STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Report S 350 (VP 25)
SP 91 ar

Some Queueing Models with Dependent Service Times

Preliminary report

by

F. Göbel

(first draft, July 1965)

Corrections to
Some Queueing Models with Dependent Service Times
by
F. Göbel

Report S 350	Beograd, paper 110	
p.3, lines 8 & 9	p.3, lines 12 & 13	Delete "when all tanks are infinite".
p.5, formula 7	p.5, formula 7	$\sum_{k=1}^{\infty} \check{C}_{kj}(t)z^k = z \frac{(1-u_j)(\lambda-t)\check{C}_{1j}(t)-tu_j\check{C}_{1j}(\lambda-\lambda u_j)}{(1-u_j)\{\lambda-t-z\lambda\check{B}_j(t)\}}$
p.7, line 13	p.7, line 13	...$\underline{r}_{jk}$, an... → ...$\underline{r}_{jk}$, of which an...
p.9, line 21	p°9, line 21	a sub-sequence of consecutive customers of one type
p.11, line 8	p.11, line 8	factor → further
p.11, line 21	p.11, line 21	...k, complexity... → ...k, the complexity...

Printed at the Mathematical Centre at Amsterdam,49,2nd Boerhaavestraat.
The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non - profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

Summary. Customers carrying goods of different types arrive at an unloading station, at which the goods are unloaded into transit sheds (tanks). The waiting time of the customers is discussed for one, two, or more infinite tanks.

Résumé. Des usageurs portant des cargaisons de types différents, arrivent à un débarcadère où ils débarquent les marchandises dans des dépôts (réservoirs). Le temps d'attente est discuté pour un, deux ou plusieurs réservoirs infinis.

1. The general model

The queueing model to be described presently is not intended as the introduction to the formulation of a problem, but rather as a framework into which a number of simpler problems fit, some of which we will discuss.

The arrival process has the following properties: customers (ships, trucks, etc.) arrive at a single service channel, called unloading line, carrying loads (e.g. crude oils, ore) of varying sizes and of m different types; however, each customer carries one type only; we shall use the term "j-customer" if the load is of type j. The customers of the various types arrive independently of all others, according to a Poisson process with density $\lambda_j = p_j\lambda$, $\Sigma p_j = 1$. The size [1] $\underline{s}_j$ of a load of type j has an arbitrary distribution, which only depends on the type j:

$$P\{\underline{s}_j \leq s\} = D_j(s). \qquad (1)$$

The service consists of unloading the goods into one of a number of warehouses or tanks of given capacities. Type j is unloaded at a rate τ_j (units per units of time), whereas the switch-over times from one unloading operation to another are negligible. There is only one unloading line. A customer leaves the system when the unloading has been completed.

The tanks, which can be emptied [2], can contain only one type at a time. When a tank containing type j is emptied, this occurs at a rate ω_j. Only one tank can be emptied at a time. After a tank has been emptied, it can be used for another type. It is possible to simultaneously fill and empty a tank. It is also possible to start or to continue emptying a tank, before one has completely emptied the previous tank, hence we do not a priori exclude the possibility of **preemptive** resume priorities.

---

[1] Stochastic variables are underlined.

[2] We will make a consistent distinction between unloading (of customers) and emptying (of tanks).

Figure 1 gives a schematic representation of the system.

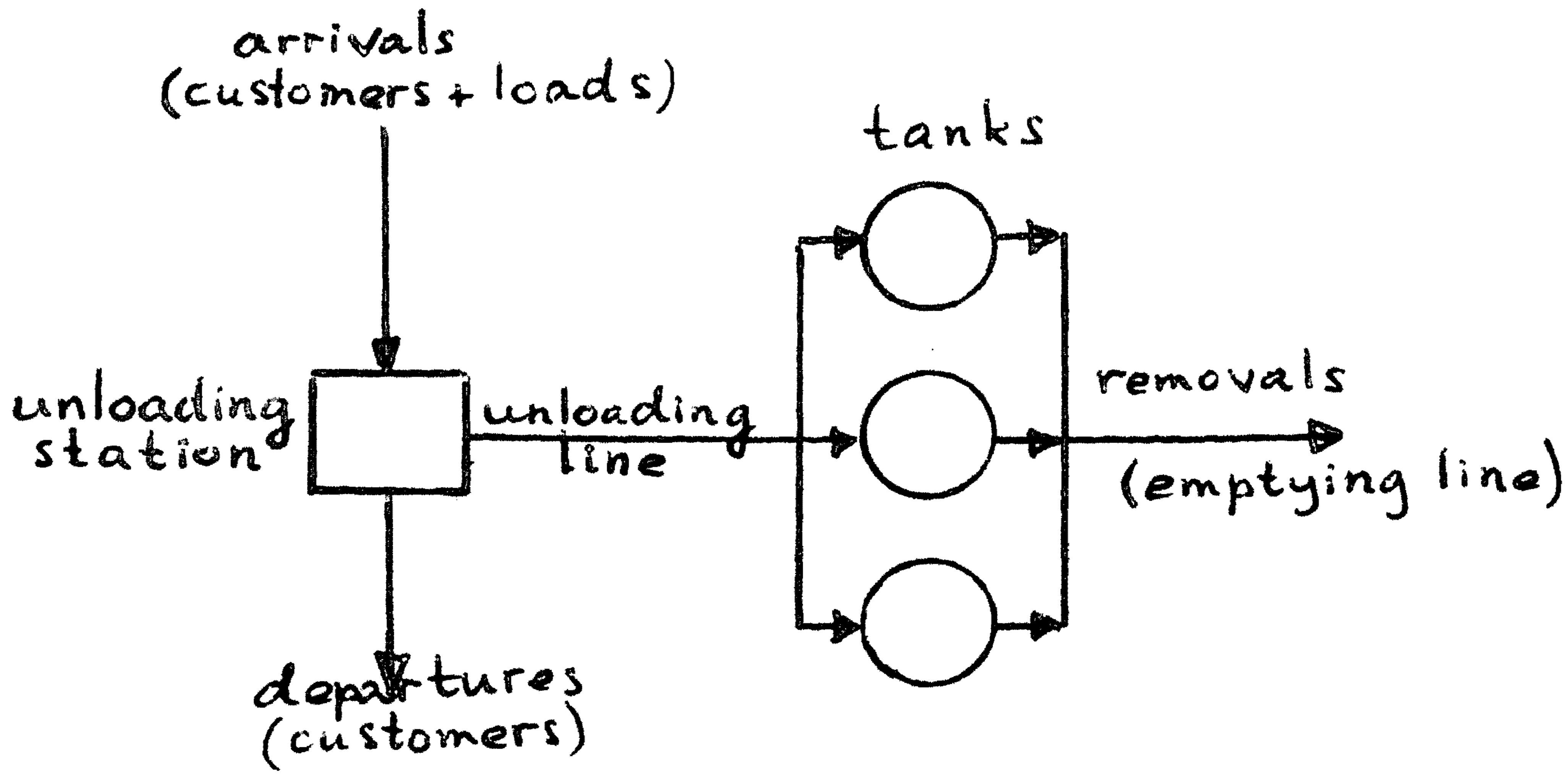

Figure 1

Although the situation is a very general one, it is clear that at some points quite arbitrary restrictions are made (like the number of unloading lines being 1).

A related model has been considered by GAVER [1], the main difference being that his "orientation times" do not accumulate like our "emptying times" do.

2. Some special cases

If the number of tanks is equal to (or even greater than) the number m of different types, and if moreover each tank has infinite capacity, then the problem is a well-known one, as far as the waiting-times of the customers and the queue length are concerned. For, in KENDALL's notation, this is the M/G/1 situation (exponential interarrival times, general service times, one service channel), the service times being the unloading times and the one service channel, of course, the unloading line. A less well-known problem arises when one gets interested in the process of the amounts in the tanks (given a rule for emptying them).

Specializing further still, by taking $m = 1$ in the preceding case (one type and one infinite tank), and $\tau_1 = \infty$, one obtains the problem of the infinite dam (MORAN [3], TAKÁCS [5]) provided the rule for emptying the tank is: "Empty the tank whenever possible".

A special case in which the customers do experience the influence of the tanks is the following: there is one infinite tank, $m > 1$; $\tau_j > \omega_j$ for all j. In section 3 this case will be treated in some detail.

In sections 4 and 5 some results will be given for two or more infinite tanks, while $\tau_j = \infty$ for all j.

Before proceeding to the special cases, let us mention that the distribution of the duration of the wet periods (i.e. the periods in which at **least** one of the tanks is not empty), is known when all tanks are infinite. The following relation is given by TAKÁCS [5].

$$P\{\underline{\theta} \leq t\} = \frac{1}{\lambda} \int_0^t \frac{1}{u} P\{u < \underline{\chi}(u) \leq u + du\}$$

where $\underline{\theta}$ is the length of a wet period, and

$$\underline{\chi}(u) = \underline{\chi}_1 + \underline{\chi}_2 + \dots + \underline{\chi}_{\underline{\nu}(u)}$$

where $\underline{\nu}(u)$ has a Poisson distribution with parameter λu and $\underline{\chi}_i$ is the amount carried by an arbitrary customer, so that $\underline{\chi}_i$ has the distribution function $\sum_j p_j D_j(\omega_j s)$.

3. One infinite tank, $\tau_j > \omega_j$

In the case of one tank it is natural to assume that the tank is emptied whenever it contains a positive quantity. Other strategies are possible, but this one is certainly not worse than any other.

Things are quite different with the unloading. Although our considerations will be limited to the first-come first-served discipline, it is clear that there are other methods, which lead to shorter waiting-times. It would be interesting to know, for example, how the (expected) waiting-time will be influenced if one adopts "opportunist" priorities, by which we mean that priority is given to customers who carry a load of the type in the tank, if the tank is not empty. (If there is more than one customer waiting who carries that type, or when the tank is empty, one could apply the first-come first-served discipline again.) This method, of opportunist priorities, has the unusual feature that it is not possible to once-and-for-all rank the types, i.e. the classes to which the customers belong, in order of priority.

Whatever the discipline adopted, a stationary state can exist only when every arriving unit can eventually be taken out of the tank (since $\tau_j > \omega_j$, this is indeed the bottle-neck); hence a necessary (and presumably sufficient) condition for the existence of a stationary state is:

$$\lambda\mu^{*} < 1 \tag{2}$$

where

$$\mu^{*} = \sum_j p_j \, \omega_j^{-1} \mathcal{E} \, \underline{s}_j . \tag{3}$$

Suppose now that we adopt the first-come first-served discipline. When a customer is admitted to the unloading station, he will still have to wait till the tank has been emptied, unless it contains goods of the same type, in which case the unloading can start immediately.

Assuming that the process is stationary, we will determine the <u>waiting-time</u> of an arbitrary customer. To this end, we shall consider, beside the "original process", a "modified process" in which a customer <u>never</u> starts unloading before the tank has been emptied. Now, in the modified process we have the familiar M/G/1 situation, the arrival times being the same as in the original process, and the service times being the times required to clear the tank from the load of an arbitrary customer. Hence, the distribution function D(s) of the service times is given by

$$D(s) = \sum_j p_j \, D_j(\omega_j s) . \tag{4}$$

For this process the stationary waiting-time distribution can be found; its Laplace Stieltjes transform $\check{H}(t)$ is given by the Pollaczek formula:

$$\check{H}(t) = \frac{1 - \rho}{1 - \rho \dfrac{1 - \check{D}(t)}{\mu^{*} t}} \tag{5}$$

where $\rho = \lambda\mu^{*}$ and $\check{D}(t)$ is the Laplace Stieltjes transform of D(s) (see e.g. KENDALL [2]).

Consider a customer who is of a type which differs from the type of his immediate predecessor. For such a customer, not only the arrival time, but also the moment at which his service starts in the modified process coincides with the corresponding moment in the original process. Therefore, the first customer of a sequence of consecutive customers of the same type, has the same waiting-time in the two processes. (We might add that the distribution of the waiting-time in the modified process is independent of the type, which information must eventually be used in the original process. This can be seen by noting that it is not necessary to consider the type before the service starts, so that the waiting-time, which has been completed by then, cannot possibly depend on the type).

Now, going back to the original process, consider a customer who is the k-th in a sequence of consecutive j-customers. The probability that an arbitrary customer has this property is $(1 - p_j)p_j^k$, since the event in question is equivalent to: the customer is of type j, the k-1 preceding customers are of type j, and the k-th customer, counting backwards, is of a different type. Of course, a stationary state is never reached within a sequence; here we are concerned with the transient behavior. Since a customer leaves the system after having completed the unloading operation, his service time has the distribution function

$$B_j(s) = D_j(\tau_j s) \tag{6}$$

Let $\check{C}_{kj}(t)$ be the Laplace Stieltjes transform of the waiting-time of the k-th customer in a j-sequence. Then

$$\sum_{k=1}^{\infty} \check{C}_{kj}(t) z^k = \frac{z(1-u_j)(\lambda-t)\check{C}_{1j}(t) - tu_j \check{C}_{1j}(\lambda-\lambda u_j)}{(1-u_j)\{\lambda-t-2\lambda\check{B}_j(t)\}} \tag{7}$$

where u_j is the root with the smallest absolute value satisfying the equation in u:

$$u = z\,\check{B}_j(\lambda-\lambda u), \tag{8}$$

and where $\check{B}_j(t)$ is the Laplace Stieltjes transform of $B_j(s)$. (Cf. TAKÁCS [4], p. 55-56).

The function $\check{C}_{kj}(t)$ occurs in (7) in a rather implicit manner: first we have to solve the equation (8), and then we still only have a generating function for the required sequence of functions. The latter difficulty, however, disappears as we turn our attention to $\check{C}_j(t)$, the Laplace Stieltjes transform of the waiting-time $\underline{w}_j$ of an arbitrary j-customer.

In (7) and (8) we replace z by p_j and $\check{C}_{1j}$ by $\check{H}$, and multiply by $p_j^{-1} q_j = p_j^{-1}(1-p_j)$, thus obtaining:

$$\check{C}_j(t) = \frac{q_j}{1-u_j} \circ \frac{(1-u_j)(\lambda-t)\check{H}(t) - tu_j\check{H}(\lambda-\lambda u_j)}{\lambda-t-\lambda p_j\check{B}_j(t)} \tag{9}$$

where u_j is now the root with the smallest absolute value satisfying the equation in u:

$$u = p_j\ \check{B}_j(\lambda-\lambda u). \tag{10}$$

Hence, the Laplace Stieltjes transform $\check{C}(t)$ of the waiting-time $\underline{w}$ of an arbitrary customer is given by

$$\check{C}(t) = \sum_j \frac{p_j q_j}{1-u_j} \circ \frac{(1-u_j)(\lambda-t)\check{H}(t) - tu_j\check{H}(\lambda-\lambda u_j)}{\lambda-t-\lambda p_j\check{B}_j(t)} \circ \tag{11}$$

By differentiating and substituting t = 0, we obtain the following formula for the expected waiting-time of an arbitrary customer:

$$\mathcal{E}\underline{w} = \mathcal{E}\,\underline{h} - \sum_j p_j\ \lambda^{-1}\ \left\{\frac{p_j}{q_j}\,(1-\lambda\,\mathcal{E}\,\underline{b}_j) - \frac{u_j}{1-u_j}\,\check{H}(\lambda-\lambda u_j)\right\} \tag{12}$$

where $\underline{h}$ is the waiting-time in the modified process, and $\underline{b}_j$ the service time (unloading-time) of a j-customer.

In the special case $\tau_j = \infty$ for all j, it follows at once from (10) that $u_j = p_j$, while $\mathcal{E}\,\underline{b}_j = 0$, hence

$$\check{C}(t) = \sum_j p_j\ \frac{q_j(\lambda-t)\check{H}(t) - tp_j\check{H}(\lambda q_j)}{\lambda q_j - t} \tag{13}$$

and

$$\mathcal{E}\underline{w} = \mathcal{E}\underline{h} - \frac{1}{\lambda}\sum p_j^2 q_j^{-1}\,\{1 - \check{H}(\lambda q_j)\}. \tag{14}$$

So far, we have not been able to solve the problem of the amounts in the tank. Suppose we consider the amount in the tank at the times when a customer leaves the system (in the original process). The collection of these times includes all times at which the amount in the tank is maximum. Although this is not the simplest way to obtain results on the amounts, it is presumably necessary to solve this problem before results on the queue lengths can be obtained. We will just indicate now, how the above quantities are related to quantities that play a role in the determination of the waiting times.

Let the amount in the tank on the departure of the k-th customer of a j-sequence be denoted by $\underline{r}_{jk}$, an amount $\underline{r}''_{jk}$ was in the tank at the beginning of the unloading, and let $\underline{r}'_{jk} = \underline{r}_{jk} - \underline{r}''_{jk}$. Then [1)]

$$\underline{r}'_{jk} \triangleq \underline{r}_{j1} \triangleq (1 - \omega_j \tau_j^{-1})\underline{s}_j,$$

and

$$\underline{r}''_{jk} = \omega_j\, \underline{v}_{jk},$$

where $\underline{v}_{jk}$ is the time by which the unloading of the k-th customer of a j-sequence can be advanced when going back from the modified process to the original process. Omitting the subscript j, we have $\underline{v}_1 = 0$ and $\underline{v}_k \leq \underline{h}_k$, where $\underline{h}_k$ is the waiting time in the modified process of the k-th customer (of a j-sequence). One can verify that, for $k \geq 2$, $\underline{v}_k = \underline{h}_k$ unless $\underline{h}_k > \underline{v}_{k-1} + (\omega^{-1} - \tau^{-1})\underline{s}_{k-1}$ in which case $\underline{v}_k = \underline{v}_{k-1} + (\omega^{-1} + \tau^{-1})\underline{s}_{k-1}$, hence

$$\underline{v}_k = \min\,\{\underline{h}_k,\ \underline{v}_{k-1} + (\omega^{-1} - \tau^{-1})\underline{s}_{k-1}\} \quad (k \geq 2).$$

---

1) $\underline{x} \triangleq \underline{y}$ means: $\underline{x}$ and $\underline{y}$ are identically distributed.

4. Two infinite tanks

We will now consider the case where the customers can unload into one of two infinite tanks, and the number of different goods is greater than two, with the additional restriction $\tau_j = \infty$ for all j. Since we shall not be interested in the amounts in the tanks, it can be assumed without loss of generality that all ω_j are 1 (by adjusting the sizes of the loads).

As soon as the number of tanks is greater than 1, it is no longer obvious how to define a stationary state. Of course, we will require the expected queue-length to be finite. However, the values of the various parameters may be such that a finite expected queue length can only be achieved by allowing the average amount of goods in one of the tanks to be infinite. Since this is not intended, we will require that there exists a method of emptying the tanks such that the expected stock size is finite. This again leads to the condition $\lambda\mu^{**} < 1$.

Another consequence of introducing a second tank is that one has to choose a strategy for emptying the tanks. We will assume that the loads are taken away in order of arrival.

We now modify the process to the effect that every customer is forced to postpone his unloading operation until one tank is entirely empty while the other tank contains at most one load. As in the one-tank case, the types play no role in the modified process.

After a customer has joined the queue, the load that he carries will be in one of the following states: it waits until it can enter a tank, it waits until it is taken out of the tank, or it is being taken out of the tank (disregarding the unloading itself which takes no time). The corresponding time intervals will be denoted by $\underline{h}$, $\underline{e}$, $\underline{s}$. Putting $\underline{h} + \underline{e} = \underline{z}$, the following relation holds for the total waiting-times of the n-th and (n+1)-st loads:

$$\underline{z}_{n+1} = \max(0, \underline{z}_n + \underline{s}_n - \underline{y}_n) \tag{15}$$

where $\underline{y}_n$ is the interarrival time between the n-th and the (n+1)-st customers. Hence, the Laplace Stieltjes transform of the stationary distribution of $\underline{z}$ is given by:

$$\check{Z}(t) = \frac{1 - \rho}{1 - \rho \dfrac{1 - \check{D}(t)}{\mu^{*} t}} \quad . \tag{16}$$

On the other hand, when the quantities $\underline{e}_n$ are considered as service times, we obtain the relation

$$\underline{h}_{n+1} = \max(0, \underline{h}_n + \underline{e}_n - \underline{y}_n) = \max(0, \underline{z}_n - \underline{y}_n) \tag{17}$$

where $\underline{z}_n$ and $\underline{y}_n$ are independent.

From (17) one has

$$\check{H}_{n+1}(t) = \frac{\lambda \check{Z}_n(t) - t \check{Z}_n(\lambda)}{\lambda - t} \tag{18}$$

where $\check{H}_{n+1}(t)$ is the Laplace Stieltjes transform of $H_{n+1}(h) = P\{\underline{h}_{n+1} \leq h\}$, and $\check{Z}_n(t)$ that of $Z_n(z) = P\{\underline{z}_n \leq z\}$. Hence, letting $n \to \infty$,

$$\check{H}(t) = \frac{\lambda \check{Z}(t) - t \check{Z}(\lambda)}{\lambda - t} \tag{19}$$

$$\mathcal{E}\underline{h} = \mathcal{E}\underline{z} - \frac{1 - \check{Z}(\lambda)}{\lambda} \quad . \tag{20}$$

In the model we are considering, it is not necessary that a customer waits until there is only one load left in the tank: some unloading times can be advanced. However, the sequences [1] in the original process are very complicated. For example, a sequence can contain all types, in sharp contrast to the one-tank case. Moreover, it is not at all clear whether the first-come first-served principle, which is in many cases adopted for the sake of simplicity, had not better be dropped in favor of some form of (opportunist) priorities. For the moment, let us concentrate on the unloadings that can be advanced without much difficulty, namely the unloadings corresponding to a sub-sequence of customers of one type. We are then back in the situation of section 3, and the Laplace Stieltjes transform of the waiting-time, and the expected waiting-time are given by (13) and (14), the

---

1) A precise definition of a sequence can be given as follows: a new sequence is initiated by every customer whose departures in the modified and the original process coincide.

only difference being, of course, the meaning of $\tilde{H}$ and of $\mathcal{E}\,\underline{h}$. Substituting (19) and (20) into (13) and (14), we obtain

$$\tilde{\mathcal{C}}(t) = \sum p_j \frac{\lambda q_j \tilde{Z}(t) - t\tilde{Z}(\lambda q_j)}{\lambda q_j - t} \tag{21}$$

$$\mathcal{E}\underline{w} = \mathcal{E}\underline{z} - \frac{1}{\lambda} \sum \frac{p_j}{q_j} \{1 - \tilde{Z}(\lambda q_j)\} \tag{22}$$

where $\underline{w}$ is the waiting-time of an arbitrary customer for the abovementioned way of advancing the unloading times.

On comparing (14) and (22), it can be seen at once that with the present policy, the expected waiting-time is uniformly less than the expected waiting-time given by (14), in accordance with what one would intuitively expect.

5. Three or more infinite tanks

When there are k tanks ($k \geq 3$), it is still possible to modify the process in a manner similar to the modification in section 4, namely by allowing at most k loads in the tanks. Hence, when a load enters its tank, at most k-1 loads are present. Each of these loads has to be removed before the load just entered can be removed. Hence, the total waiting-time $\underline{z}_{n,1}$ of the n-th load can be split as follows:

$$\underline{z}_{n,1} = \underline{h}_{n,k} + \underline{e}_{n,k-1} + \ldots + \underline{e}_{n,1} \tag{23}$$

where $\underline{h}_{n,k}$ is the waiting-time of the n-th customer until the unloading takes place, and $\underline{e}_{n,i}$ the time the n-th customer has to wait for the removal of the i-th load present ($i = 1, \ldots, k-1$), where the loads present are counted in order of arrival. Of course, when less than k-1 loads are present just before the load in question enters its tank, some of the $\underline{e}_{n,i}$ are 0. More generally, define

$$\underline{z}_{n,i} = \underline{h}_{n,k} + \underline{e}_{n,k-1} + \ldots + \underline{e}_{n,i} \quad (i = 1,\ldots,k-1) . \tag{24}$$

Then, by the same argument as in section 4,

$$\underline{h}_{n+1,k} = \max(0, \underline{z}_{n,k-1} - \underline{y}_n) \tag{25}$$

$$\underline{z}_{n+1,i} = \max(0, \underline{z}_{n,i-1} - \underline{y}_n) \quad (i=2,\ldots,k-1) \tag{26}$$

$$\underline{z}_{n+1,1} = \max(0, \underline{z}_{n,1} + \underline{s}_n - \underline{y}_n) \tag{27}$$

where $\underline{s}_n$ is the service time of the n-th customer and $\underline{y}_n$ the inter-arrival time between the n-th and (n+1)-st customers.

Taking the Laplace Stieltjes transform of (25) and (26), and letting $n \to \infty$, one obtains after some factor calculations

$$\mathcal{E}\,\underline{h}_k = \mathcal{E}\,\underline{z}_1 - \frac{k-1}{\lambda} + \frac{1}{\lambda}\sum_{i=0}^{k-2} (k-i-1)\,\frac{(-\lambda)^i}{i!}\,\check{Z}_1^{(i)}(\lambda) \tag{28}$$

where $\check{Z}_1^{(i)}(\lambda)$ is an abbreviation for $\left[\frac{d^i}{dt^i}\check{Z}_1(t)\right]_{t=\lambda}$, and $\underline{h}_k = \lim_{n\to\infty} \underline{h}_{n,k}$, $\underline{z}_1 = \lim_{n\to\infty} \underline{z}_{n,1}$; hence, $\mathcal{E}\,\underline{h}_k$ is the expected waiting-time of an arbitrary customer in the modified process when there are k tanks.

It may be noted that $\underline{h}_k$ has an interpretation in the M/G/1 situation: $\underline{h}_k$ is the waiting-time from the moment of arrival until the first moment at which there are at most k-1 other customers in the system (including the customer being served).

Here too, it is possible to determine the expected waiting-time of an arbitrary customer when his unloading is advanced in the manner mentioned in section 4. For k = 3, one has

$$\mathcal{E}\,\underline{w} = \mathcal{E}\,\underline{z}_1 - \frac{1}{\lambda}\sum \frac{1}{q_j}\,\{1 - \check{Z}(\lambda q_j)\} + \frac{m-1}{\lambda}\,\{1 - \check{Z}_1(\lambda)\}. \tag{29}$$

For larger values of k, complexity of the formula increases rapidly. The convexity on $(0,\infty)$ of the Laplace Stieltjes transform is now required to verify that $\mathcal{E}\,\underline{w}$ as given by (29) is uniformly smaller than the corresponding result for two tanks (cf. (22)).

Acknowledgements. I am grateful to Prof. J.Th. RUNNENBURG for his guidance. It is a pleasure to thank him and Drs. W. MOLENAAR for many helpful discussions.

References.

1. D.P. GAVER, A comparison of queue disciplines when service orientation times occur; Naval Research Logistics Quarterly 10 (1963), p. 219-235.

2. D.G. KENDALL, Some problems in the theory of queues; J. Royal Stat. Soc. B13 (1952), p. 151-185.

3. P.A.P. MORAN, The theory of storage; Methuen & Co, London 1959.

4. L. TAKÁCS, Introduction to the theory of queues; Oxford University Press, New York 1962.

5. L. TAKÁCS, Combinatorial methods in the theory of dams, J. Appl. Prob. 1 (1964), p. 69-76.