A MULTISERVER QUEUING SYSTEM WITH PREEMPTIVE PRIORITY

Joachim G. Brandt

Technical University Stuttgart

Stuttgart, Federal Republic of Germany

ABSTRACT

The investigation of waiting-lines is a basic
problem in computerized data-transmitting net-
works, in high-organized computer-systems and
in many other technical, but also economical
situations. In this paper the solution of a wai-
ting-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-pla-
ces is dealt with. This loss-delay-system inclu-
des two important special cases:

(1) For s=0 we have no waiting possibility so
    that we have the preemptive loss-system,
    where the interrupted calls are lost.

(2) For s → ∞ no call is lost so that we have
    the preemptive delay-system.

In my doctor-thesis /1/ 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.

A first class of traffic characteristics results
from the probabilities of state. Further traffic
characteristics can be calculated by the RANDOM-
WALK principle. As shown in /1/, the RANDOM-WALK
principle is not confined to the loss-delay-sys-
tem dealt with, but represents a method of how
to determine the traffic-characteristics of gen-
eral MARKOV-systems.

The basic function $\phi_{\nu,\kappa}$, which in /1/ has been
introduced to simplify the known formulae of the
loss-delay-system without priorities proves it-
self to be an important tool for handling the
preemptive loss-delay-system.

I INTRODUCTION

The loss-delay-system has n servers and s wait-
ing places. We have r priority classes so that
class i has absolute priority over all calls of
class i+1, i+2,... up to class r. Joining the
calls of class 1,2,... up to class i, we speak
of class ≤i, accordingly of class <i or >i. A
call of class ≤r is any call without regard to
its individual priority class. A call of class i
is shortly denoted as "i-call".
The traffic-characteristics of a certain pri-
ority class are labeled with the class index
preceding the abbreviation of the characteri-
stic. For instance the probability that a call
of class ≤i is lost is denoted by $_{\leq i}B$.

We investigate the preemptive priority case as-
suming the first come - first served rule with-
in 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 ν≤n represents a server and a place
ν>n a waiting place.

This order of precedence has to be described in
the following two situations:

1. There are j calls in the system and one of
   the served calls terminates its service.
   Therefore, a place ν (with ν≤n) becomes free.
   The calls, having until now occupied the pla-
   ces ν+1, ν+2,...,j, advance one place so
   that they occupy the places ν, ν+1,...,j-1.
   The call of place n+1 proceeds to place n,
   that means this call now finds 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
   ≤ i-calls. If all n+s places are occupied
   by ≤i-calls, then the arriving call is lost.
   If the call occupies a place ν≤j, then the
   calls in place ν, ν+1,...up to j are pushed
   back one place so that they now occupy the
   places ν+1, ν+2,...,j+1. A call which is
   pushed back from place n to place n+1, inter-
   rupts its service and has to wait 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 distribu-
ted negative exponentially. The mean number of
arriving calls per unit of time is denoted as
the total arrival rate $_{\leq r}\lambda$. The share of class i
is denoted by $_ip$ so that:

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

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

$$_i\lambda = {}_ip \cdot {}_{\leq r}\lambda \tag{2}$$

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 is made up of the following: Let us consider a call which is still in service. The call's remaining service time does not depend on the time which has already been spent in service. The remaining service time is distributed again negative exponentially with the mean h. This 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. Consequently, we need not distinguish non-interrupted and interrupted calls.

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

$$\mu_j = \begin{cases} \dfrac{j}{h} & j \leq n \\ \dfrac{n}{h} & j > n \end{cases} \qquad (3)$$

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

$$_iA = h \cdot {}_i\lambda \qquad (4)$$

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

II THE PROBABILITIES OF STATE AND THE RELATED TRAFFIC CHARACTERISTICS

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

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

$$_{\leq i}P_j(t) = P\{_{\leq i}X(t) = j\} \qquad (5)$$

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:

(1) from state j-1 with probability
$_{\leq i}\lambda \cdot \Delta t + o(\Delta t)$

(2) from state j+1 with probability
$\mu_{j+1} \cdot \Delta t + o(\Delta t)$

(3) from state j for j < n+s with probability
$1 - (_{\leq i}\lambda + \mu_j) \cdot \Delta t + o(\Delta t)$

and for j = n+s with probability
$1 - \mu_{n+s} \cdot \Delta t + o(\Delta t)$

(4) from another 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:

$$\begin{aligned}
_{\leq i}P_0'(t) &= -_{\leq i}\lambda \cdot {}_{\leq i}P_0(t) + \mu_1 \cdot {}_{\leq i}P_1(t) \\
_{\leq i}P_j'(t) &= {}_{\leq i}\lambda \cdot {}_{\leq i}P_{j-1}(t) - (_{\leq i}\lambda + \mu_j) \cdot {}_{\leq i}P_j(t) + \\
&\quad + \mu_{j+1} \cdot {}_{\leq i}P_{j+1}(t) \qquad (6) \\
_{\leq i}P_{n+s}'(t) &= {}_{\leq i}\lambda \cdot {}_{\leq i}P_{n+s-1}(t) - \mu_{n+s} \cdot {}_{\leq i}P_{n+s}(t)
\end{aligned}$$

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

$$\sum_{j=m}^{n+s} {}_{\leq i}P_j'(t) = {}_{\leq i}\lambda \cdot {}_{\leq i}P_{m-1}(t) - \mu_m \cdot {}_{\leq i}P_m(t) \qquad (7)$$

Summing up these equations again, we get:

$$\sum_{m=1}^{n+s} \sum_{j=m}^{n+s} {}_{\leq i}P_j'(t) = \qquad (8)$$

$$_{\leq i}\lambda \cdot \sum_{m=0}^{n+s-1} {}_{\leq i}P_m(t) - \sum_{m=1}^{n+s} \mu_m \cdot {}_{\leq i}P_m(t)$$

For i=r the system (6) of equations is identical with the system for the probabilities of state $P_j(t)$ of the priorityless loss-delay-system. Therewith we conclude that the total state process $_{\leq r}X(t)$ of the preemptive system behaves, as if the priority classification does not exist. This theorem is analogously valid for the state process $_{\leq i}X(t)$ by replacing the arrival rate: This process $_{\leq i}X(t)$ of the preemptive system behaves as does the state process of the priorityless system, if only the calls of class $\leq i$ are offered. Consequently, the process $_{\leq i}X(t)$ behaves, as if the calls of class >i were not existent.

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 following determinant:

$$_iD(\xi) = \begin{vmatrix} _{\leq i}\lambda+\xi & -\mu_1 & 0 & \cdots & & 0 \\ -_{\leq i}\lambda & _{\leq i}\lambda+\mu_1+\xi & -\mu_2 & \cdots & & 0 \\ 0 & -_{\leq i}\lambda & & \ddots & & 0 \\ \vdots & \vdots & & \ddots & & -\mu_{n+s} \\ 0 & 0 & & -_{\leq i}\lambda & & _{\leq i}\lambda+\mu_{n+s}+\xi \end{vmatrix} \qquad (9)$$

One root of this polynomial is $_i\xi=0$, which corresponds to the stationary solution $_{\leq i}P_j$. 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 following form:

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

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

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

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

$$_{\leq i}\lambda \cdot {}_{\leq i}P_{m-1} = \mu_m \cdot {}_{\leq i}P_m \qquad (12)$$

The probabilities of state satisfy the condition

$$\sum_{j=0}^{n+s} {}_{\leq i}P_j = 1 \qquad (13)$$

Using the traffic-parameter $_{\leq i}\beta_j$:

$$_{\leq i}\beta_j = \frac{\mu_j}{_{\leq i}\lambda} \qquad (14)$$

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

$$_{\leq i}P_j = \frac{\prod_{m=j+1}^{n+s} {}_{\leq i}\beta_m}{\sum_{j=0}^{n+s} \prod_{m=j+1}^{n+s} {}_{\leq i}\beta_m} \qquad (15)$$

The denominator of this expression agrees with 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 following abbreviations.

$$\Phi_\kappa = \Phi_{0,\kappa} \qquad (16)$$

$$_{\leq i}\Phi_{\nu,\kappa} = {}_{\leq i}\Phi_{\nu,\kappa}(_{\leq i}A) \qquad (17)$$

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

$$\Psi_\rho = \phi_{\nu,\nu+\rho} \qquad (18)$$

Using the basic function $\phi_{\nu,\kappa}$, the probability of state can be stated in the following form:

$$_{\le i}P_{n+s} = \frac{1}{_{\le i}\phi_{n+s}} \qquad (19)$$

$$_{\le i}P_j = {}_{\le i}P_{n+s} \cdot \prod_{m=j+1}^{n+s} {}_{\le i}\beta_j$$

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

$$_iF_j = {}_{\le i}P_j \qquad (20)$$

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:

$$_iC = {}_{\le i}P_{n+s} = \frac{1}{_{\le i}\phi_{n+s}} \qquad (21)$$

and satisfies - in accordance with the priority rule - the relation:

$$_1C < {}_2C < \ldots < {}_rC \quad \text{with } {}_rC = B \qquad (22)$$

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:

$$_iR = \frac{n}{_{\le i}A} \cdot \frac{_{\le i}\Psi_{s-1}}{_{\le i}\phi_{n+s}} \qquad (23)$$

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

$$_{\le i}X_j = \sum_{\nu=j}^{n+s} {}_{\le i}P_\nu = \frac{_{\le i}\phi_{j,n+s}}{_{\le i}\phi_{n+s}} \qquad (24)$$

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

$$_{\le i}X = \sum_{j=1}^{n+s} j \cdot {}_{\le i}P_j \qquad (25)$$

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

$$_{\le i}X = \sum_{j=1}^{n+s} {}_{\le i}X_j \qquad (26)$$

Therefore, $_{\le i}X_j$ can be interpreted as the traffic of class i carried by place j.
Analogically, we gain the traffic $_{\le i}Y$ of class $\le i$ carried by the servers:

$$_{\le i}Y = \sum_{j=1}^{n} {}_{\le i}X_j \qquad (27)$$

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

$$_{\le i}Y = h \cdot \sum_{j=1}^{n+s} \mu_j \cdot {}_{\le i}P_j \qquad (28)$$

Introducing the stationary conditions (11) into equation (8), we obtain:

$$_{\le i}Y = {}_{\le i}A \cdot (1 - {}_{\le i}P_{n+s}) \qquad (29)$$

We will return to this formula in chapter IV.
The mean queue length of class $\le i$ results in:

$$_{\le i}A = n \qquad _{\le i}Q = \frac{_{\le i}A}{n - {}_{\le i}A} \cdot \frac{_{\le i}\Psi_{s} - (s+1)}{_{\le i}\phi_{n+s}} \qquad (30)$$

$$_{\le i}A \ne n \qquad _{\le i}Q = \frac{1}{2} \cdot s \cdot (s+1) \cdot \frac{1}{_{\le i}\phi_{n+s}}$$

We have the relation:

$$_{\le i}X = {}_{\le i}Y + {}_{\le i}Q \qquad (31)$$

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

$$_iX_j = {}_{\le i}X_j - {}_{<i}X_j = \frac{_{\le i}\phi_{j,n+s}}{_{\le i}\phi_{n+s}} - \frac{_{<i}\phi_{j,n+s}}{_{<i}\phi_{n+s}} \qquad (32)$$

III THE RANDOM-WALK PRINCIPLE

After having calculated the probabilities of state in the last chapter, 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, which 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 to these problems in /1/. The RANDOM-WALK principle shall be dealt with in this paper as far as it is necessary for the preemptive loss-delay-system in case of an infinite number of sources.
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 formulae 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,... 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 chapter IV, 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:

$$P\{Y(t+\Delta t)=j \mid Y(t)=k\} = u_{k,j} \cdot \Delta t + o(\Delta t) \qquad (33)$$

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

$$\sum_{j=0}^{n+s+1} P\{Y(t+\Delta t)=j \mid Y(t)=k\} = 1 \qquad (34)$$

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

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:

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

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

$$P\{Y(t+\Delta t)=k \mid Y(t)=k\} = 1 - u_{k,k} \cdot \Delta x + o(\Delta x) \qquad (36)$$

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 following system of equations:

$$u_{j,j} \cdot E_j - \sum_{\substack{k=1 \\ k \neq j}}^{n+s} u_{j,k} \cdot E_k = u_{j,o} \qquad (37)$$

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:

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

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

$$Z_j(x=0) = E_j \qquad (39)$$

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:

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

This equation results by integration from the differential system (38). The system of linear equations (4o) for the Z_j is identical with the system (37) for the probabilities E_j of success, but with other "right hand sides".

By means of the systems (37), (38) and (40), we can calculate the probability E_j of success in place j, the distribution $Z_j(x)$ and the mean time Z_j of a successful call in place j related to all calls in place j. In the derivation of these results we do not need to assume that the process is stationary. These conditioned characteristics are valid also for the time-dependent process.

IV 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 the following: An i-call in place j proceeds to place j-1, when one of the j-1 calls, which precede it, terminates service:

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

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

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

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:

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

Substituting these jump rates in (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:

$${}_iG_1'(x) = -({}_{<i}\lambda+\mu_1)\cdot{}_iG_1(x) + {}_{<i}\lambda\cdot{}_iG_2(x)$$
$${}_iG_j'(x) = \mu_{j-1}\cdot{}_iG_{j-1}(x) + {}_{<i}\lambda\cdot{}_iG_{j+1}(x) +$$
$$\qquad - ({}_{<i}\lambda+\mu_j)\cdot{}_iG_j(x) \qquad (44)$$
$${}_iG_{n+s}'(x) = \mu_{n+s-1}\cdot{}_iG_{n+s-1}(x) +$$
$$\qquad - ({}_{<i}\lambda+\mu_{n+s})\cdot{}_iG_{n+s}(x)$$

Since the success probabilities of this RANDOM-WALK equal 1, we get as initial values:

$$G_j(x=0) = 1 \qquad (45)$$

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 equation (40):

$$({}_{<i}\lambda+\mu_1)\cdot{}_iG_1 - {}_{<i}\lambda\cdot{}_iG_2 = 1$$
$$- \mu_{j-1}\cdot{}_iG_{j-1} + ({}_{<i}\lambda+\mu_j)\cdot{}_iG_j - {}_{<i}\lambda\cdot{}_iG_{j+1} = 1 \quad (46)$$
$$- \mu_{n+s-1}\cdot{}_iG_{n+s-1} + ({}_{<i}\lambda+\mu_{n+s})\cdot{}_iG_{n+s} = 1$$

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 general resulting 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 non-conditional traffic-characteristics by substituting the actual "right hand side" C_j.

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 equation (95). If we notice equation (26) for the traffic ${}_iX$ carried by the system, we obtain the relation:

$${}_iX = {}_i\lambda \cdot {}_iG \qquad (47)$$

Equation (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. Equation (47) proves that it holds true also in the preemptive case where displacements and interruptions occur.

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

$${}_{\leq r}X = \sum_{i=1}^{r} {}_i\lambda \cdot {}_iG = {}_{\leq r}\lambda \cdot {}_{\leq r}G \qquad (48)$$

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:

$$u_{n+s,o} = {}_{<i}\lambda \qquad (49)$$

All other jump-rates $u_{j,o}$ 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 equation (93):

$${}_iV_j = \frac{{}_{<i}\Phi_{j-1}}{{}_{<i}\Phi_{n+s}} \qquad (50)$$

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

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

Using equation (19) we can write:

$${}_i\lambda \cdot {}_iV = {}_{\leq i}\lambda \cdot \{{}_{\leq i}P_{n+s} - {}_{<i}P_{n+s}\} \qquad (52)$$

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 $\le i$-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:

$$_iB = {_iC} + {_iV} \qquad (53)$$

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

$$_{\le i}B = \sum_{\nu=1}^{i} {_\nu\lambda} \cdot {_\nu B} = {_iC} \qquad (54)$$

and in particular:

$$_{\le r}B = {_rC} = B \qquad (55)$$

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

0,9 PROBABILITY

n = 3
s = 5
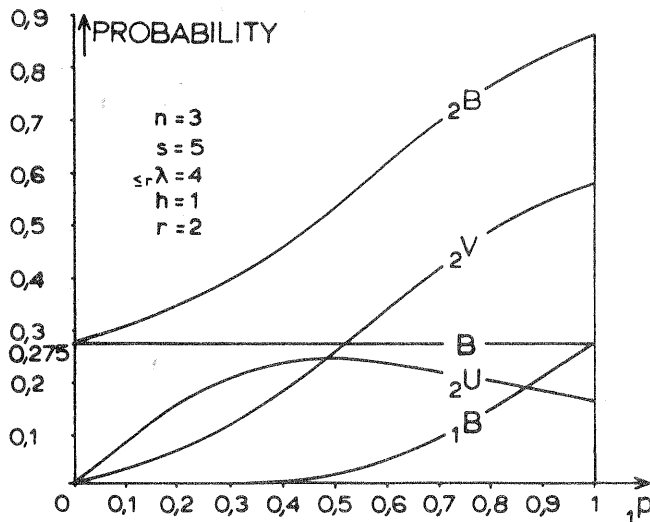$_{\le r}\lambda = 4$
h = 1
r = 2

FIGURE 1: Total probability of loss B, loss probability $_1B$ and $_2B$ of class 1 and 2, respectively, displacement probability $_2V$ and interruption probability $_2U$ of class 2 as functions of the share $_1p$ of class 1.

For a loss-delay-system with r=2 priority classes figure 1 shows some traffic-characteristics as a function of the share $_1p$ of class 1. According to equation (55), the overall probability of loss $_{\le 2}B$ is equal to the probability B of the priorityless system. The loss-probability $_1B$ of class 1 is less than B and $_2B$ is greater than B for all values of $_1p$. For $_1p=0$ no calls of class 1 are offered so that the traffic-characteristics of class 2 equal those of the priorityless system. For $_1p=0$, the displacement probability $_2V$ and the interruption probability $_2U$ of class 2 (which will be calculated later on) are equal to 0.

Together with $_2V$, the probability $_2B$ increases as $_{\le i}\lambda$ increases. So does the loss-probability $_1B$ until it reaches (for $_1p=1$) the value B. The interruption probability $_2U$ decreases after having reached a maximum, since for greater values of $_1p$, class 1 displaces the calls of class 2 out of the servers so that fewer interruptions occur.

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

$$_{\le i}Y = {_iA} \cdot (1 - {_{\le i}B}) \qquad (56)$$
$$_iY = {_iA} \cdot (1 - {_iB}) \qquad (57)$$

This relation (57) is not so obvious as the corresponding one in the system without prio-

rities, 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 follows:

$$_iY = {_iP} \cdot {_{\le r}Y} + {_iA} \cdot ({_{\le r}B} - {_iB}) \qquad (58)$$

$_iP \cdot {_{\le r}Y}$ 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 $B = {_{\le r}B}$ of the priorityless loss-delay-system. To clarify the meaning of the different total-times which the calls spent 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 our 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 "successful" 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 correspondencies:

$_iC = z_0/z$ $_iV = z_1/z$ $_iB = (z_0 + z_1)$
$_iG = x/z$ $_iJ = x_1/z$ $_iL = x_1/z_1$ $_iD = x_2/z_2$

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

The mean time $_iD$ (or $_iL$, respectively), which an unsuccessful (or successful resp.) call stays in the system cannot be calculated by means of the total-time $_iG$ of an arbitrary i-call. In order to get $_iD$ (the mean total-time of an uncuccessful 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 following initial values:

$$C_j = {_iV_j} \qquad (60)$$

The mean total-time $_iJ$ of an unsuccessful i-call related to all i-calls results from equation (95) with the "right hand sides":

$$_iJ_j(x=0) = {_iV_j} \qquad (59)$$

We get:

$$_i\lambda \cdot {_iJ} = \sum_{j=1}^{n+s} {_iX_j} \cdot {_iV_j} \qquad (61)$$

In equation (61) we use the term $_iX_j$, which has been introduced in equation (32) as the traffic of class i carried by place j. The term $_iX_j \cdot {_iV_j}$ 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 equation (61) as the unsuccessful share of traffic carried by the system. This unsuccessful traffic is not equal to $_iV \cdot {_iX}$. This is obvious, since the displacement-probabilities of the various places have not the same quantity. The greater j, the greater is the displacement probability $_iV_j$ of place j. We get the mean total-time $_iD$, which a displaced call stays in the system, by dividing $_iJ$ by $_iV$:

$$_iD = \frac{_iJ}{_iV} \qquad (62)$$

Correspondingly, the mean total-time $_iL$, which a successful call stays in the system, results in:

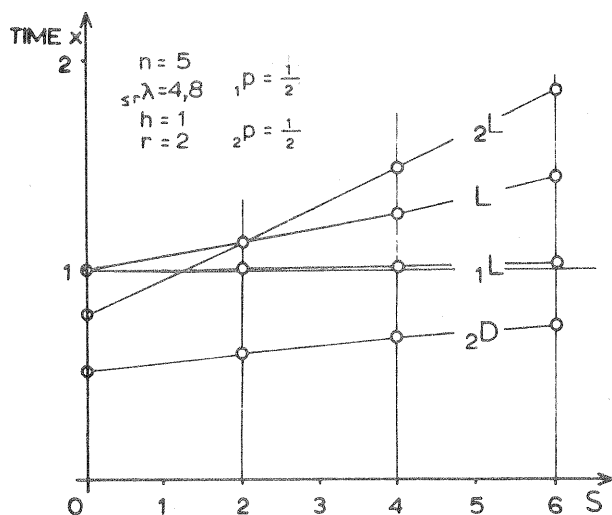$$_iL = \frac{_iG - _iJ}{_iS} \qquad (63)$$

FIGURE 2: The total-time L of a successful call in the priorityless system, the total-time $_1L$ 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 figure 2 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.

IV   WAITING-CHARACTERISTICS

In this chapter those traffic-characteristics are dealt with, 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. The second and further sections of the waiting-time are distributed as the first waiting-time of those calls, which arrive at place n+1.

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 explicite 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 (95) the following formulae can be derived:

$$_i\lambda \cdot {}_iT = {}_{\leq i}Q - \frac{_iE}{_{i-1}E} \cdot {}_{\leq i}Q \qquad (64)$$

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

$$_iE = {}_iC + {}_iR = \frac{_{\leq i}\Psi_s}{_{\leq i}\Phi_{n+s}} \qquad (65)$$

Please notice that we do not have an analogous result, as in the priorityless system ($\lambda \cdot T = Q$). 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 formulae:

$$_1\lambda \cdot {}_1T = {}_1Q \qquad (66)$$

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

$$_iT_{n+1} = \frac{1}{_{\leq i}\lambda} \cdot \frac{_{\leq i}\Phi_{n+s}}{_{\leq i}\Psi_s} \cdot {}_{\leq i}Q \qquad (67)$$

$_iT_{n+1}$ 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:

$$_iK_{n+j} = \frac{_{<i}\Psi_{k-1}}{_{<i}\Psi_s} \qquad (68)$$

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

$$_iK = \frac{_{\leq i}p}{_ip} \cdot \left\{ \frac{1}{_{\leq i}\Phi_{n+1}} - \frac{_iE}{_{i-1}E} \cdot \frac{1}{_{<i}\Phi_{n+s}} \right\} \qquad (69)$$

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

VI   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 equation (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.

$$(\mu_1 + {}_i\lambda) \cdot {}_iU_1 - {}_{<i}\lambda \cdot {}_iU_2 = 0$$
$$-\mu_{j-1} \cdot {}_iU_{j-1} + (\mu_j + {}_i\lambda) \cdot {}_iU_j - {}_{<i}\lambda \cdot {}_iU_{j+1} = 0 \quad j \neq n$$
$$-\mu_{n-1} \cdot {}_iU_{n-1} + (\mu_n + {}_{<i}\lambda) \cdot {}_iU_n = {}_{<i}\lambda \qquad (70)$$
$$-\mu_{n+s-1} \cdot {}_iU_{n+s-1} + (\mu_n + {}_{<i}\lambda) \cdot {}_iU_{n+s} = 0$$

In the appendix the solution of this system is also given for an arbitrary "right hand side" C_j. With the "right hand side" of equation (70) the general solution (97) of the appendix results in:

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

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

The last formulae 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 equation (68). Now, the i-call starts service in place n. From here the probability of interruption is $_iU_n$. It follows from equation (72) that both these events are independent. Before using this fact, we calculate the unconditional probability $_iU$ that an i-call is interrupted:

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

The sum can be evaluated using the equation (95) in the appendix:

$$\sum_{j=1}^{n} {}_iF_j \cdot {}_iU_j = \frac{<i\lambda}{i\lambda} \cdot \frac{n^s}{\leq_i A^s} \cdot \frac{1}{\leq_i \Phi_{n+s}} \cdot \left\{ 1 - \frac{\leq_i \Phi_n}{<i\Phi_n} \right\} \qquad (74)$$
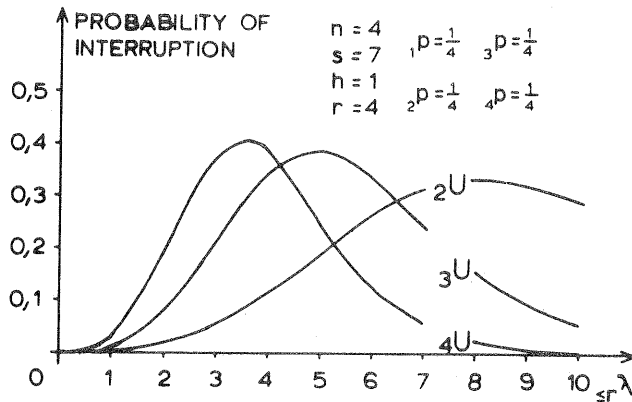
FIGURE 3: Interruption probabilities ${}_iU$ as a function of the total arrival rate $\leq_r\lambda$.

For the loss-delay-system as in figure 3 the probability of interruption ${}_iU$ for class i=2,3,4 is shown as a function of the total arrival rate $\leq_r\lambda$. For small $\leq_r\lambda$ the probabilities ${}_iU$ increase with increasing $\leq_r\lambda$. Then after having reached a maximum, they decrease and tend towards 0, since with increasing $\leq_i\lambda$, 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:

$${}_iW = {}_iR + \sum_{j=1}^{n} {}_iF_j \cdot {}_iU_j \qquad (75)$$

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

$${}_i^{\nu}U = {}_iU \cdot {}_iU_{n+1}^{\nu-1} \cdot (1 - {}_iU_{n+1}) \qquad (76)$$

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

$${}_iN = \sum_{\nu=1}^{\infty} \nu \cdot {}_i^{\nu}U = \frac{{}_iU}{1 - {}_iU_{n+1}} \qquad (77)$$

The mean number of interruptions of an arbitrary call amounts to:

$$\leq_r N = \sum_{i=1}^{r} {}_iP \cdot {}_iN \qquad (78)$$

In chapter V we calculated in equation (64) the mean first waiting-time ${}_iT$ and in equation (67) the mean waiting-time in place n+1, which is valid for interrupted calls. Noticing that the mean number ${}_iN$ of interruptions for an arbitrary i-call is given by (75), we are able to calculate the mean total waiting-time ${}_iM$ of an i-call:

$${}_iM = {}_iT + {}_iN \cdot {}_iT_{n+1} \qquad (79)$$

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

$$Z = \frac{\leq_i X}{\leq_i G} \cdot \leq_r N \qquad (80)$$

Substituting equation (78) and (48) into (80) we get the following result for the mean number of interruptions per unit of time:
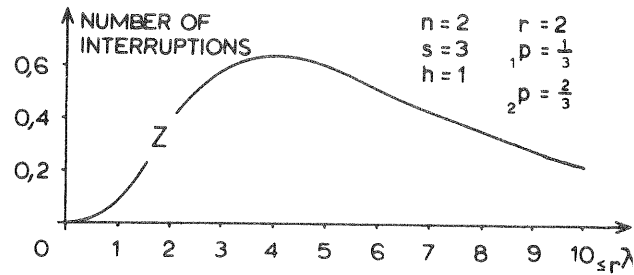
$$Z = \leq_r \lambda \cdot \leq_r N \qquad (81)$$

FIGURE 4: The mean number of interruptions per time-unit as a function of the total arrival rate $\leq_r\lambda$.

Figure 4 shows the mean number Z of interruptions per time unit as a function of the total arrival rate $\leq_r\lambda$. For small $\leq_r\lambda$ only few interruptions occur. As $\leq_r\lambda$ 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.

VII CONCLUSION

The traffic-characteristics of the preemptive loss-delay-system are calculated exactly in this paper for an infinite number of POISSON-sources.

These characteristics have been classified into the following four groups:

- State-probability characteristics (Chapter II)

- Displacement characteristics (Chapter IV)

- Waiting characteristics (Chapter V)

- Interruption characteristics (Chapter VI)

The derivations presented are shown in more detail in /1/, for an infinite as well as a finite number of POISSON-sources.

In order to facilitate practical applications ALGOL-programs have been written for the numerical evaluation. Besides the analysis of traffic-characteristics for given systems and fixed traffic values, these calculation methods enable the investigation of the trends and of the interdependences of the various traffic-characteristics. This may be important for the synthesis of systems. E.g. a system can be planned such that only a suitable number of interruptions per time-unit occurs (confer figure 4).

APPENDIX

In the appendix the solution 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 $\Phi_{\nu,\kappa}$. For this basic function the definition and some calculating rules are stated, at first.

We will define the basic function $\Phi_{\nu,\kappa}$ as a function of the traffic offered A and of the traffic parameter β_ρ, respectively. According to equation (14) we have:

$$\beta_\rho = \frac{\mu_\rho}{\lambda} \qquad (82)$$

The basic function $\Phi_{\nu,\kappa}$ is given by the following definition:

$$\Phi_{\nu,\kappa} = \sum_{\sigma=\nu}^{\kappa} \prod_{\rho=\sigma+1}^{\kappa} \beta_\rho \qquad (83)$$

The basic function $\leq_i\Phi_{\nu,\kappa}$ results by substituting A by $\leq_i$A (and as a consequence β_ρ by $\leq_i\beta_\rho$). If $\nu=0$, then we use the abbreviation of equation (16):

$$\Phi_{0,\kappa} = \Phi_\kappa \qquad (84)$$

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

$$\nu \geq n: \qquad \Psi_\rho = \Phi_{\nu,\nu+\rho} = \sum_{\sigma=0}^{\rho} \beta_n^\sigma \qquad (85)$$

The basic function satisfies the following relation:

$$\Phi_{\nu,\kappa} = \Phi_{\sigma,\kappa} + \Phi_{\nu,\sigma-1} \prod_{\rho=\sigma}^{\kappa} \beta_\rho \qquad (86)$$

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

$$\Phi_{\nu,\kappa} = \Phi_{\nu+1,\kappa} + \prod_{\rho=\nu+1}^{\kappa} \beta_\rho \qquad (87)$$

$$\Phi_{\nu,\kappa} = 1 + \beta_\kappa \cdot \Phi_{\nu,\kappa-1} \qquad (88)$$

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

$$\kappa \leq n: \qquad \Phi_{\nu,\kappa} = \frac{\kappa!}{A^\kappa} \sum_{\sigma=\nu}^{\kappa} \frac{A^\sigma}{\sigma!} \qquad (89)$$

$$\nu \geq n: \qquad \Phi_{\nu,\kappa} = \Psi_{\kappa-\nu} = \begin{cases} \dfrac{A}{n-A} \cdot (\dfrac{n^{\kappa-\nu+1}}{A^{\kappa-\nu+1}} - 1) & A \neq n \\ \\ \kappa-\nu+1 & A=n \end{cases} \qquad (90)$$

$$\nu < n < \kappa: \qquad \Phi_{\nu,\kappa} = \Psi_{\kappa-n-1} + (\frac{n}{A})^{\kappa-n} \cdot \Phi_{\nu,n} \qquad (91)$$

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

The first system of equations, whose solution shall be given in this appendix, reads as follows (cf. equation 46):

$$(\mu_1+_i\lambda) \cdot_i X_1 - _i\lambda \cdot_i X_2 = C_1$$
$$- \mu_{k-1} \cdot_i X_{k-1} + (\mu_k+_i\lambda) \cdot_i X_k - _i\lambda \cdot_i X_{k+1} = C_k \qquad (92)$$
$$- \mu_{n+s-1} \cdot_i X_{n+s-1} + (\mu_{n+s}+_i\lambda) \cdot_i X_{n+s} = C_{n+s}$$

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

$$_iX_k = \frac{1}{_i\lambda \cdot_i\Phi_{n+s}} \cdot \{_i\Phi_{k,n+s} \cdot \sum_{\nu=1}^{k-1} C_\nu \cdot_i\Phi_{\nu-1} \cdot \prod_{\rho=\nu}^{k-1} {}_i\beta_\rho + {}_i\Phi_{k-1} \cdot \sum_{\nu=k}^{n+s} C_\nu \cdot_i\Phi_{\nu,n+s}\} \qquad (93)$$

The index k of $_iX_k$ represents the starting place, to which the characteristic $_iX_k$ is related. We get the corresponding characteristic $_iX$, which is averaged among all starting places by means of the probability $_iF_k$, that an i-call starts in place j:

$$_iX = \sum_{k=1}^{n+s} {}_iF_k \cdot {}_iX_k \qquad (94)$$

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

$$_iX = \sum_{\nu=1}^{n+s} C_\nu \cdot \{\frac{\leq_i\Phi_{\nu,n+s}}{\leq_i\Phi_{n+s}} - \frac{<_i\Phi_{\nu,n+s}}{<_i\Phi_{n+s}}\} \qquad (95)$$

In order to calculate the interuption probabilitiy, we have to solve the following system of linear equations (cf. equation 70):

$$(\mu_1+_i\lambda) \cdot_i U_1 - _i\lambda \cdot_i U_2 = C_1$$
$$-\mu_{n-1} \cdot_i U_{n-1} + (\mu_n+_i\lambda) \cdot_i U_n = C_n \qquad (96)$$
$$-\mu_{k-1} \cdot_i U_{k-1} + (\mu_k+_i\lambda) \cdot_i U_k - _i\lambda \cdot_i U_{k+1} = C_k$$
$$-\mu_{n+s-1} \cdot_i U_{n+s-1} + (\mu_n+_i\lambda) \cdot_i U_{n+s} = C_{n+s}$$

The solution of this system is given by:

$$_iX_k = \frac{1}{_i\lambda \cdot_i\Phi_n} \cdot \{_i\Phi_{k,n} \cdot \sum_{\nu=1}^{k-1} C_\nu \cdot_i\Phi_{\nu-1} \cdot \prod_{\rho=\nu}^{k-1} {}_i\beta_\rho + {}_i\Phi_{k-1} \cdot \sum_{\nu=k}^{n} C_\nu \cdot_i\Phi_{\nu,n} \qquad k=1,2,\ldots,n \qquad (97)$$

$$_iX_k = \frac{<_i\beta_n \cdot <_i\Psi_{s-k}}{\Psi_s} \cdot \frac{1}{_i\lambda \cdot_i\Phi_n} \cdot \sum_{\kappa=1}^{n} C_\kappa \cdot_i\Phi_{\kappa-1} \cdot \prod_{\rho=\kappa}^{n-1} {}_i\beta_\rho + \qquad k=1,2,\ldots,s$$

$$+ \frac{1}{_i\lambda \cdot_i\Psi_s} \cdot \{_i\Psi_{s-k} \cdot \sum_{\kappa=1}^{n-1} C_{n+\kappa} <_i\beta_n \cdot <_i\Psi_{\kappa-1} + {}_i\Psi_{k-1} \cdot \sum_{\kappa=k}^{s} C_{n+\kappa} <_i\Psi_{s-\kappa}\}$$

Substituting "actual" right hand sides C_j into (93), (95) and (97) resp., we obtain the various traffic characteristics, which were presented in this paper.

REFERENCES

/1/ G.J.BRANDT   Das preemptive Warte-Verlust-System
                  Ph.D.Thesis,1970,University of Stuttgart

/2/ W.FELLER   An Introduction to Probability Theory and its Application
                  John Wiley,New York

/3/ N.U.PRABHU   Stochastic Processes
                  The Macmillan Comp.,New York

/4/ J.RIORDAN   Stochastic Service Systems
                  John Wiley,New York

/5/ W.WAGNER   Ober ein kombiniertes Warte-Verlustsystem mit Prioritäten
                  Ph.D.Thesis,1968,University of Stuttgart