ON QUEUEING WITH REGULAR SERVICE INTERRUPTIONS

by

P. Robillard and P. Naor

University of North Carolina

Institute of Statistics Mimeo Series No. 489

August 1966

ABSTRACT

A queueing situation is considered where customers arrive in a stationary Poisson stream and their service times are independently, identically and exponentially distributed. The single station present in the system interrupts and renews its service activities in a regular cyclic fashion. Average queue sizes at various epochs and during different periods are derived by approximative procedures. A number of exact relations between these quantities is established.

This research was supported by the Office of Naval Research under Contract No. Nonr-855(09).

DEPARTMENT OF STATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

ON QUEUEING WITH REGULAR SERVICE INTERRUPTIONS

by

P. Robillard and P. Naor [1]

- -

I. INTRODUCTION

In industrial, traffic, military, medical (hospital) and other practice queueing situations are frequently encountered in which the service station discontinues and resumes its operation intermittingly. Thus, for instance, the station may physically break down and repairs have to be carried out before resumption of service; in the meantime a queue is being built up. Or again a train of high priority customers may seize upon the station and effectively displace an accumulating queue of low priority customers from service for some time. In both of these types of station breakdown (and in some others) it is not unreasonable to assume that the disruption of service was brought about in some Poissonion fashion. Models of this character have received rather extensive treatment in some recent communications, e.g. in studies by Gaver (1962) and by Avi-Itzhak and Naor (1963).

A number of cases presents itself in which the mode of interruption possesses a different character. Thus, for instance, the discontinuation of service (for some time) may be associated with some savings as well with

---

1. On leave of absence from Technion, Israel Institute of Technology, Haifa, Israel.

additional expense due to increased average queue size. Hence if a decision maker is able to exercise detailed control an optimal policy may be to close down the service station when no customers are in the system and reopen it after a queue of prescribed size has accumulated. A rather general model pertaining to such a situation (as well as optimisation procedures) was presented by Yadin and Naor (1963). However, it may not always be possible to control the system in the fashion described above. The "next best" procedure appears to be installing a periodic cycle made up of prescribed service and shutdown phases. This is also a very reasonable course of action whenever the decision maker has to apportion the time of the service station to a number of "mutually exclusive" customer streams. This set of problems is easily exemplified by an intersection where the various streams of vehicles are regulated by a "fixed-cycle" traffic signal. The fixed-cycle traffic signal has been approached and discussed from various viewpoints; suffice it to quote two recent papers (Buckley and Wheeler (1964) and Newell (1965)) in which - additional to the original contribution presented - older work is extensively reviewed and discussed. One view that has emerged in a number of papers is that many results are not very sensitive to distribution assumptions of arrivals and departures. If that is indeed the case it appears to be advantageous to study the queueing system with regular service interruptions under those distribution assumptions which are most easily amenable to mathematical analysis, to wit: Poisson arrival and negative-exponential service. Surprisingly enough this approach appears not to have been taken before. Poisson arrivals have been considered in this context but most authors assumed the departure pattern to possess deterministic character. This latter assumption is neither warranted on using experimental evidence from a traffic intersection nor is it mathematically convenient.

One purpose of this communication is to derive properties - some exact, some approximate - of a queueing system in which the (single) station is interrupted from time to time in a fixed-cycle mode. It is not assumed that the model under discussion is a faithful representation of vehicular traffic passing through an intersection regulated by a fixed-cycle traffic signal. Rather it may serve for the purposes of approximation and bounding when such a traffic situation is evaluated. Again in contexts other than traffic - allocation of times to various computing jobs, say- the apportioning of uninterrupted service times within a general cyclic arrangement may be of some interest.

No optimization procedures will be presented here since they depend, perforce, on two additional assumptions: a) a cost structure relating to the waiting time of differing customer streams; b) dead time due to re-orientation and/or switching costs.

II. SOME BASIC RELATIONS

The notation and the nomenclature will be introduced as the occasion will warrant. They will be (more or less) in conformance with the general usage in the literature on queueing theory.

Consider a cycle of duration T. Each cycle is made up of two phases - one during which the station is capable of rendering service and one during which the station is shut down. For convenience of discussion they will be referred to as green and red phases respectively, and their durations (prescribed constants under our control) will be denoted by θ and τ respectively.

$$\theta + \tau = T \tag{1}$$

As a matter of convention we shall stipulate that a cycle starts with the green phase and ends with a red one.

Customers arrive at the station in a steady Poisson stream at a rate λ which is independent of time. They are being discharged from the station (if present) during green periods only and the mode of service is such that

a) all service times are identically distributed,

b) the distribution is negative-exponential with parameter μ, and

c) any two service times are independent of each other.

Without formal proof we state that (as in all other analogous systems) a necessary and sufficient condition for a steady state regime to exist is non-saturation; that is: the average arrival rate of customers must fall short of the station's average capacity to discharge customers on completion of service. Now within the duration of a cycle λT customers make their appearance (on the average); the station (if called upon) can render service to an average of $\mu\theta$ customers. If we designate $\frac{\lambda}{\mu}$ and $\frac{\lambda T}{\mu\theta}$ by b and B respectively the necessary and sufficient condition for steady state to prevail is represented by

$$b \leq B < 1 \tag{2}$$

The physical meaning of b and B is the average "station busy-ness" * and average "station busy-ness" during green phases, respectively. Hence another wording for the above-mentioned criterion of steady state is that during (some) green periods "patches" of idle station time must exist.

The steady state is characterized in the following fashion: Let $p_i(t)$ $(i = 0, 1, 2, \ldots ; t \geq 0)$ be the probability of i customers - including the one receiving service - queueing up at the station at time t. Typically $p_i(t_1)$ is not equal to $p_i(t_2)$; however if and only if the following relation holds for all values of i and t

$$p_i(t + k\,T) = p_i(t) \qquad \text{(integral } k\text{)} \tag{3}$$

a steady state regime pertaining to the system is said to exist.

We are now in a position to state (and derive) a number of basic relations. Consider a green phase; from the model assumptions it is clear that the relations between the various states are governed by a birth-and-death process.

$$\frac{d\,p_o(t)}{dt} = -\ \lambda\, p_o(t) + \mu\, p_1(t) \qquad (0 \leq t \leq \theta) \tag{4a}$$

$$\frac{d\,p_i(t)}{dt} = (\lambda + \mu)\; p_i(t) + \lambda\, p_{i-1}(t) + \mu\, p_{i+1}(t) \qquad (i = 1, 2, \ldots)\quad (0 \leq t \leq \theta) \tag{4b}$$

* Average "station busy-ness" is the fraction of time during which the station actually renders service.

Next we define the average station "busy-ness" at time t as

$$\sum_{i=1}^{\infty} p_i(t) = b(t) \qquad (5)$$

The average total queue q(t) at time t is given by

$$q(t) = E\{i|t\} = \sum_{i=0}^{\infty} i\, p_i(t) \qquad (6)$$

As steady state conditions are assumed to **exist it is clear** that the average increase, R, of queue size during a red phase equals

$$R = q(T) - q(\theta) = q(0) - q(\theta) = \lambda\tau \qquad (7)$$

Obviously R is also equal to the average decrease of queue size during a green phase; in short it is a measure of the non-random pulsations of the queue due to intermittent opening and closing-down of the serive station.

If equation (4b) is multiplied by i and the result summed over all values of i the following interesting relation is obtained

$$\frac{dq(t)}{dt} = \lambda - \mu\ b(t) \qquad (8)$$

Equation (8) has an immediate physical interpretation: the change of the average queue in unit time is made up of two components (of opposite sign): a) the number of arrivals in unit time λ; b) the product of service rate μ and the probability of rendering service b(t), that is, the effective service rate μ b(t).

If differential equation (8) is integrated between 0 and θ we derive

$$\int_0^{\Theta} d\, q(t) = q(\Theta) - q(0) = -\lambda\tau$$

$$= \int_0^{\Theta} (\lambda - \mu\, b(t))\, dt = \lambda\,\Theta - \mu \int_0^{\Theta} b(t)\, dt \tag{9}$$

which yields after a little manipulation

$$\int_0^{\Theta} b(t)\, dt = \frac{\lambda T}{\mu} \tag{10}$$

But the ℓ.h.s. of (10) is the time invested by the station (during a green phase or equivalently during a whole cycle) in rendering service. This equation then is consistent with the definitions (given before) of the "average busy-ness", b, and of the "average busy-ness" during a green phase, B.

Next consider equation (4b) again, multiply by i(i+1) and sum over all i. This operation results in

$$\frac{d}{dt} E\left\{i(i+1)\,|t\right\} = 2\left[\lambda - q(t)\,(\mu - \lambda)\right] \tag{11}$$

Utilizing relation (8), the ℓ.h.s. of equation (11) can be rewritten as

$$\frac{d}{dt} E\left\{i(i+1)\,|t\right\} = \frac{d}{dt}\left[V\left\{i\,|t\right\} + q^2(t) + q(t)\right] =$$

$$= \frac{d}{dt} V\left\{i\,|t\right\} + \frac{d\,q(t)}{dt}\left[2q(t) + 1\right] = \tag{12}$$

$$= \frac{d}{dt} V\left\{i\,|t\right\} + \left[\lambda - \mu\, b(t)\right]\left[2q(t) + 1\right] =$$

$$= 2\left[\lambda - q(t)\,(\mu - \lambda)\right]$$

After some further rearrangements this relation is transformed into

$$\frac{d}{dt} V\left\{i \,|\, t\right\} = 2\lambda - 2\mu\, q(t)\, p_o(t) - \frac{d\, q(t)}{dt} \tag{13}$$

It is not difficult to perform an integration on (13) between 0 and Θ. We have already utilized the fact-in relation (7)- that the average queue size at the beginning of the green phase is equal to the sum of the average queue size at the beginning of the red phase and the average number of arrivals during the red phase. Closer inspection reveals that the random variable "queue size at $t = 0$" may be considered as the sum of two independent random variables: a) "queue size at $t = \Theta$" and b) "number of arrivals during the red phase"- the latter being Poisson distributed with parameter $\lambda\tau$ Hence not only relation (7) holds but the variances, too, are additive

$$V\left\{i \,|\, 0\right\} = V\left\{i \,|\, \Theta\right\} + \lambda\tau \tag{14}$$

Hence integration of (13) yields

$$\int_0^{\Theta} \frac{d}{dt} V\left\{i \,|\, t\right\} dt = -\lambda\tau = $$

$$= 2\lambda\Theta - 2\mu \int_0^{\Theta} q(t)\, p_o(t)\, dt + \lambda\tau \tag{15}$$

This in turn may be developed into the following very interesting relation

$$\frac{1}{\Theta} \int_0^{\Theta} q(t)\, p_o(t)\, dt = \frac{\lambda T}{\mu\Theta} = B \tag{16}$$

Equation (16) serves as starting point for the generation of some very useful formulas. We shall designate the average queue sizes within a red phase and within a green phase (these are quantities averaged over time) by q_r and q_g, respectively.

Clearly

$$q_r = \frac{q(\Theta) + q(0)}{2} = q(\Theta) + \frac{R}{2} = q(0) - \frac{R}{2} \tag{17}$$

The quantity q_g bears no such simple relationship to the average queue sizes at the special time epochs: $t = 0$ and $t = \Theta$. It is obtained by integration and on using equations (16) and (8) the following relation is generated

$$\begin{aligned}
q_g &= \frac{1}{\Theta}\int_0^{\Theta} q(t)\,dt = \frac{1}{\Theta}\int_0^{\Theta} q(t)\left[p_o(t) + b(t)\right] dt = \\
&= B + \frac{1}{\Theta}\int_0^{\Theta} q(t)\,b(t)\,dt = B + \frac{1}{\Theta}\int_0^{\Theta} q(t)\left(b - \frac{1}{\mu}\frac{d\,q(t)}{dt}\right) dt = \\
&= B + b\,q_g - \frac{1}{\mu\Theta}\int_0^{\Theta} q(t)\frac{d\,q(t)}{dt}\,dt = \\
&= B + b\,q_g - \frac{1}{2\mu\Theta}\left[q^2(\Theta) - q^2(0)\right] = \\
&= B + b\,q_g - \frac{1}{\mu\Theta}\,\frac{q(\Theta) + q(0)}{2}\left[q(\Theta) - q(0)\right] = \\
&= B + b\,q_g + \frac{\lambda\tau}{\mu\Theta}\,q_r = \\
&= B + b\,q_g + (B - b)\,q_r
\end{aligned} \tag{18}$$

Hence we obtain

$$q_g(1 - b) - q_r(B - b) = B \tag{19}$$

This can be rewritten as

$$\frac{q_g(1 - b) + q_r b}{1 + q_r} = B \tag{20}$$

The average queue(that is: averaged over time) is defined by

$$q = \frac{\Theta}{T} q_g + \frac{\tau}{T} q_r \tag{21}$$

If we multiply (21) by B and add (19) another interesting relationship is derived

$$B(1 + q) = (1 - b)q_g - (B - b)q_r + b\ q_g + (B - b)q_r = q_g \tag{22}$$

Equation (22) is of particular interest. Clearly each of the three quantities q, q_g, q_r is a function of b, B and R. The meaning of (22) is that once q_g is determined, q is uniquely fixed by the value of B. In other words: suppose the value of q_g has been determined by a triplet (b, B, R); if then B is held constant and values b and R are varied simultaneously in such a way so as to keep q_g unchanged then-at the same time-the total average queue, q, is held constant as well.

Judicious combination of (20) and (22) results in

$$q = \frac{b}{1-b} + \frac{B-b}{B(1-b)} \quad q_r \tag{23}$$

Relation (22) can be usefully employed in an argument establishing an inequality: For a given B the value of q cannot fall short of $\frac{B}{1-B}$; this lower bound is attained (in a "thought experiment", of course) if service availability and discontinuation are evened out in some sense and non-random pulsations are not permitted. Thus, for instance, we could visualize (again in a thought experiment) the service station to oscillate very rapidly between the two states such that during a prescribed proportion of time the station is active and during the complementary fraction no service can be rendered. This is equivalent to generating a simple queueing situation with arrival rate λ and effective service rate $\mu \frac{\Theta}{T}$. The associated queue length is then equal to $\frac{B}{1-B}$. If now the oscillations are less rapid (non-random) pulsations make additional contributions to average queue size. What we have established then is the following

$$q(b,B,R) \geq q(b,B,0) = \frac{B}{1-B} \tag{24}$$

Manipulation of (24) and insertion of (22) yields

$$q \geq B(1+q) = q_g \qquad (25)$$

Since q is a weighted average of q_g and q_r this inequality is equivalent to

$$q_g \leq q \leq q_r \qquad (26)$$

An alternative line of argument leading to (26) would be to utilize the property of $p_o(t)$ being a continuous and increasing function within the interval of customer discharge $(0,\Theta)$. Differentiation of both sides of (8) renders

$$\frac{d^2\, q(t)}{dt^2} = -\mu \frac{d\, b(t)}{dt} = +\mu \frac{d\, p_o(t)}{dt} \qquad (27)$$

Hence the second derivative of q(t) is positive in the interval under consideration and q(t) is then a strictly convex (and decreasing) function. However, within the interval (Θ,T) q(t) is-by virtue of Poisson arrivals and absence of service-a strictly linear function. The behavior of q(t) within the period (0,T) is schematically represented in Fig. 1. Inequality (26) is an immediate result.

Analogous arguments enable us to set bounds for the ratios $\frac{q_g}{q_r}$, $\frac{q_g}{q}$ and $\frac{q}{q_r}$. From equation (20) we obtain

$$\frac{B-b}{1-b} = 1 - \frac{1-B}{1-b} < \frac{q_g}{q_r} \leq 1 \qquad (28)$$

The lower bound can be approached as closely as desired by increasing

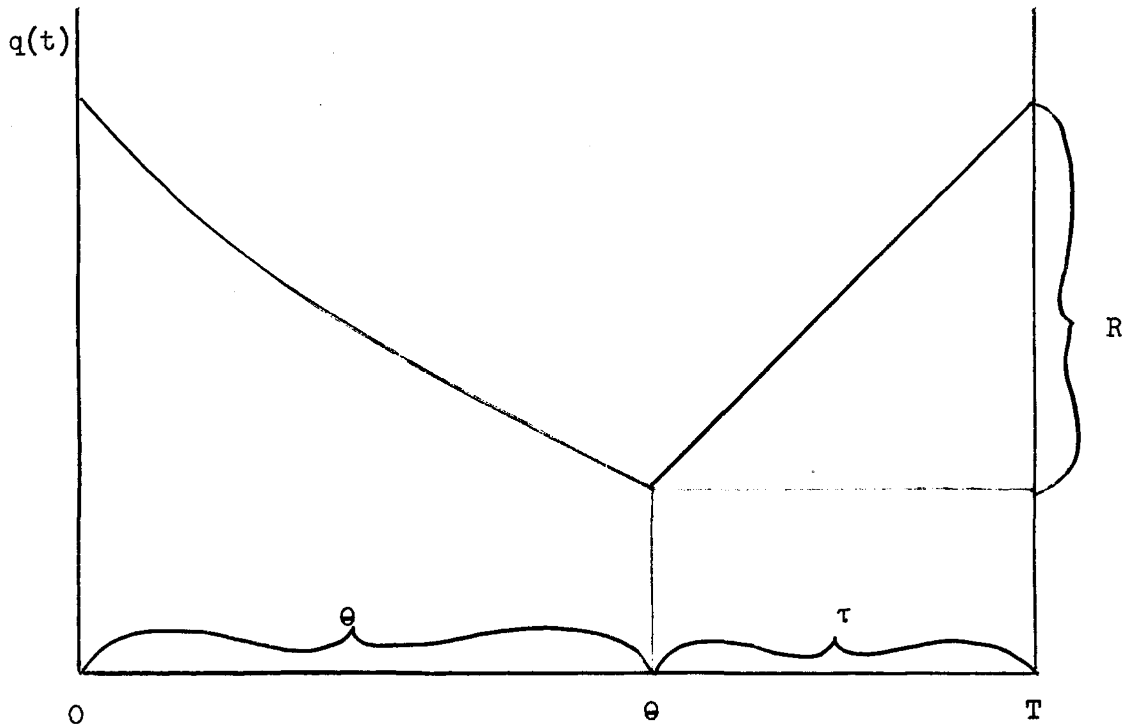

Fig. 1 Average queue length as a function of time within a cycle of duration T

the value of R.

In a similar fashion we can derive

$$B < \frac{q_g}{q} \leq 1 \tag{29}$$

and

$$\frac{B-b}{B-Bb} = 1 - (\frac{b}{1-b}) / (\frac{B}{1-B}) < \frac{q}{q_r} \leq 1 \tag{30}$$

Finally, in this section, an exact relation (that is, exact under the assumptions of our model) between b(0) and b(Θ) is presented. Clearly for zero customers to be present in the system at the beginning of a green phase the preceding green phase had to terminate with an empty station and no customers could have arrived in the interceding red phase. Hence the following must hold

$$p_o(0) = p_o(\Theta)\ e^{-\lambda\tau} \tag{31}$$

or, noting that $b(t) = 1 - p_o(t)$, equivalently

$$b(\Theta) = 1 - (1-b(0))\ e^{R} \tag{32}$$

III. APPROXIMATIONS TO AVERAGE QUEUE SIZE

The various relatively simple relationships presented in the previous section do not suffice for the generation of an exact formula of the average queue size in terms of b, B and R. In principle, of course, one could proceed in the following fashion. First a matrix of transition probabilities $\{p_{ij}^{(0,\Theta)}\}$ is established relating to states at the beginning and the end of a green phase; these transition probabilities are expressible by rather complicated but well-known formulas. Secondly the matrix of transition probabilities $\{p_{jk}^{(\Theta,0)}\}$ - relating to states at the beginning and the end of a red phase - is set up; these transition probabilities are of rather simple character since changes (k-j) follow a Poisson distribution. Thirdly, the two matrices are multiplied - in the above order, say - and that probability vector is sought which represents the (existing and unique) fixed point of the matrix product. This is the set $\{p_i(0)\}$. All quantities of interest in this study can be derived from this set.

Unfortunately, it is not practicable to carry out this programme. The entries $p_{ij}^{(0,\Theta)}$ are of functional complexity and the number of rows and columns of the matrix is infinite. Hence it appears that the problem of evaluating the average queue size has to be approached on other avenues

Two modes of approximate evaluation will be presented here. One pertains to the case of rapid pulsations (that is: frequent interruptions and renewals of service) whereas the second mode is most useful when the oscillations are rather slow.

Consider the function b(t) within the interval (0,Θ). Clearly the average "busy-ness" of the service station decreases in time. However, the rate of decrease diminishes, that is the second derivative of b(t) is positive and extrapolation of $b(t)$ beyond Θ ("forgetting" to close down the service station at time Θ, as it were) would let this function approach the value $\frac{\lambda}{\mu}$ (=b) in an asymptotic fashion. Hence a reasonable approximation of this function is given by

$$b(t) = b + (b(0) - b)\, e^{-\zeta t} \tag{33}$$

where ζ is a constant which will be determined by imposing two constraints:

(1) At times $t = 0$ and $t = \Theta$ the values of the function $b(t)$ must be consistent with the (exact) condition expressed by (32).

(2) The average value (in time) of $b(t)$ within interval $(0,\Theta)$ must equal B.

Hence we obtain

$$\frac{1}{\Theta} \int_0^{\Theta} b(t)\, dt = b + \frac{1}{\zeta\Theta} (b(0) - b)\, (1 - e^{-\zeta\Theta}) = B \tag{34}$$

or, alternatively

$$Z(B-b) = (b(0) - b)\, (1 - e^{-Z}) \tag{35}$$

where the dimensionless constant Z is defined as

$$Z = \zeta\Theta \tag{36}$$

This quantity Z - a constant for any given system - depends on b, B, and R and can be determined in the following fashion: we examine the value of $b(t)$ at time $(t = \Theta)$ and set it equal to that obtained in equation (32)

$$b(\Theta) = b + (b(0) - b)e^{-Z} = 1 - (1 - b(0))e^{R} \tag{37}$$

This can be rewritten as

$$1 - b(\Theta) = (1 - b)(1 - e^{-Z}) + (1 - b(0))e^{-Z} = (1 - b(0))e^{R} \tag{38}$$

and further rearrangement yields

$$p_o(0) = 1 - b(0) = \frac{(1 - b)(1 - e^{-Z})}{e^{R} - e^{-Z}} \tag{39}$$

Now if equation (39) is combined with (35) this results in

$$\begin{aligned} Z(B - b) &= \left[(1 - b) - (1 - b(0) \right] (1 - e^{-Z}) = \\ &= (1 - b(0))(e^{R} - e^{-Z}) - (1 - b(0))(1 - e^{-Z}) = \\ &= (1 - b(0))(e^{R} - 1) \end{aligned} \tag{40}$$

However further use can be made of (32) in order to modify relation (40) into

$$\begin{aligned} Z(B - b) &= (1 - b(0))e^{R} - (1 - b(0)) = (1 - b(\Theta)) - (1 - b(0)) = \\ &= b(0) - b(\Theta) \end{aligned} \tag{41}$$

While the various equations presented above exhibit a number of simple relationships between quantities of interest no single equation has been put forward yet which describes Z as a function - implicit, at least - of b, B and R. Such an equation can be derived by manipulating (39) and (40)

$$1 - b(0) = \frac{(1 - b)(1 - e^{-Z})}{e^{R} - e^{-Z}} = \frac{Z(B - b)}{e^{R} - 1} \qquad (42)$$

This latter equation may be transformed into

$$\frac{1 - b}{B - b} \frac{1}{Z} = \frac{1}{e^{R} - 1} + \frac{1}{1 - e^{-Z}} \qquad (43)$$

A convenient form for numerical computations is now obtained

$$1 - e^{-Z} = \frac{Z}{\frac{1 - b}{B - b} - Z \frac{1}{e^{R} - 1}} \qquad (44)$$

It is easy to verify that one and only one solution of (44) exists with the property $Z > 0$. Also an upper bound for the solution is immediately established

$$Z < (1 - e^{-R}) \frac{1 - b}{B - b} \qquad (45)$$

on observing that the numerator of the r.h.s. of (44) falls short of the denominator. Inequality (45) may be advantageously employed for that stage of computational work at which a first approximation to the solution is sought.

It is both useful and interesting to describe the behavior of Z in dependence on R with quantities B and b fixed but arbitrary (and, of course, consistent with inequality (2)). Since Z is an implicit function of R the necessary derivations are rather long and some results only will be reported here:

a) Z is a concave and monotone increasing function of $R(\geq 0)$.

b) If R is set equal to 0, Z attains the value 0 as well.

c) In the neighborhood of the origin a linear Mac Laurin expansion yields

$$Z \gtrsim \frac{1 - B}{B - b} \quad R \tag{46}$$

Clearly the coefficient of R is positive and unbounded.

d) With increasing argument R the function Z tends to an asymptotic value, Z_∞ say, which is the solution of

$$\frac{B - b}{1 - b} \; Z_\infty + e^{-Z_\infty} = 1 \tag{47}$$

This latter function (Z_∞ dependent on the coefficient $\frac{B-b}{1-b}$) was tabulated by Barton et al (1960). For finite R, the value of Z must fall short of Z_∞; hence we have

$$Z < Z_\infty = (1 - e^{-Z_\infty}) \frac{1-b}{B-b} \tag{48}$$

which - under some easily specified circumstances - may be more useful than (45).

Let the case of (sufficiently) small R and Z be considered. We may rewrite (35) as

$$b(0) = b + \frac{Z}{1 - e^{-Z}} (B - b) \doteq b + (1 + \frac{Z}{2})(B-b) \qquad (49)$$

Combination of (49) and (46) yields

$$b(0) \doteq b + (B-b) + \frac{1}{2} \frac{1-B}{B-b} (B-b) R = B + (1-B) \frac{R}{2} \qquad (50)$$

If now the value of b(0) - as obtained in (50) - is inserted in (32) an approximate formula for $b(\theta)$ is derived

$$b(\theta) \doteq 1 - (1-B) (1 + \frac{R}{2}) = B - (1-B)\frac{R}{2} \qquad (51)$$

It is recalled that relations (50) and (51) hold for sufficiently small values of R only. Under such circumstances the value of b(t) decreases linearly in the interval $(0,\theta)$ and "jumps back" to the original value b(0) during the interval (θ,T). Inspection of (50) immediately clarifies the meaning of the phrase "sufficiently small R". The average increase of b with respect to its (time) average value B in $(0,\theta)$ has to be very small and obviously under no circumstances may $\frac{R}{2}$ exceed the value one. Hence we may sum up these two conditions as

$$\frac{R}{2} \ll \frac{B}{1-B} \qquad (52)$$

and

$$\frac{R}{2} < 1 \qquad (53)$$

The physical significance of (52) is that the average increase and decrease of queue size (to wit: $\frac{R}{2}$) during a pulsation has to be small with respect to average queue size under condtions of extremely rapid oscillations (to wit: $\frac{B}{1-B}$); the significance of (53) is that even if the average queue size were to be very large, $\frac{R}{2}$ is not permitted to take a value larger than 1.

Our next step will be the evaluation of the various average queue lengths of interest. Within the interval $(0,\Theta)$ the average queue size at some arbitrary (but fixed) time t can be derived from (8), (33) and (35)

$$
\begin{aligned}
q(t) &= q(0) + \int_0^t \frac{d\,q(t')}{dt'}\,dt' = \\
&= q(0) + \int_0^t \left[\lambda - \mu\, b(t')\right] dt' = \\
&\ q(0) + \int_0^t \left\{\lambda - \mu\left[b + (b(0) - b)e^{-\zeta t'}\right]\right\} dt' = \\
&= q(0) - (b(0) - b)\,\mu \int_0^t e^{-\zeta t'}\,dt' = q(0) - \frac{\mu}{\zeta}(b(0) - b)(1 - e^{-\zeta t}) = \\
&= q(0) - \frac{R}{Z(B-b)}\,(b(0) - b)\,(1 - e^{-\zeta t}) = \\
&= q(0) - R\frac{1 - e^{-\zeta t}}{1 - e^{-Z}}
\end{aligned}
\qquad (54)
$$

This formula yields the proper result (7), if the particular value $t = \Theta$ is chosen.

The average value (in time) of the queue, q_g , during a green phase is then obtained as

$$q_g = \frac{1}{\Theta} \int_0^{\Theta} q(t')dt' = q(0) - \frac{R}{1-e^{-Z}} + \frac{R}{(1-e^{-Z})\Theta} \int_0^{\Theta} e^{-\zeta t} dt =$$

$$= q(0) - \frac{R}{1-e^{-Z}} + \frac{R}{Z} = q(0) - R(\frac{1}{1-e^{-Z}} - \frac{1}{Z}) \tag{55}$$

The corresponding quantity, q_r, for the red phase is presented in equation (17).

It is useful to consider the coefficient of R in (55) as a function, ψ say, of Z.

$$\psi = \frac{1}{1-e^{-Z}} - \frac{1}{Z} \tag{56}$$

It is not difficult to verify that ψ is a monotonically increasing function of Z such that $\psi(0) = \frac{1}{2}$, $(\frac{d\psi}{dZ})_{Z=0} = \frac{1}{12}$, $(\frac{d^2\psi}{dZ^2})_{Z=0} = 0$ and $\psi(\infty) = 1$.

The latter value of ψ is of little consequence since Z itself is a bounded function of R.

The two equations

$$q_g = q(0) - \psi R \tag{57}$$

and

$$q_r = q(0) - \frac{1}{2} R \tag{58}$$

are now combined with (19)

$$(q(0) - \psi R)(1-b) - (q(0) - \tfrac{1}{2} R)(B-b) =$$

$$= q(0)(1-B) - R\left[\psi(1-b) - \tfrac{1}{2}(B-b)\right] = B \tag{59}$$

Hence the average queue length at the beginning of the green phase equals

$$q(0) = \frac{B}{1-B} + \frac{\psi(1-b) - \frac{1}{2}(B-b)}{1-B} R \tag{60}$$

and the other average quantities of interest are immediately obtained

$$q(\theta) = \frac{B}{1-B} - \frac{\frac{1}{2}\left[(1-B) - (1-b)(2\psi - 1)\right]}{1-B} R \tag{61}$$

$$q_r = \frac{B}{1-B} + \frac{(\psi - \frac{1}{2})(1-b)}{1-B} R \tag{62}$$

$$q_g = \frac{B}{1-B} + \frac{(\psi - \frac{1}{2})(B-b)}{1-B} R \tag{63}$$

$$q = \frac{B}{1-B} + \frac{(\psi - \frac{1}{2})}{1-B} R \tag{64}$$

If R is sufficiently small we can approximate the function ψ in the neighborhood of 0 in the following fashion

$$\psi = \frac{1}{2} + \frac{Z}{12} = \frac{1}{2} + \frac{1-B}{B-b}\frac{R}{12} \tag{65}$$

Hence if oscillations are very rapid - that is close-down and reopening of the station occurs rather frequently - the five average queue sizes of interest may be (approximately) evaluated as

$$q(0) \doteq \frac{B}{1-B} + \frac{R}{2} + \frac{1-b}{B-b}\frac{R^2}{12} \tag{66}$$

$$q(\theta) \doteq \frac{B}{1-B} - \frac{R}{2} + \frac{1-b}{B-b}\frac{R^2}{12} \tag{67}$$

$$q_r \doteq \frac{B}{1-B} + \frac{1-b}{B-b}\frac{R^2}{12} \tag{68}$$

$$q_g \doteq \frac{B}{1-B} + \frac{R^2}{12} \tag{69}$$

$$q \doteq \frac{B}{1-B} + \frac{1}{B}\frac{R^2}{12} \tag{70}$$

Equation (70) is of particular interest since a special feature of the deterministic interruption pattern is demonstrated. In the neighborhood of the origin the average queue does not increase linearly with R (which - we recollect - is an inverse measure of oscillation rapidity). The derivative of the average queue with respect to R equals zero at the location R = 0. In cruder terms: if oscillations possess "infinite frequency" the average queue size equals B/(1-B); if now the frequency is made finite but still very large then - to a first approximation - no change in average queue size can be recorded. This is strikingly different from the case of Poisson station breakdowns (instead of deterministic breakdowns) which was dealt with by Gaver (1962) and Avi-Itzhak and Naor (1963). The analogous formula for the average queue size in this latter case can be derived from equation (24) in Avi-Itzhak's and Naor's communication as

$$q = \frac{B}{1-B} + \frac{(B-b)}{B(1-B)} \frac{R}{2} \tag{71}$$

This formula holds rigorously over the whole domain of possible values of B, b and R and therefore, also in the neighborhood of R = 0.

It is to be expected that the assumption of linear change in b (equation (50) and (51)) should yield proper results (that is: correct formulas for average queue sizes) for sufficiently small values of R and all permissible values of b and B. Furthermore approximation (33) should be associated with numerically good results for "medium" values of R. This is the case since (33) is a specification - negative exponential, that is - of a function b(t) of which we know a) the ratio $\frac{1-b(\Theta)}{1-b(0)}$; b) the integral $\int_0^\Theta b(t)\, dt$;

and c) that it is convex. Clearly in practice the negative exponential specification can differ only slightly from any other reasonable alternative specifications - precisely because of the ratio and integral constraints. Hence, as long as 0 and Θ are not too far apart the shape of b(t) is almost invariant with respect to the set of alternative specifications, and perforce the same statement holds good for the shape of q(t) as well as for q_g, q_r and q. If, however, 0 and Θ are far apart, (that is if the case of large R is considered) we can no longer expect that the true shape of b(t) and the one derived from the exponential decay assumption will be close to each other. Indeed it is possible to demonstrate that for a sufficiently large R a patently wrong value of $q(\Theta)$ is obtained. This becomes obvious on examination of (61): The coefficient $\frac{\frac{1}{2}[(1-B) - (1-b)(2\psi - 1)]}{1 - B}$ tends to a (positive) constant when R increases. Hence for sufficiently large values of R the average queue size at the end of the green phase falls short of any arbitrarily selected finite bound—even a negative one. This is not in accord with physical reality and therefore the queueing formulas derived up to now are valid for a limited domain only of possible R-values. There exists a value of R, R_1 say, beyond which formula (61) must yield non-feasible values of $q(\Theta)$. It is clear that under <u>all</u> circumstances the following inequalities must hold

$$\frac{b}{1-b} \leq q(\Theta) \leq \frac{B}{1-B} \tag{72}$$

While mathematical analysis would render (72) as strict inequalities the equality sign was included for practical considerations. $q(\Theta)$ takes the value on the r.h.s. in the case of extremely rapid oscillations (R very small); it is equal to the ℓ.h.s. of (72) if the oscillations are very infrequent (R very large). The set of equations (60)-(64) is certainly not applicable if equation (61) does not obey the left inequality (72).

The above considerations regarding the general behavior of $q(\Theta)$ serve as a basis of approximating the various average queue sizes by a new approach. $q(\Theta)$ is a monotonically decreasing function of R with upper bound

$$q(\Theta | R \rightarrow 0) \doteq \frac{B}{1-B} \tag{73}$$

and lower bound

$$q(\Theta \mid R \to \infty) \doteq \frac{b}{1-b} \qquad (74)$$

Our present programme is to find a suitable monotonically increasing (in dependence on R) function $\Phi(R, B, b)$ possessing values

$$\Phi(0, B, b) = 0 \qquad (75)$$

and

$$\Phi(\infty, B, b) = 1 \qquad (76)$$

independent of B and b.

For notational convenience define a quantity D as

$$D = \frac{B}{1-B} - \frac{b}{1-b} = \frac{1}{1-B} - \frac{1}{1-b} = \frac{B-b}{(1-B)(1-b)} \qquad (77)$$

The average queue at the end of the green phase, $q(\Theta)$, is then determinable by

$$q(\Theta) = \frac{B}{1-B} - \Phi D \qquad (78)$$

At the origin the value of the derivative of $q(\Theta)$ with respect to R is known (e.g. from (67)) to be $-\frac{1}{2}$.

Hence one criterion for the suitability of function Φ should be

$$\left(\frac{\delta q(\Theta)}{\delta R}\right)_{R=0} = -D\left(\frac{\delta \Phi}{\delta R}\right)_{R=0} = -\frac{1}{2} \qquad (79)$$

If (as an example) an exponential form is selected for Φ criterion (79) prescribes that

$$\Phi(R, B, b) = 1 - e^{-\frac{R}{2D}} \tag{80}$$

This is one of several "reasonable" choices for the functional form of Φ; in the following, whenever Φ appears in an equation, specialization (80) is not necessarily assumed. However it should be emphasized that the various alternative forms for Φ typically render similar shapes. Furthermore, within that part of the domain of R in which alternative functions Φ_1 and Φ_2 may attain truly distinct values the relative contribution of Φ to the average queue size q is not expected to be appreciable.

Utilizing the various exact relations obtained before we derive the following equations for average queue lengths of interest.

$$q(0) = \frac{B}{1-B} - \Phi D + R \tag{81}$$

$$q_r = \frac{B}{1-B} - \Phi D + \frac{R}{2} \tag{82}$$

$$q_g = \frac{B}{1-B} + \left(\frac{R}{2} - \Phi D\right) \frac{(B-b)}{1-b} \tag{83}$$

$$q = \frac{B}{1-B} + \left(\frac{R}{2} - \Phi D\right) \frac{(B-b)}{B(1-b)} \tag{84}$$

For sufficiently large R the function Φ practically attains value 1. In this case equation (84) tends to

$$q \doteq \frac{b}{1-b} + \frac{b}{1-b} \cdot \frac{(B-b)}{B(1-b)} + \frac{R}{2}\frac{(B-b)}{B(1-b)} =$$

$$= \frac{b(2B - b - Bb)}{B(1 - b)^2} + \frac{R}{2}\frac{(B-b)}{B(1-b)} \tag{85}$$

It is interesting to compare (85) with equation (71) which represents the exact solution of a model with Poisson arrivals, exponentially distributed service times, Poisson breakdowns of the service station and constant shut-down times.

The present approximative procedure for the evaluation of average queue size has been constructed in order to obtain results for large R - a region where the first approximative procedure failed. However the "suitability" criterion (79) which was required for the construction of function Φ ascribes the proper first derivative (with respect to R, at location R = 0) to all queue sizes. Hence a linear Mac Laurin expansion of queue length certainly yields correct results and the present approximation is valid not only for large values of R but also for very small values of that variable. Should we require -as an additional suitability criterion- that the function's second derivative too (at location R = 0) should have the correct value then it is frequently possible to obtain a "formula" which represents queue size very well over the whole domain of R. However typically not many simple functional forms for Φ are at our disposition.

As an additional example for a suitable function Φ we select

$$\Phi = 1 - e^{-\frac{R}{2D}} \cosh \frac{R}{2D}\sqrt{\frac{B-\frac{1}{3}}{1 - B}} \tag{86}$$

This function (which, incidentally, is meaningful only if $\frac{1}{3} < B < \frac{2}{3}$) renders good approximations when introduced in equation (84) and its associates. However frequently the application of rather complicated functions (such as (86)) is superfluous and one can arrive at useful (approximate) results employing graphical methods. The methods revolve around the following pivots: The value of the function q at $R = 0$ is known as well as the function's first two derivatives; the straight line (85) is approached asymptotically when R increases.

The model described, developed, and discussed in this communication may have direct applicability in some queueing situations and formulas derived here may possibly be employed as they stand. However the main purpose of this study was not to provide precise formulas to a rather narrow class of situations described in our model: a) Poisson arrivals; b) service time exponentially distributed; c) regular (that is cyclic) service interruptions and renewals. Rather our aims were the following: 1) It has been shown (or at least claimed) in the literature that many results are (almost) invariant under "moderate" changes of assumptions. If indeed such robustness exists the optimal approach to analysis is to make the simplest assumptions and use standard methods for the derivation of the desired results. Letting up the differential equations pertaining to the birth and death process appears to be a simple proceedure. 2) Even if the desired robustness of results is not a characteristic of a large segment of "situation space" the analysis presented here may be utilized for bounding the proper solutions and -at a later phase- arriving at approximations for the concrete situation at hand. 3) In a wider class of situations where quasi-invariance of solutions no longer prevails and even bounding by the special class under discussion is not possible, the methods employed here and types of results attained may serve as guide lines. Thus, for instance, the type of exact relation as presented in (20) and the ensuing approximation (85) of the average queue length are in all likelihood obtainable in other situations either more general or differently specified.

We have made no attempt to develop optimization procedures. To do this further assumptions relating both to costs and to feasible allocations of service to various customer streams must be made . Some work is in progress in this subject area.

REFERENCES

B. Avi-Itzhak and P. Naor, "Some Queueing Problems with the Service Station Subject to Breakdown", Opns. Res. 11, 303-320 (1963).

D. E. Barton, F. N. David and M. Merrington, "Tables for the Solution of the Exponential Equation, exp(-a) + ka=1", Biometrika 47, 439-445 (1960).

D. J. Buckley and R. C. Wheeler, "Some Results for Fixed-Time Traffic Signals", J. Roy. Statist. Soc. B 26, 133-140 (1964).

D. P. Gaver, "A Waiting Line with Interrupted Service, including Priorities", J. Roy. Statist. Soc. B 24, 73-90 (1962).

G. F. Newell, "Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light", SIAM Review 7, 223-239 (1965).

M. Yadin and P. Naor, "Queueing Systems with a Removable Service Station", Opns. Res. Quart. 14, 393-405 (1963).