} = \begin{cases} \dfrac{A}{n-A}\left(\dfrac{n^{\varkappa-\nu+1}}{A^{\varkappa-\nu+1}} - 1\right) & \text{for} \quad A \neq n \,, \\ \varkappa - \nu + 1 & \text{for} \quad A = n \,, \end{cases}$$

$$ \nu < n < \varkappa: \qquad \Phi_{\nu,\varkappa} = \Psi_{\varkappa-n-1} + (n/A)^{\varkappa-n}\, \Phi_{\nu,n} \,. \tag{91}$$

The relations (83) up to (91) were important in order to prove various formulas of this paper. In particular, the solution of the following systems of eqs. (92) and (96) is based on the "calculating-tool" presented by the basic function $\Phi_{\nu,\varkappa}$.

The system of equations, solution of which shall be given in this Appendix, reads as follows (cf. eq. (46)):

$$ \begin{aligned} (\mu_1 + {}_{<i}\lambda)\, {}_{i}X_1 - {}_{<i}\lambda\, {}_{i}X_2 &= C_1 \,, \\ -\mu_{k-1}\, {}_{i}X_{k-1} + (\mu_k + {}_{<i}\lambda)\, {}_{i}X_k - {}_{<i}\lambda\, {}_{i}X_{k+1} &= C_k \,, \\ -\mu_{n+s-1}\, {}_{i}X_{n+s-1} + (\mu_{n+s} + {}_{<i}\lambda)\, {}_{i}X_{n+s} &= C_{n+s} \,. \end{aligned} \tag{92}$$

For an arbitrary "right-hand side" C_j the solution of the system (92) of equations is given by

$$ {}_{i}X_k = \frac{1}{{}_{<i}\lambda\, {}_{<i}\Phi_{n+s}} \times \tag{93}$$

$$ \times \left({}_{<i}\Phi_{k,n+s} \sum_{\nu=1}^{k-1} C_\nu\, {}_{<i}\Phi_{\nu-1} \prod_{\varrho=\nu}^{k-1} {}_{<i}\beta_\varrho + {}_{<i}\Phi_{k-1} \sum_{\nu=k}^{n+s} C_\nu\, {}_{<i}\Phi_{\nu,n+s} \right).$$

The index k of ${}_{i}X_k$ represents the starting place, to which the characteristic ${}_{i}X_k$ is related. We get the corresponding characteristic ${}_{i}X$, which is averaged among all starting places by means of the probability ${}_{i}F_k$, that an i-call starts in place j:

$$ {}_{i}X = \sum_{k=1}^{n+s} {}_{i}F_k\, {}_{i}X_k \,. \tag{94}$$

An extensive calculation results in the following relation, which allows the direct calculation of many important traffic characteristics:

$$ {}_{i}X = \sum_{\nu=1}^{n+s} C_\nu \left(\frac{{}_{\leqq i}\Phi_{\nu,n+s}}{{}_{\leqq i}\Phi_{n+s}} - \frac{{}_{<i}\Phi_{\nu,n+s}}{{}_{<i}\Phi_{n+s}} \right). \tag{95}$$

Substituting "actual" right-hand sides C_j into eqs. (93) and (95), we obtain the various traffic characteristics presented in this paper.

References

[1] Brandt, G. J., Das preemptive Warte-Verlust-System. Thesis, University of Stuttgart 1971.

[2] Feller, W., An introduction to probability theory and its application. J. Wiley & Sons, New York 1968.

[3] Prabhu, N. U., Stochastic processes. Macmillan Co., New York 1965.

[4] Riordan, J., Stochastic service systems. J. Wiley & Sons, New York 1962.

[5] Wagner, W., Über ein kombiniertes Warte-Verlustsystem mit Prioritäten. Thesis, University of Stuttgart 1968.