# TWO-STAGE QUEUING SYSTEM WITH SAMPLED PARALLEL INPUT QUEUES

M. Langenbach-Belz
University of Stuttgart
Stuttgart, Federal Republic of Germany

ABSTRACT

A two-stage queuing model is dealt with, which corresponds to a simplified model of a modern central controlled telephone or data switching system. The primary system consists of g parallel storages (input queues) corresponding to different peripheral devices. The secondary system consists of a buffer and a service unit corresponding to the central processor unit. The input process of the primary system is a Poisson process. The holding time of the service unit in the secondary system is constant. Always at fixed, equidistantly distributed instants the input queues are inspected simultaneously for waiting informations. At those inspection instants from each input queue i there are transferred at most $n_i$ informations to the secondary system.
The analysis of this model is based on two separate single-stage models, which were earlier dealt with by the author in /1/. However, the two-stage model treated in this paper takes into consideration the dependency of the state of the secondary system on the state of the primary system.

In this paper, first the two single-stage models mentioned above and their main results are briefly described. Then, the two-stage model is studied in detail.
For the primary system exact solutions for probabilities of state, mean waiting time, waiting time distribution function etc. are available.

For the secondary system exact solutions for certain probabilities of state, mean queue length, mean waiting time and probability of waiting are derived. For certain parameter combinations an approximate solution for the mean waiting time in the secondary system is given.

Furthermore, another operating mode of the two-stage model is considered. For the case, that the input queues are inspected not simultaneously but cyclically, an approximation is given for the mean waiting time in the secondary system.
Finally, several numerical results are presented.

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

## 1. INTRODUCTION

Since the last few years digital computers are used more and more as central control in modern telephone and data switching systems, in communication networks etc. The structure of such central controlled systems can be subdivided into two main functional parts, namely the periphery and the central unit. The informations, generated in the peripheral devices, must be transferred to the central unit, where they are processed. On the other hand, instructions generated in the central unit must be transferred to the peripheral devices. In this paper, the transfer of informations from the periphery to the central unit is considered.
In principle, two different modes are possible for this transfer:
Firstly, the informations can be transferred to the central unit directly after their generation in the periphery. But each input/output operation requires a certain amount of organization time in the central unit (overhead) and the currently processed program must be interrupted. Therefore, the disadvantage of this method is, that each randomly generated information causes an interrupt in the central processor unit. This would result in a relatively large amount of overhead and so the waiting times of the informations would be increased in particular for a large offered traffic.
Secondly, the informations may be buffered in the periphery after their generation. The central unit may inspect the periphery for waiting informations in adequate time intervals. This method is mostly used in modern central controlled telephone or data switching systems. The time interval between two "inspection instants" is constant in most of the realized systems. Thus, the information transfer between the periphery and the central processor unit (CPU) takes place in fixed intervals of time. The basic structure of such a "sampled queuing system" is shown in Fig.1. The clocked information transfer between the periphery and the CPU is indicated by a sampling switch (gate).

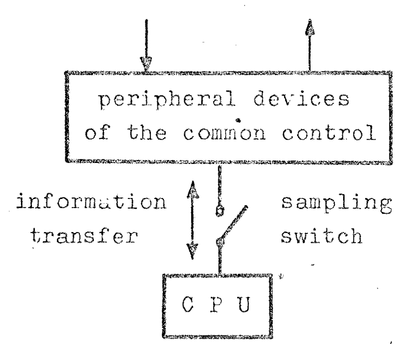


Fig.1: Basic structure of a sampled queuing system

In this paper in particular the waiting times of the informations in systems according to Fig.1

are analyzed by the aid of simplified mathematical models. For this analysis, two single-stage models are used as a basis, which were already dealt with by the author in /1/. One of these models corresponds to a peripheral device, the other one corresponds to the CPU. In this paper, however, a two-stage model is dealt with, which corresponds to the general system of Fig.1 and which takes into account the dependency of the state of the CPU on the periphery.

## 2. DESCRIPTION OF MODELS

### 2.1. SINGLE-STAGE MODELS FOR SUB-SYSTEMS

Because the two single-stage models mentioned above are used as a basis for the investigations in the further sections, their structure and their main results are briefly described in the following. Their detailed analysis and further results can be found in /1/.

#### 2.1.1. MODEL CORRESPONDING TO A PERIPHERAL DEVICE (MODEL A)

The configuration of a model corresponding to a peripheral device is shown in Fig.2 (Model A).

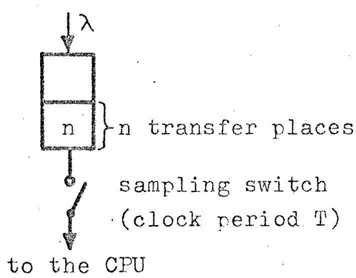


Fig.2: Configuration of Model A

The calls arrive according to a Poisson-process with the arrival rate $\lambda$. They have to wait in the queue in the order of their arrival (probability of waiting = 1). It is assumed, that there is an infinite number of waiting places. Always after a fixed interval of time T the sampling switch is closed and the waiting calls situated in the n "transfer places" in front of the queue are transferred e.g. to the buffer of a CPU. It is assumed, that this transfer cannot be blocked.

The principle way of solution for this model is the following (cf. /1/):
The state x of the system is considered always at fixed instants $t_v$ ($0 \leq t_v \leq T$) before the next sampling clock. The state x at these instants forms an imbedded Markov chain. Then, the generating function $Gx(z,t_v)$ of the probabilities $p(x,t_v)$ of state x at the instant $t_v$ before the next sampling clock is derived for the case of statistical equilibrium. The definition of the generating function $Gx(z,t_v)$ is

$$Gx(z,t_v) = \sum_{x=0}^{\infty} p(x,t_v)\, z^x \qquad (1)$$

The generating function $Gx(z,t_v)$ derived in /1/ is:

$$Gx(z,t_v) = \frac{n-\lambda T}{\prod_{\nu=1}^{n-1}(1-z_\nu)} \cdot \frac{\prod_{\nu=0}^{n-1}(z-z_\nu)}{z^n e^{\lambda T(1-z)}-1} \cdot e^{\lambda t_v(1-z)} \qquad (2)$$

where $z_\nu$ are the roots of $z^n e^{\lambda T(1-z)} - 1 = 0$
<u>inside and on</u> the unit circle
(note: $z_0 = 1$ is always a root)

The mean queue length $E[x,t_v]$ at an arbitrary instant $t_v$ before the next sampling clock is the first derivative of $Gx(z,t_v)$ with respect to z at the point z=1. It is (cf. /1/):

$$E[x,t_v] = \sum_{\nu=1}^{n-1} \frac{1}{1-z_\nu} + \frac{(n-\lambda T)^2 - n}{2(\lambda T - n)} - \lambda t_v \qquad (3)$$

By the aid of the mean queue length $E[x,t_v]$ the mean waiting time w can be obtained. It is

$$w = \frac{1}{\lambda}\left[\sum_{\nu=1}^{n-1} \frac{1}{1-z_\nu} + \frac{n}{2}\left(\frac{1}{n-\lambda T} - 1\right)\right] \qquad (4)$$

The probabilities of state $p(x,t_v)$ at an arbitrary instant $t_v$ before the next sampling clock are achieved by a serial expansion of the generating function $Gx(z,t_v)$ of equ.(2) and a coefficient comparison with its definition in equ.(1). The result is according to /1/:

$$p(x,t_v) = \frac{\lambda T - n}{\prod_{\nu=1}^{n-1}(1-z_\nu)} \cdot e^{\lambda t_v} \left[ \sum_{j=0}^{b-1} \left( e^{j\lambda T} \sum_{\mu=0}^{n} (-1)^{n-\mu} S_{n-\mu} \frac{[-\lambda(jT+t_v)]^{x-jn-\mu}}{(x-jn-\mu)!} \right) + e^{b\lambda T} \sum_{\mu=0}^{x-bn} (-1)^{n-\mu} S_{n-\mu} \frac{[-\lambda(bT+t_v)]^{x-bn-\mu}}{(x-bn-\mu)!} \right] \qquad (5A)$$

for $bn \leq x < (b+1)n$, $b = 0,1,2,..$
and with

$$\left.\begin{aligned} S_0 &= 1 \\ S_1 &= (z_0+z_1+z_2+\ldots+z_{n-1}) = \sum_{i=0}^{n-1} z_i \\ S_2 &= (z_0z_1+z_0z_2+\ldots+z_{n-2}z_{n-1}) = \sum_{\substack{i1,i2=0 \\ (i1<i2)}}^{n-1} z_{i1}z_{i2} \\ &\vdots \\ S_n &= z_0z_1z_2\ldots z_{n-1} \end{aligned}\right\} \qquad (5B)$$

and additionally $\sum_{i}^{j}(\ldots) = 0$ for $j<i$ (5C)

In /1/, however, further results were derived, e.g. an explicit formula for the waiting time distribution function, but the above formulae are sufficient for the next sections.

#### 2.1.2. MODEL CORRESPONDING TO THE CENTRAL PROCESSING UNIT (MODEL B)

The configuration of a model corresponding to the central processing unit is shown in Fig.3 (Model B).

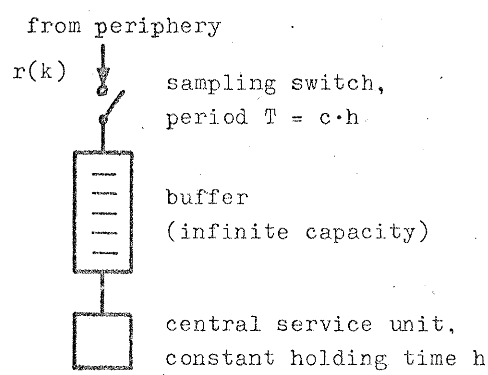


Fig.3: Configuration of Model B

The central service unit has a constant service time h. Always after a fixed period of c service times (c=integral number $\geq 1$) the sampling switch is closed and a group of $k=0,1,2,\ldots,m$ calls arrives from the periphery at the buffer (clocked group arrivals). The probability of k arriving calls per sampling clock be r(k). It is assumed in this Model B, that the probabilities r(k) are independent in successive instants of the sampling clock. The calls waiting in the buffer are served serially from the central service unit with discipline first come, first served.

The principle way of solution for this system is similar as for Model A in section 2.1.1. From the variety of results only the following two are necessary for this paper.

The mean queue length $E[x,0]$ (including a call in the service unit) directly before the sampling clock ($t_v=0$) is according to /1/:

$$E[x,0] = \sum_{\mu=c}^{m-1} \frac{1}{z_\mu - 1} \qquad (6)$$

where $z_\mu$ are the roots of $z^c - \sum_{k=0}^{m} r(k) z^k = 0$ outside the unit circle.

The mean waiting time w of all calls then is (cf. /1/):

$$w = E[x,0]\cdot h + \frac{1}{2}\left(\frac{Var[k]}{E[k]} + E[k] - 1\right)\cdot h \qquad (7)$$

where $E[k] = \sum_{k=0}^{m} k\cdot r(k)$ = mean of k (8A)

and $Var[k] = \sum_{k=0}^{m} (k-E[k])^2 r(k)$ = variance of k, (8B)

In equ.(7), $E[x,0]\cdot h$ is the mean waiting time of the first call of an arriving group, whereas the second term results from the additional waiting time of the other calls of the group.

### 2.2. MODEL OF A TWO-STAGE QUEUING SYSTEM WITH SAMPLED PARALLEL INPUT QUEUES (MODEL C)

The model described in this section is a combination of Model A and Model B of section 2.1. It is named Model C. Model C corresponds to a system with the basic structure of Fig.1 and it is a simplified model of e.g. a common controlled telephone or data switching system. The configuration of Model C is shown in Fig.4.

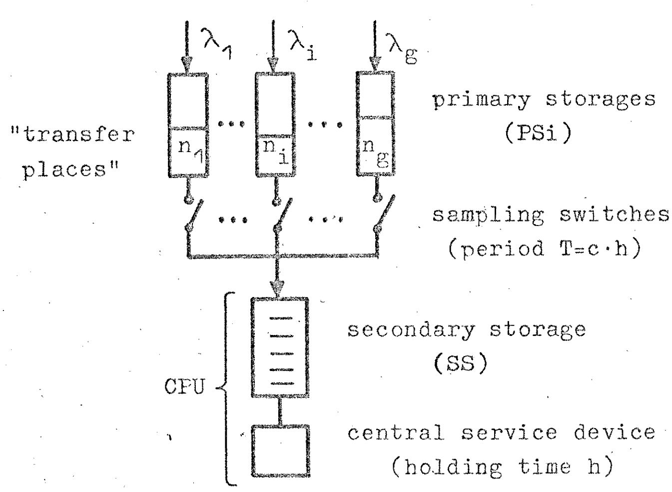


Fig.4: Configuration of Model C (all storages are assumed to have infinite capacity)

The primary storages (PSi, i=1,2,...,g) correspond to different peripheral devices. The secondary storage (SS) and the central service device correspond to the central processor unit (CPU). The primary storages altogether are named primary system, the secondary storage and the central service unit form the secondary system. Primary and secondary system form together the general system.
The operating mode of this system is the following:
The calls arrive according to a Poisson-process with the arrival rate $\lambda_i$ at the PSi and they have to wait in the queue of PSi. Always after a fixed time T all sampling switches are closed simultaneously. When the sampling switches are closed, the waiting calls situated in the $n_i$ transfer places of each PSi are transferred to the SS. It is assumed, that the calls of PS2 are filed into the SS behind the calls of PS1 etc. without delay during one sampling instant. The calls waiting in the SS are served serially from the central service device with discipline first come, first served. The central service device has a constant service time h. The sampling period T is an integral multiple c ($c \geq 1$) of the holding time h.
The following sections of this paper are dealing with this Model C.

## 3. ANALYTICAL TREATMENT OF MODEL C

### 3.1. EXACT SOLUTION FOR THE PRIMARY SYSTEM

With each sampling clock there are removed at most $n_i$ calls from the primary storage PSi and are transferred to the secondary storage SS (see Fig.4). Because the SS is assumed to have an infinite number of waiting places, there exists never a blocking of this transfer from the PSi to the SS. Therefore, the state of any PSi is not influenced at all by the state of all other primary storages or the secondary system. Hence, the structure and the input/output process of a single PSi is identical with the configuration of Model A of section 2.1.1. Thus, all characteristic traffic values of interest may be obtained from Model A ( cf. section 2.1.1. and /1/).

### 3.2. EXACT SOLUTION FOR THE SECONDARY SYSTEM

The structure of the secondary system is identical with the structure of Model B (see Fig.4 and Fig.3). But there is quite a difference between the arrival process in the SS and the arrival process in Model B. In Model B it was assumed, that the probabilities r(k) for an arrival of a group with k calls per sampling clock are independent in successive sampling instants. This condition is not fulfilled in the secondary system for the groups of calls coming from the primary system with each sampling clock. Let us see this fact by an example:
The primary system shall consist of only one PS with n transfer places. If directly before the sampling clock there are less than n calls in the PS, then all calls are transferred with the sampling clock to the SS and the PS is empty thereafter. Therefore, the size of the group transferred with the successive sampling clock depends only on the Poisson arrival process of the PS. If exactly n calls are transferred with a sampling clock, then more than 0 calls may be left in the PS. Therefore, the size of the group transferred with the successive sampling clock depends not only on the Poisson arrival process of the PS, but also on the state of the PS directly after the sampling clock. Thus, the probability of the transferred group size depends on the group size transferred with the previous sampling clock.

Normally, multi-dimensional probabilities of state must be used for the solution of Model C. With the assumption, however, that each number of transfer places $n_i$ is greater or equal to c, important characteristic traffic values can be obtained by only one-dimensional probabilities of state. The analysis of such systems with $c \leq Min[n_i]$ (where $Min[n_i]$ is the smallest of all $n_i$, i=1,2,...,g) is given in sections 3.2.1. through 3.2.3.

#### 3.2.1. PROBABILITIES OF STATE

To find a solution for the secondary system, first the state of the general system (primary system + secondary system) is considered always directly after the sampling clock. For reasons of a uniform nomenclature one can say equivalently, that the general system is considered at the time T before the next sampling clock.

The probability of $x_{G,i}$ or $x_{G,i+1}$ calls in the general system at time T before the next sampling clock i or i+1 shall be $p_{G,i}(x,T)$ or $p_{G,i+1}(x,T)$, respectively ( cf. Fig.5).

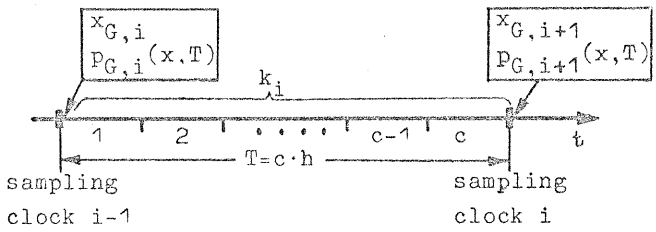


Fig.5: Imbedded Markov chain

The state $x_G$ at these instants forms an imbedded Markov chain. The number of calls arriving at the primary system during the time T shall be $k_i$. These $k_i$ arriving calls have to wait in the primary system and they cannot be transferred to the secondary system before clock i. Hence, only those calls can be served by the service unit during the time between clock i-1 and i, which were already present in the secondary system directly after clock i-1.
With each sampling clock, $n_i \geq c$ calls can be transferred at most from PSi to the SS. If directly after clock i-1 there are $x_{G,i} \leq \mathrm{Min}[n_i]$ calls in the general system, then all $x_{G,i}$ calls must be situated in the secondary system and the primary system must be empty. If directly after clock i-1 there are $x_{G,i} > \mathrm{Min}[n_i]$ calls in the general system, then at least $\mathrm{Min}[n_i]$ calls must be situated in the secondary system.
The considerations above lead to the following facts:
If directly after clock i-1 there are $x_{G,i} \leq c$ ($c \leq \mathrm{Min}[n_i]$) calls in the general system, then all $x_{G,i}$ calls are served until clock i. If directly after clock i-1 there are $x_{G,i} > c$ calls in the general system, then exactly c calls are served until clock i. Therefore, the following equation for the state of the <u>general system</u> directly <u>after</u> the sampling clock holds true:

$$x_{G,i+1} = \mathrm{Max}\left[(x_{G,i}-c),0\right] + k_i \tag{9}$$

with

$$\mathrm{Max}\left[(x_{G,i}-c),0\right] = \begin{cases} x_{G,i}-c & \text{for } x_{G,i} \geq c \\ 0 & \text{for } x_{G,i} < c \end{cases}$$

Equ.(9) is completely analog to an equation, which was obtained in /1/ for the state of the <u>Model A</u> directly <u>before</u> the sampling clock ($t_v=0$). In this equation of Model A, only the number of transfer places n must be replaced by c to get equ.(9). In equ.(9) the number $k_i$ of calls arriving at the primary system during the time T depends on the total Poisson arrival process with the arrival rate

$$\Lambda = \sum_{i=1}^{g} \lambda_i \tag{10}$$

Because of the analogy with Model A mentioned above, the derivation of the generating function $Gx_G(z,T)$ of the probabilities of state $p_G(x,T)$ directly after the sampling clock in the case of statistical equilibrium can be performed completely analog to Model A. Thus, analogously to equ.(2), one gets for the generating function $Gx_G(z,T)$ of the probabilities $p_G(x,T)$ in the general system directly after the sampling clock

$$Gx_G(z,T) = \frac{c-\Lambda T}{\prod_{\nu=1}^{c-1}(1-z_\nu)} \cdot \frac{\prod_{\nu=0}^{c-1}(z-z_\nu)}{z^c - e^{-\Lambda T(1-z)}}\, e^{-\Lambda T(1-z)} \tag{11}$$

where $z_\nu$ are the roots of $z^c - e^{-\Lambda T(1-z)} = 0$ <u>inside and on</u> the unit circle (with $z_0=1$)
(In equ.(2): $t_v=0$, $n \neq c$, $\lambda \neq \Lambda$ leads to equ.(11))

Analogously to equ.(3), the expectation of the number of calls in the general system directly after the sampling clock then is

$$E_G[x,T] = \sum_{\nu=1}^{c-1} \frac{1}{1-z_\nu} + \frac{(c-\Lambda T)^2 - c}{2(\Lambda T - c)} \tag{12}$$

Finally, the probabilities $p_G(x,T)$ of the general system directly after the sampling clock are in analogy to equ.(5A):

$$p_G(x,T) = \frac{\Lambda T - c}{\prod_{\nu=1}^{c-1}(1-z_\nu)} \left[ \sum_{j=0}^{b-1} \left( e^{j\Lambda T} \sum_{\mu=0}^{c} (-1)^{c-\mu} S_{c-\mu} \frac{(-\Lambda j T)^{x-jc-\mu}}{(x-jc-\mu)!} \right) + e^{b\Lambda T} \sum_{\mu=0}^{x-bc} (-1)^{c-\mu} S_{c-\mu} \frac{(-\Lambda b T)^{x-bc-\mu}}{(x-bc-\mu)!} \right] \tag{13}$$

with $bc \leq x < (b+1)c$ and additionally equs.(5B) and (5C)

As already mentioned above, the primary system must be empty, if directly after the sampling clock there are $x < \mathrm{Min}[n_i]$ calls in the secondary system. So in this case the general system contains also $x_G = x$ calls. Therefore, the probabilities of state $p_S(x,T)$ of the secondary system directly after the sampling clock are in the case of $x < \mathrm{Min}[n_i]$:

$$p_S(x,T) = p_G(x,T) \quad \text{for } x < \mathrm{Min}[n_i] \tag{14}$$

For the special case of $x < c$ it follows from equ.(14) and equ.(13) (with b=0):

$$p_S(x,T) = \frac{\Lambda T - c}{\prod_{\nu=1}^{c-1}(1-z_\nu)} (-1)^{c-x} S_{c-x} \quad \text{for } x < c \tag{15}$$

($S_i$ from equ.(5B))

### 3.2.2. MEAN QUEUE LENGTH AND MEAN WAITING TIME

The mean number of calls $E_G[x,T]$ in the general system directly after the sampling clock is given by equ.(12). On the other hand, the mean queue length $E_{Pi}[x,T]$ in a PSi directly after the sampling clock is available from equ.(3) with $t_v=T$. Then, the mean queue length $E_S[x,T]$ in the secondary system directly after the sampling clock is

$$E_S[x,T] = E_G[x,T] - \sum_{i=1}^{g} E_{Pi}[x,T] \tag{16}$$

The mean number $\Omega_S$ of calls <u>waiting</u> in the secondary storage, averaged over the whole sampling period T, is approached by the following considerations:
The state of the secondary system can change only at integral multiples of the holding time before or after the sampling clock. Therefore, the probabilities of state and the mean queue length are step functions of time between two sampling instants ( cf. Fig.6)

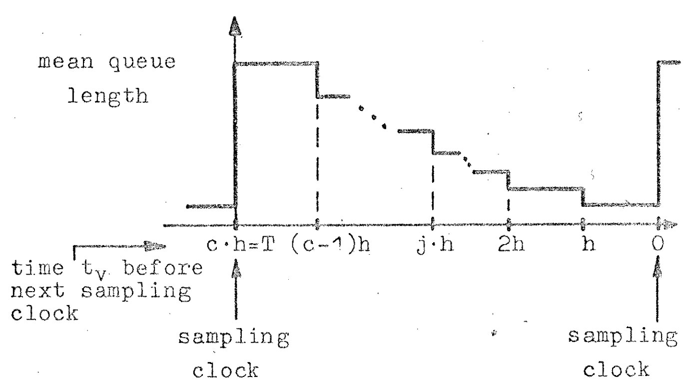


Fig.6: Mean queue length in the secondary system

So the secondary system is considered now at time instants $t_v = j \cdot h$ ($j=0,1,2,\ldots,c$) before the next sampling clock. Since the last sampling clock, c-j calls may be served at most until the instant $t_v = j \cdot h$ before the next sampling clock.

Then, for the probabilities of state $p_S(x,jh)$ of the secondary system at the time $t_v=jh$ before the next sampling clock the following equation holds true:

$$\left.\begin{aligned} p_S(0,jh) &= \sum_{x=0}^{c-j} p_S(x,T) \\ p_S(x,jh) &= p_S(x+c-j,T) \text{ for } x>0 \end{aligned}\right\} j=0,1,\ldots,c \qquad (17)$$

The probabilities on the right hand side of equ.(17) are known from equ.(15) (up to $x+c-j<\text{Min}[n_i]$).

If the secondary storage contains $x>0$ calls at an arbitrary instant, then the service unit is occupied and the secondary system contains $x+1$ calls. Therefore, the mean number of calls $\Omega_S(jh)$ waiting in the secondary storage (SS) at an instant $t_v=jh$ before the next sampling clock is

$$\Omega_S(jh) = \sum_{x=1}^{\infty} x p_S(x+1,jh) \qquad (18)$$

The mean number of calls $\Omega_S$ waiting in the SS, averaged over a whole sampling interval T is

$$\Omega_S = \frac{1}{c}\sum_{j=1}^{c} \Omega_S(jh)$$

or with equ.(18)

$$\Omega_S = \frac{1}{c}\sum_{j=1}^{c}\sum_{x=1}^{\infty} x p_S(x+1,jh) \qquad (19)$$

Substituting equ.(17) into equ.(19) leads after several transformations to

$$\Omega_S = E_S[x,T] - \frac{1}{c}\sum_{j=1}^{c}\left[j-\sum_{x=0}^{j-1}(j-x)p_S(x,T)\right] \qquad (20)$$

$E_S[x,T]$ is known from equ.(16), $p_S(x,T)$ is known from equ.(15).

With the general equation $\Omega_S=\Lambda\cdot w_S$ and with equ.(20) the mean waiting time $w_S$ of all calls in the secondary system is derived as

$$w_S = \frac{1}{\Lambda}\left[E_S[x,T] - \frac{1}{c}\sum_{j=1}^{c}\left(j-\sum_{x=0}^{j-1}(j-x)p_S(x,T)\right)\right] \qquad (21)$$

### 3.2.3. PROBABILITY OF WAITING

At first, the probability $W_S(=0)$ is determined, that a call arriving at the secondary system must not wait.

Only the first call of an arriving group must not wait, when the secondary system is empty directly before the sampling clock. The probability of an empty secondary system directly before the sampling clock is obtained from equ.(17) with j=0:

$$p_S(0,0) = \sum_{x=0}^{c} p_S(x,T) \qquad (22)$$

If the secondary system is empty directly before the sampling clock i, then directly after the previous sampling clock i-1 there were $\leq c$ calls in the secondary system. In this case, the whole primary system was empty directly after the previous sampling clock i-1 (cf. section 3.2.1). Then, the probability that a call is transferred at all from the primary to the secondary system with sampling clock i is because of the Poisson arrival process

$$1 - e^{-\Lambda T} \quad .$$

Hence, the probability that a group with at least one call arrives with the sampling clock and finds the secondary system empty, is

$$p_S(0,0)\cdot(1 - e^{-\Lambda T}) \quad .$$

Because in this case only the first call of the arriving group must not wait,

$$p_S(0,0)\cdot(1 - e^{-\Lambda T})\cdot 1$$

calls arrive in the mean per sampling clock at the secondary system, which do not have to wait. Altogether $\Lambda T$ calls arrive per sampling clock in the mean. Therefore, the probability $W_S(=0)$ is

$$W_S(=0) = \frac{p_S(0,0)\cdot(1-e^{-\Lambda T})}{\Lambda T} \qquad (23)$$

The probability of waiting $W_S(>0)$ then is

$$W_S(>0) = 1 - W_S(=0) \qquad (24)$$

Equs.(22) and (23) can be inserted into equ.(24).

## 3.3. APPROXIMATE SOLUTIONS FOR THE SECONDARY SYSTEM

### 3.3.1. APPROXIMATION OF THE MEAN WAITING TIME, WHEN THE NUMBER OF TRANSFER PLACES PER PS IS SMALLER THAN c.

If the number of transfer places $n_i$ per PS is smaller than c, the calculation method shown in section 3.2 for systems with $n_i\geq c$ is not applicable. The reason is explained by an example. The number of transfer places of all PSi shall be uniformly $n<c$. Directly after a sampling clock there shall be $x_G$ calls in the general system, with $n<x_G<c$. Then, either all $x_G$ calls may be situated in the secondary system and the primary system is empty, or less than $x_G$ but more than or equal to n calls may be situated in the secondary system and the rest is situated in the primary system. The latter case occurs, if e.g. directly before the sampling clock one PSi contains $x_G$ calls and all other PSi and the secondary system are empty. Hence, for the determination of the number of calls, served until the next sampling clock, it is not sufficient to know only the state of the general system directly after the sampling clock, but the states of all PSi and the secondary system are used. This requires the application of multi-dimensional probabilities of state, as already mentioned in section 3.2.

For symmetrical systems ($n_i=n$ and $\lambda_i=\lambda$ for $i=1,2,\ldots,g$), however, an approximate solution for the mean waiting time $w_S$ of all calls in the secondary system was found. This method is based on Model B and it is described in the following.

The absolute probabilities r(k), that k calls arrive per sampling clock at the secondary system, can be calculated by the aid of the probabilities of state $p_{Pi}(x,0)$ of the PSi directly before the sampling clock (c.f. equ.(5A,B,C) with $t_v=0$). The probability, that from one PSi $x<n$ calls are transferred, is identical with the probability $p_{Pi}(x,0)$, that the PSi contains exactly x calls directly before the sampling clock. The probability, that from one PSi exactly n calls are transferred, is identical with the probability, that the PSi contains more than or equal to n calls directly before the sampling clock, i.e.

$$1 - \sum_{x=0}^{n-1} p_{Pi}(x,0) \quad .$$

Then, the probabilities r(k) of the global group transferred from the whole primary system to the secondary system are obtained by a convolution of the above probabilities. The limits for the group size k are

$$0 \leq k \leq \sum_{i=1}^{g} n_i \quad .$$

The <u>first simple approximation</u> is, to assume the probabilities r(k) as independent from each other in successive sampling instants. Then, Model B is applicable with

$$m = \sum_{i=1}^{g} n_i = g\cdot n$$

and so the approximate mean waiting time $w_S$ is obtained from equ.(7). Simulation runs on digital computers have shown, that this simple approximation underestimates the mean waiting time $w_S$ in particular for a large offered traffic. The reason for this underestimation can be ex-

plained qualitatively:
If the maximum of n calls is transferred from one PSi to the SS, the following two facts are important:

-Firstly, the transferred group is relatively large, so that directly before the next sampling clock normally the secondary system contains relatively many calls.

-Secondly, there is a certain probability, that calls are left in the PSi and therefore a group of at least one call is transferred with the succeeding sampling clock. The waiting time of the first call of this group is relatively large because of the first fact.

In Model B, this correlation is neglected.

Now, the idea for a better approximation is to increase the mean waiting time of the first call of a group, arriving at the SS, by a factor $F_S$ against that one of Model B. Thus, from equ.(7) it follows for the approximate mean waiting time $w_S$ (cf. also remark below equ.(8)):

$$w_S = F_S \cdot E[x,0] \cdot h + \frac{1}{2}\left(\frac{Var[k]}{E[k]} + E[k] - 1\right) \cdot h \qquad (25)$$

The factor $F_S$ takes into account the correlation mentioned above. One measure for this correlation is the probability, that at least one call is left in the primary system directly after the sampling clock, i.e. the primary system contains at least one call at the time $t_v = T$ before the next sampling clock. Because of the assumption of a symmetrical system, this probability is

$$1 - (p_P(0,T))^g \ .$$

$p_P(0,T)$ is the probability, that one PS contains 0 calls directly after the sampling clock ( cf. equ.(5A,B,C) with $t_v = T$).
Lots of simulation runs have proofed, that the factor $F_S$ may be assumed generally as

$$F_S = 1 + 1.9\left[1 - (p_P(0,T))^g\right] \qquad (26)$$

In equ.(25), the factor $F_S$ is inserted from equ.(26). All other terms are calculated with equs.(6) and (8A,B) of Model B (with m=gn).
The accuracy of this approximation is sufficient for practical applications. In all cases, proofed by a lot of simulation runs, the difference between the approximate values of equ.(25) and the simulated values for $w_S$ was less than 10%.

Finally it should be noted, that the case of a maximum group size $m = gn \leq c$ is trivial, because in this case always the group of arriving calls is served completely until the next sampling clock. Therefore, the arriving groups always find an empty secondary system and so equ.(7) of Model B is applicable for $w_S$ with $E[x,0] = 0$.

### 3.3.2. APPROXIMATION OF THE MEAN WAITING TIME, WHEN THE CLOCK PERIOD IS NOT AN INTEGRAL MULTIPLE OF THE HOLDING TIME

If the sampling clock period T is not an integral multiple of the holding time h, the following approximation for the mean waiting time $w_S$ of all calls in the secondary system is applicable:
The sampling clock period shall be T=ch, where c is a real value between the two successive integral values $c_u$ and $c_o$. Then, the mean waiting times $w_{S,cu}$ and $w_{S,co}$ are calculated each according to equ.(21) or equ.(25) once with $c_u$ and once with $c_o$, respectively, without change of all other parameters. The mean waiting time $w_S$ of the real system then is obtained approximately by linear interpolation between $w_{S,cu}$ and $w_{S,co}$:

$$w_S = (c - c_u) w_{S,co} + (c_o - c) w_{S,cu} \qquad (27)$$

Simulation runs on digital computers have shown, that the approximate values of equ.(27) are rather close to the simulated values (in most cases <5% difference).

### 3.4. MEAN WAITING TIME IN THE GENERAL SYSTEM

The mean waiting time of the calls in PSi shall be $w_{Pi}$, which can be calculated from equ.(4). Then, the mean waiting time $w_P$ of all calls in the whole primary system is the weighted sum

$$w_P = \sum_{i=1}^{g} \frac{\lambda_i}{\Lambda} w_{Pi} \qquad (28)$$

The mean waiting time $w_G$ of all calls in the general system is the sum of $w_P$ and $w_S$

$$w_G = w_P + w_S \qquad (29)$$

$w_G$ is the waiting time, which a call has to wait in the mean from its arrival at the primary system until the beginning of its service in the central service unit.

### 3.5. MODIFICATION OF MODEL C WITH CYCLICALLY CLOSED SAMPLING SWITCHES.

In Model C, the sampling switches of all PSi are closed simultaneously ( cf. Fig.4). Another operating mode of such a system may be, that only one sampling switch is closed per sampling instant in a cyclical manner.
Then, again the characteristic traffic values for each primary storage PSi can be obtained from the results of Model A when taking into consideration, that the time interval between two successive sampling instants of the same PSi is $T = g \cdot c \cdot h$.

For the mean waiting time $w_S$ of all calls in the secondary system a similar approximation as in section 3.3.1 is applicable for symmetrical systems ($n_i = n$ and $\lambda_i = \lambda$ for $i = 1,2,\ldots,g$), using Model B as a basis. The absolute probabilities r(k), that k calls arrive per sampling clock at the SS, depend on the probabilities of state of only one PSi directly before the sampling clock coordinated to this PSi. They are known by the aid of equ.(5A,B,C) with $t_v = 0$ and T=gch ( cf. also section 3.3.1). The limits for the group size now are $0 \leq k \leq n$.
Similar cosiderations as in section 3.3.1, concerning the first call of an arriving group, lead to the following formula for the approximate mean waiting time $w_S$:

$$w_S = F_C \cdot E[x,0] \cdot h + \frac{1}{2}\left(\frac{Var[k]}{E[k]} + E[k] - 1\right) h \qquad (30)$$

All terms on the right hand side of equ.(30), except the factor $F_C$, are calculated with equs.(6) and (8A,B) of Model B (with m=n).

Lots of simulation runs on digital computers have proofed, that for the following different cases the factor $F_C$ in equ.(30) may be assumed as

$$\left.\begin{array}{ll} F_C = 1+5.5(1-p_P(0,T)) & \text{for } c=1 \\ F_C = 1+2.8(1-p_P(0,T)) & \text{for } c=2 \\ F_C = 1+1.9(1-p_P(0,T)) & \text{for } c \geq 3 \end{array}\right\} \underline{\text{for } g \cdot c < 30} \qquad \begin{array}{r} (31A) \\ (31B) \\ (31C) \end{array}$$

$$F_C = 1 \qquad \underline{\text{for } g \cdot c \geq 30} \qquad (31D)$$

$p_P(0,T)$ is the probability, that a PS contains 0 calls directly after the sampling clock coordinated to this PS (c.f. equ.(5A,B,C) with $t_v = T$ and $T = g \cdot c \cdot h$).

In all cases, proofed by a lot of simulation runs, the difference between the approximate values of equ.(30) and the simulated values for $w_S$ was less than 10%.

## 4. EVALUATIONS AND NUMERICAL RESULTS

Fig.7 shows the mean waiting time $w_{Pi}$ of a primary storage PSi according to equ.(4) as a function of the arrival rate $\lambda_i$. The number of transfer places $n_i$ is used as parameter.

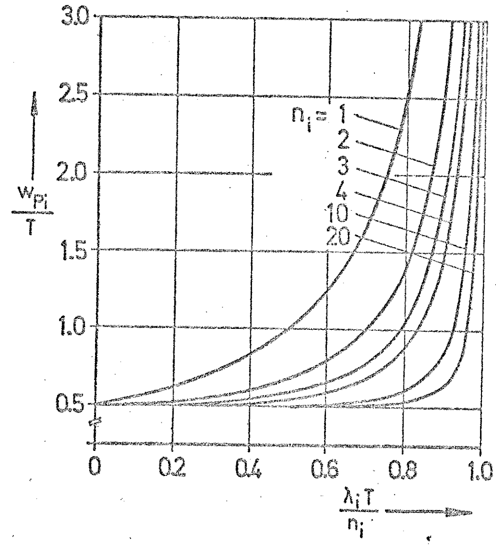


Fig.7: Mean waiting time $w_{Pi}$ of a primary storage PSi

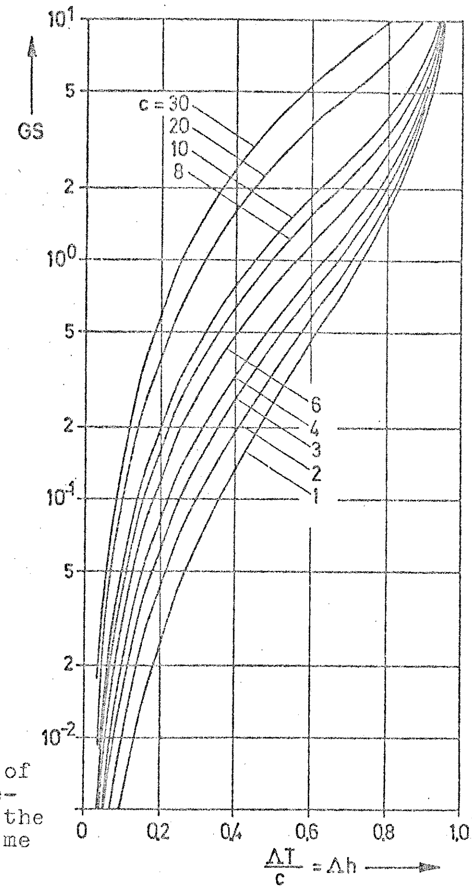


Fig.8: Term GS of equ.(32) for determination of the mean waiting time $w_S$

The mean waiting time $w_S$ of the secondary system depends on such a variety of parameters, that it is impossible to draw only one chart of $w_S$ for all parameter combinations. But by the aid of only two charts, however, the exact mean waiting time $w_S$ for systems with $c \leq \mathrm{Min}[n_i]$ can be obtained numerically as described in the following. Substituting equs.(14) and (16) into equ.(21), one gets

$$w_S = \frac{1}{\Lambda}\Bigg[\underbrace{E_G[x,T] - \frac{1}{c}\sum_{j=1}^{c}\Big(j - \sum_{x=0}^{j-1}(j-x)p_G(x,T)\Big)}_{GS} - \sum_{i=1}^{g} E_{Pi}[x,T]\Bigg] \quad (32)$$

The term indicated with GS in equ.(32) depends only on the parameters $\Lambda$ and $T=ch$ of the general system. Fig.8 shows the numerical values of GS as a function of $\Lambda$ with parameter c. Furthermore, for each PSi the mean queue length $E_{Pi}[x,T]$ directly after the sampling clock can be calculated separately according to equ.(3) with $t_V=T$. Fig.9 shows the numerical values of $E_{Pi}[x,T]$ as a function of $\lambda_i$ with parameter $n_i$. Thus, by the aid of Fig.8 and Fig.9 the mean waiting time $w_S$ can be obtained numerically according to equ.(32).

The probability of waiting $W_S(>0)$ in the secondary system according to equ.(24) is shown in Fig.10 for systems with $c<\mathrm{Min}[n_i]$. It depends only on the parameters $\Lambda$ and c of the general system.

Finally, an example for the approximate mean waiting time $w_S$ in the secondary system of a symmetrical system with $n<c$ and simultaneously closed sampling switches is given in Fig.11, according to equ.(25). Many tests have shown, that the simple approximation with Model B is sufficient for practical applications in the case of an offered traffic $\Lambda \cdot h<0.5$ (cf. also dashed line a in Fig.11). However, for an offered traffic $\Lambda \cdot h \geq 0.5$ the more accurate approximation according to equ.(25) should be applied (cf. also solid line b in Fig.11).

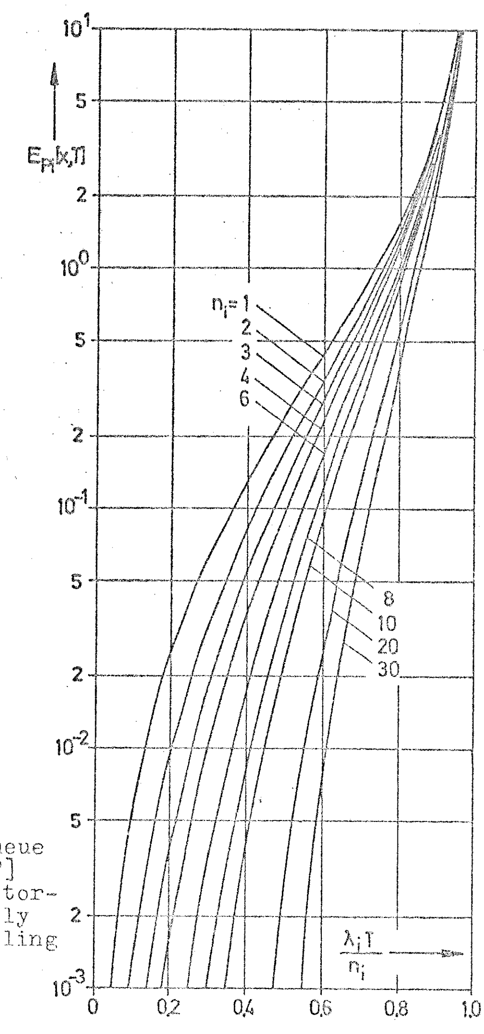


Fig.9: Mean queue length $E_{Pi}[x,T]$ of a primary storage PSi directly after the sampling clock

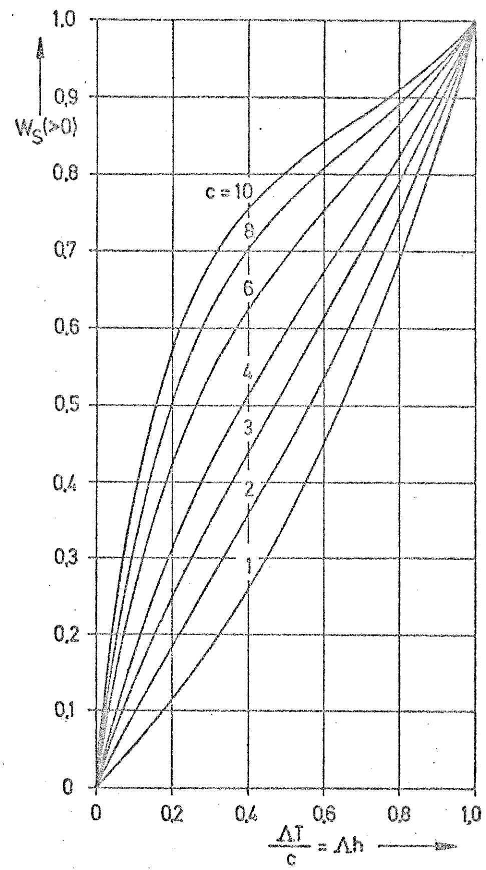


Fig. 10: Probability of waiting $W_S(>0)$ in the secondary system for systems with $c < Min[n_i]$

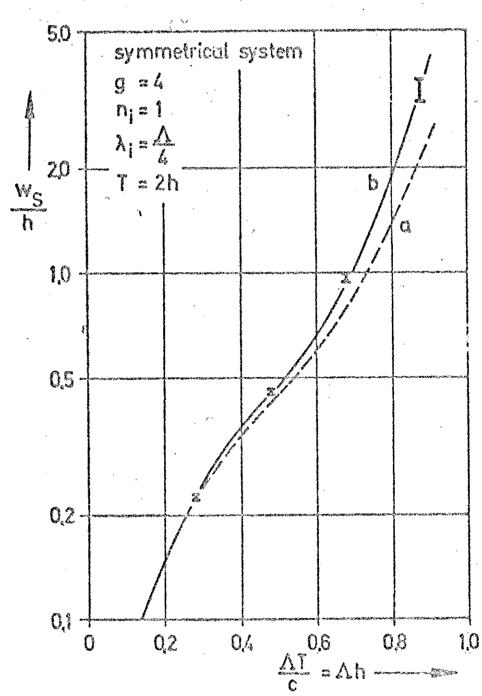


Fig. 11: Example for the mean waiting time $w_S$ in the secondary system (with $c > Min[n_i]$)

I simulation values (95% confidence interval)

dashed line a: approximation with Model B

solid line b: approximation according to equ. (25)

## 5. CONCLUSION

A two-stage queuing system is dealt with, which corresponds to a simplified model of a central controlled telephone or data switching system. The characteristical operating mode of this model is the clocked transfer of informations from the periphery to the central unit.
In principle, this model is a combination of two different single-stage models, which were earlier dealt with independently from each other in /1/. However, this two-stage model takes into account the dependency of the state of the central unit on the periphery. Exact solutions as well as approximate solutions are derived in particular for the mean waiting times.
Further investigations, to be published at some future time, will concern with a modification of the service process in the central service unit. Different classes of informations with constant, but different holding times will be taken into account.

## REFERENCES


/1/ Langenbach-Belz, M.: Sampled queuing systems. Proceedings of the Symposium on Computer-Communications Networks and Teletraffic, Volume XXII of the MRI Symposia Series, Polytechnic Press of the Polytechnic Institute of Brooklyn, Brooklyn, New York 1972.

/2/ Powell, B.A., Avi-Itzhak, B.: Queuing systems with enforced idle time. Oper. Res. 15(1967), 1145-1156.

/3/ Delbrouck, L.E.N.: A multi server queue with enforced idle times. Oper. Res. 17(1969), 506-518.

/4/ Bailey, N.T.J.: On queuing processes with bulk service. Journ.Roy.Stat.Soc., Ser.B, 16(1954), 80-87.

/5/ Downton, F.: Waiting time in bulk service queues. Journ.Roy.Stat.Soc., Ser.B, 17(1955), 256-261.

/6/ Chan, W.-C.: Computer-controlled queuing system with constant-access cycle and general service times. Proc. IEE, 117(1970), 927-930.

/7/ Miller, R.G.: A contribution to the theory of bulk queues. Journ.Roy.Stat.Soc., Ser.B, 21(1969), 320-337.

/8/ Boudreau, P.E., Kac, M.: Analysis of a basic queuing problem arising in computer systems. IBM Journ. Res. and Dev., 5(1961), 132-140.

/9/ Crommelin, C.D.: Delay probability formula when the holding times are constant. Post. Off. Elect. Engrs., 25(1932), 41-50 and 26(1933/34), 266-274.

/10/ Pollaczek, F.: Zur Theorie des Wartens vor Schaltergruppen. Elekt. Nachr. Techn., 9(1932), 434-454.

/11/ Müller, B.: Simulationsprogramm für abgetastete Wartespeicher. Institute for Switching and Data Technics, University of Stuttgart, Monograph 1970.

/12/ Toll, P.: Berechnung von Wartezeiten und Zustandswahrscheinlichkeiten bei abgetasteten Wartespeichern. Institute for Switching and Data Technics, University of Stuttgart, Monograph 1972.

/13/ Blass, Ch.: Exakte und näherungsweise Berechnung der mittleren Wartezeit im Sekundärteil eines zweistufigen, getakteten Wartesystems. Institute for Switching and Data Technics, University of Stuttgart, Monograph to be published 1973.

/14/ Feller, W.: An introduction to probability theory and its applications. Volume 1, John Wiley & Sons, N.Y., 1957.