

Equilibrium Properties of the $M/G/1$ Queue

Søren Asmussen

Institute of Mathematical Statistics, University of Copenhagen, 5 Universitetsparken,
DK-2100 Copenhagen Ø, Denmark

Summary. Various aspects of the equilibrium $M/G/1$ queue at large values are studied subject to a condition on the service time distribution closely related to the tail to decrease exponentially fast. A simple case considered is the supplementary variables (age and residual life of the current service period), the distribution of which conditioned upon queue length n is shown to have a limit as $n \rightarrow \infty$. Similar results hold when conditioning upon large virtual waiting times. More generally, a number of results are given which describe the input and output streams prior to large values e.g. in the sense of weak convergence of the associated point processes and incremental processes. Typically, the behaviour is shown to be that of a different transient $M/G/1$ queueing model with a certain stochastically larger service time distribution and a larger arrival intensity. The basis of the asymptotic results is a geometrical approximation for the tail of the equilibrium queue length distribution, pointed out here for the $GI/G/1$ queue as well.

1. Introduction

We consider the $M/G/1$ queue and let α denote the arrival intensity, G the service time distribution and

$$\rho = \alpha v \left(v = v^1 = \int_0^{\infty} x dG(x) = \int_0^{\infty} (1 - G(x)) dx \right)$$

the traffic intensity. We assume throughout $\rho < 1$. It is then well-known that a number of quantities associated with the queueing process at time t converge in distribution as $t \rightarrow \infty$. E.g. this holds for the queue size Q_t , the virtual waiting time v_t (residual amount of work in the system), the age A_t of the present service time and the residual service time B_t (the A_t , B_t are defective, being defined on $\{Q_t > 0\}$ only). The limiting distributions are the *equilibrium*

distributions (e.d.) or steady states and a standard point of view in queueing theory is to measure the characteristics of the system by means of the e.d., cf. e.g. Cox and Smith (1961). This motivates a detailed study of the process at equilibrium.

To facilitate notation, we let P_e, E_e refer to the equilibrium case so that e.g.

$$\pi_n = P_e(Q_t = n) = \lim_{s \rightarrow \infty} P(Q_s = n).$$

It is well-known and easily proved, cf. Miller (1972), that P_e can also be interpreted as the probability law governing a strictly stationary process.

The investigations of the present paper start in Sect. 2 by a discussion of the estimate

$$\pi_n \simeq c \delta^{-n} \quad \text{as } n \rightarrow \infty \tag{1.1}$$

for the tail of the e.d. of the queue size. This formula should be compared to the exact geometric form $\pi_n = (1 - \rho) \rho^n$ in the $M/M/1$ case and explicit expressions for special cases as $M/E_k/1$ and $M/D/1$ as given e.g. in Saaty (1961) Chap. 6. The conditions for (1.1) are the same as the main ones for the rest of the paper, viz. the existence of a solution $\gamma > 0$ to the equation

$$\alpha \int_0^\infty e^{\gamma x} (1 - G(x)) dx = 1 \tag{1.2}$$

with the additional property

$$\kappa = \int_0^\infty x e^{\gamma x} (1 - G(x)) dx < \infty. \tag{1.3}$$

The connection between $\delta, \gamma, c, \kappa$ is given by

$$\alpha(\delta - 1) = \gamma, \quad c = \frac{1 - \rho}{\alpha^2 \kappa}. \tag{1.4}$$

The traditional approach to (1.1) is based upon transform methods (see e.g. Gaver (1959), Le Gall (1962)) and requires some additional analyticity conditions. The proof pointed out here is quite simple and produces also certain generalizations, e.g. a $GI/G/1$ analogue. Condition (1.2) is certainly a restriction on G (whereas (1.3) is only slightly stronger). However, it will follow (e.g. from the last part of Theorem 3.3) that (1.2) is necessary for main parts of the paper. Though not coming up very directly in that form in the present paper, it is of some interest to note that (1.2) is equivalent to $E e^{\gamma(S-T)} = 1$ (with S, T independent with $P(S \leq s) = G(s), P(T > t) = e^{-\alpha t}$), i.e. to the existence of the associated distribution in the sense of Feller (1971) pp. 406-407. For other examples of the relevance in queueing theory of this and related conditions see e.g. Kingman (1964), Cohen (1968), (1969), (1973b), Iglehart (1972) and Asmussen (1980), (1981).

In Sect. 3, we study the joint e.d. of Q_t and A_t, B_t , i.e.

$$U_n(\xi) = P_e(A_t \leq \xi | Q_t = n), \quad V_n(\xi) = P_e(B_t \leq \xi | Q_t = n).$$

The interest in these quantities arise largely from the role of either of A_t, B_t as *supplementary variables*, who in conjunction with Q_t form the minimal information needed to make the process Markovian. In fact, the e.d. of any the above quantities are derivable from the π_n, U_n . As examples, note the formulae

$$1 - V_n(\xi) = \int_0^\infty \frac{1 - G(x + \xi)}{1 - G(x)} dU_n(x), \tag{1.5}$$

$$P_e(v_t \leq \xi) = \pi_0 + \sum_{n=1}^\infty \pi_n V_n * G^{*(n-1)}(\xi). \tag{1.6}$$

However, discussions like those of Gnedenko and Kovalenko (1968) pp.157-160, Cohen (1969) II.6.2 or Hokstad (1975) place little emphasis on the supplementary variables *per se*. The main result in that direction seems to be that (up to the defect) the marginal e.d. of A_t, B_t coincide with the common e.d. of the backwards and forwards recurrence times in a renewal process with inter-arrival distribution G . That is (cf. Cohen (1976) Chap.I and Feller (1971) Chap. XI),

$$P_e(A_t \leq \xi) = \sum_{n=1}^\infty \pi_n U_n(\xi) = P_e(B_t \leq \xi) = \sum_{n=1}^\infty \pi_n V_n(\xi) = \frac{\rho}{v} \int_0^\xi (1 - G(x)) dx. \tag{1.7}$$

Section 3 starts off by computing the U_n, V_n by means of the embedded Markov chain and the basic formula

$$E_e W_t = \frac{1}{Ec} E \int_0^c W_s ds \tag{1.8}$$

(with c the busy cycle) for functionals W_t of the process which are regenerative w.r.t. the renewal process formed by the successive ends of busy cycles and satisfies some path conditions automatic in all cases considered in the present paper. Cf. Smith (1955), Feller (1971) Chap. XI, Miller (1972) and Cohen (1976). The expressions obtained are explicit, though maybe not as simple as one could have hoped from (1.7). However, as one of our main results we show that U_n, V_n have weak limits as $n \rightarrow \infty$. A corollary is a similar behaviour of the length $C_t = A_t + B_t$ of the current service period.

These results raise the more general question of the behaviour of the process prior to the large value $Q_t = n$. This problem is the topic of Sects. 4 and 5, where we obtain a number of limit results describing the entire past, e.g. in terms of the input and output point processes or the incremental processes. In addition to the queue length, we also consider the virtual waiting time, motivated, of course, from the fact that the virtual waiting time in many applications is a more relevant measure of the amount of congestion than the queue length. The results obtained seem to be of a genuinely new type (except that in the $M/M/1$ case there is a close relation to the well-known time reversibility) and a more detailed statement is deferred to the body of the paper. A typical result is, however, that the input and output point processes prior to a large virtual waiting time behave like two independent stationary point processes, which are, respectively, a Poisson process with intensity $\tilde{\alpha} = \alpha \delta$ (rather than α)

and a renewal process with interarrival distribution $d\tilde{G}(x) = \delta^{-1} e^{\gamma x} dG(x)$ (rather than $dG(x)$). Results of a similar spirit (but rather different framework) are further exploited in Asmussen (1980).

2. The Imbedded Markov Chain and the Queue Length

Unless when considering the equilibrium situation, we suppose that $Q_0 = 0$. Define $\tau(0) = 0$, $\tau(n)$ as the instant where the n^{th} service period is completed. It is then well-known, that $\{X_n\} = \{Q_{\tau(n)}\}$ is a aperiodic positive recurrent Markov chain, the e.d. of which coincides with the e.d. $\{\pi_n\}$ of the queue length Q_t . Let $p_n(t)$ be the probability of n arrivals in an interval of length t , $q_n = E p_n(S)$ the probability of n arrivals during a service period S , i.e.

$$p_n(t) = e^{-\alpha t} \frac{(\alpha t)^n}{n!}, \quad q_n = \int_0^\infty p_n(t) dG(t).$$

Also let $s_n = 1 - q_0 - \dots - q_n$. For future reference, we state

2.1. Lemma. *The expected amount of time during a service period where n customers have arrived since the start of the period is*

$$E \int_0^S I(n \text{ arrivals in } [0, t]) dt = \int_0^\infty p_n(t) (1 - G(t)) dt = \frac{S_n}{\alpha}. \tag{2.1}$$

Indeed, the first equality in (2.1) follows immediately and the second upon integration by parts, noting that

$$\frac{d}{dt} (1 - p_0(t) - \dots - p_n(t)) = \alpha p_n(t).$$

Turning next to the discussion of (1.1), various approaches apply. E.g. (1.1) could be derived from the well-known relation (e.g. Prabhu (1965) pp.127) between $\{\pi_n\}$ and the maximum of the left-continuous random walk with generic increments Y_k satisfying $P(Y_k = n - 1) = \pi_n$, in conjunction with Feller (1971) pp.411. Note in this connection that it is easy to see that (1.2) is equivalent to the defining equation $E \delta^{Y_n} = 1$ for the corresponding associated distribution. However, we shall take the opportunity to point out a $GI/G/1$ analogue of (1.1) as well as to sketch certain generalizations.

Consider thus the $GI/G/1$ queue with distribution F of the interarrival time T , suppose that F is non-lattice, that $ES < ET$ and that for some $\gamma > 0$

$$E e^{\gamma(S-T)} = 1, \quad E |S - T| e^{\gamma(S-T)} < \infty. \tag{2.2}$$

Then the e.d. of the actual waiting time W_n is well-known to be that of the maximum of a random walk with generic increments distributed as $S - T$ and appealing once more to Feller (1971) pp.411 we may deduce that for some $\tilde{d} < \infty$

$$P_e(W_n > w) \cong \tilde{d} e^{-\gamma w}. \tag{2.3}$$

Using the well-known relation between the e.d. of W_n and v_t (Cohen (1969) pp. 297) it follows easily from (2.3) that, with $\delta = \int_0^\infty e^{\gamma u} dG(u) = \left(\int_0^\infty e^{-\gamma u} dF(u) \right)^{-1}$,

$$P_e(v_t > y) \cong de^{-\gamma y}, \quad d = \bar{d} \rho(\delta - 1) / \gamma v \tag{2.4}$$

and relating v_t and Q_t by means of Cohen (1969) pp. 302 we finally get from (2.4)

$$P_e(Q_t \geq n) = \int_0^\infty P_e(v_t > x) dF^{*(n-1)}(x) \cong d\delta^{-(n-1)}, \tag{2.5}$$

$$P_e(Q_t = n) = c\delta^{-n}, \quad c = d(\delta - 1). \tag{2.6}$$

Feller's expression for \bar{d} can be rewritten in various ways (alternatively \bar{d} can be computed by applying an Abelian argument to Spitzer's identity). In the $M/G/1$ case one in fact gets $\bar{d} = d = (1 - \rho) / \alpha \gamma \kappa$ and Feller (1971) pp. 377-378 gives here an alternate derivation of (2.4) by means of the renewal equation (e.g. Cohen (1976) pp. 35) satisfied by $P_e(v_t \leq x)$.

If in the $GI/G/1$ case one replaces (2.2) by conditions on regular variation of the tail of G , the expressions in Cohen (1973a) for $P_e(W_n > w)$ apply in a rather similar manner to produce approximations for $P_e(v_t > x)$, $P_e(Q_t \geq n)$. We omit the details.

We finally refer to Asmussen (1981) for a non-equilibrium version of (1.1), derived in a somewhat more general setting.

3. The Supplementary Variables

We start off by computing U_n, V_n . Define $\pi_1^* = \pi_0 + \pi_1, \pi_n^* = \pi_n, n > 1$.

3.1. Proposition. *The distributions $U_1, U_2, \dots, V_1, V_2, \dots$ have densities u_n, v_n given by*

$$\pi_n u_n(\xi) = \alpha(1 - G(\xi)) \sum_{m=1}^n \pi_m^* p_{n-m}(\xi), \tag{3.1}$$

$$\pi_n v_n(\xi) = \alpha \int_{\xi}^{\infty} \sum_{m=1}^n \pi_m^* p_{n-m}(x - \xi) dG(x). \tag{3.2}$$

Proof. We let $W_t = I(Q_t = n, A_t \leq \xi)$ in (1.8), recall the imbedded Markov chain defined in Sect. 2 and write

$$\int_0^c W_s ds = \sum_{k=0}^{k-1} J_k, \quad \text{with } J_k = \int_{\tau(k)}^{\tau(k+1)} W_s ds \tag{3.3}$$

and k the number of customers served during the busy cycle, i.e. the time of the first return of $\{X_n\}$ to 0. Now suppose first $k \geq 1$. Then a new service period starts at time $\tau(k)$ and $X_k = m$ (say) customers are present. Thus in order for the event $\{A_t \leq \xi, Q_t = n, \tau(k) \leq t < \tau(k+1)\}$ to occur, $n - m$ new customers

must have arrived within $\bar{u} = t - \tau(k)$ time units, the service period must not have terminated and we must have $u \leq \xi$. Conditioning upon \mathcal{H}_k (the σ -algebra containing all relevant information up to time $\tau(k)$) shows that

$$EJ_k I(1 \leq k < \mathbf{k}, X_k = m) = P(1 \leq k < \mathbf{k}, X_k = m) \int_0^\xi p_{n-m}(u) (1 - G(u)) du.$$

For $k=0$ an exponentially distributed period elapses before service starts and a slight modification of the argument yields

$$EJ_0 = \int_0^\xi p_{n-1}(u) (1 - G(u)) du.$$

Thus, combining these expressions by (1.8), (3.3), $E\mathbf{c} = 1/\alpha(1 - \rho)$ and the fact that the expected number of visits of X_n to m before \mathbf{k} is π_m/π_0 , it follows that $P_e(Q_t = n, A_t \leq \xi)$ equals

$$\begin{aligned} & \alpha(1 - \rho) \int_0^\xi \left\{ p_{n-1}(u) + \sum_{m=1}^n E \sum_{k=1}^{\mathbf{k}-1} I(X_k = m) p_{n-m}(u) \right\} (1 - G(u)) du \\ &= \alpha(1 - \rho) \int_0^\xi \left\{ \left(1 + \frac{\pi_1}{\pi_0} \right) p_{n-1}(u) + \sum_{m=2}^n \frac{\pi_m}{\pi_0} p_{n-m}(u) \right\} (1 - G(u)) du \\ &= \alpha \int_0^\xi (1 - G(u)) \sum_{m=1}^n \pi_m^* p_{n-m}(u) du \end{aligned}$$

and (3.1) follows by differentiation. (3.2) could be derived in a similar manner, but follows more directly from (3.1), (1.5). We get

$$\begin{aligned} \pi_n(1 - V_n(\eta)) &= \pi_n \int_0^\infty \frac{1 - G(u + \eta)}{1 - G(u)} dU_n(u) \\ &= \alpha \int_0^\infty \sum_{m=1}^n \pi_m^* p_{n-m}(u) (1 - G(u + \eta)) du \\ &= \alpha \int_0^\infty \int_0^\infty \sum_{m=1}^n \pi_m^* p_{n-m}(x - \xi) I(\eta \leq \xi \leq x < \infty) dG(x) d\xi, \end{aligned}$$

which is the same as $\int_\eta^\infty w_n(\xi) d\xi$, with w_n the r.h.s. of (3.2). Hence $v_n = w_n$. \square

3.2. Remark. In equilibrium, the rate of upcrossings $n \rightarrow n + 1$ is the same as the rate of downcrossings $n + 1 \rightarrow n$. Hence $\alpha\pi_n = \pi_{n+1} v_{n+1}(0)$ and it follows that the equilibrium equations for $\{\pi_n\}$, $\{V_n\}$ as given by Gnedenko and Kovalenko (1968) p. 158 can be written as

$$\pi_1 v_1(x) = \alpha\pi_1^*(1 - G(x)) - \alpha\pi_1(1 - V_1(x)), \tag{3.4}$$

$$\pi_n v_n(x) = \alpha\pi_{n-1}(1 - V_{n-1}(x)) - \alpha\pi_n(1 - V_n(x)) + \alpha\pi_n(1 - G(x)). \tag{3.5}$$

An alternative verification of (3.2) is possible using (3.4), (3.5) and induction.

For some purposes (3.4), (3.5) are quite convenient. Consider e.g. μ_n^r , the r^{th} moment of V_n . Then multiplying (3.4), (3.5) by x^r and integrating yields the set

$$\pi_1 \mu_1^r = \alpha \pi_1^* \frac{v^{r+1}}{r+1} - \alpha \pi_1 \frac{\mu_1^{r+1}}{r+1}, \tag{3.6}$$

$$\pi_n \mu_n^r = \alpha \pi_{n-1} \frac{\mu_{n-1}^{r+1}}{r+1} - \alpha \pi_n \frac{\mu_n^{r+1}}{r+1} + \alpha \pi_n \frac{v^{r+1}}{r+1}, \tag{3.7}$$

of equations (with v^r the r^{th} moment of G), which combined with $\mu_n^0 = 1$ determines the μ_n . E.g. in this manner one can check (after some tedious algebra) that

$$\mu_n^1 = \frac{1-\rho}{\alpha \pi_n} \sum_{k=n+1}^{\infty} \pi_k, \quad \mu_n^2 = \frac{2(1-\rho)}{\alpha^2 \pi_n} \sum_{k=n+2}^{\infty} (k-n-1) \pi_k - \frac{v^2}{\pi_n} \sum_{k=n+1}^{\infty} \pi_k. \tag{3.8}$$

A set of equations similar to (3.4), (3.5) involving the U_n rather than V_n seems only to hold if G is absolutely continuous, cf. Cohen (1969) II.6.2 (adapted to the equilibrium situation). In any case, moments are available directly from (3.1). We omit the details.

We can now easily prove

3.3. Theorem. *If Conditions (1.2), (1.3) hold, the distributions U_∞ , V_∞ with densities*

$$u_\infty(\xi) = \alpha e^{\rho \xi} (1 - G(\xi)), \quad v_\infty(\xi) = \alpha \int_\xi^\infty e^{\rho(x-\xi)} dG(x)$$

are proper, $U_\infty(\infty) = V_\infty(\infty) = 1$, and $u_n(\xi) \rightarrow u_\infty(\xi)$, $v_n(\xi) \rightarrow v_\infty(\xi) \forall \xi \geq 0$. In particular (cf. Billingsley (1968) pp. 224), U_n and V_n converge weakly and in total variation to U_∞ , resp. V_∞ . Conversely, if U_n has a proper limit as $n \rightarrow \infty$, then Condition (1.2) holds.

Proof. That U_∞ is proper is inherent in (1.2), and that V_∞ is so follows by the obvious integration by parts. Furthermore, from (1.1) it follows that there is a constant c_1 such that for all n and k , $\pi_{n-k}^*/\pi_n \leq c_1 \delta^k$, and also that $\pi_{n-k}^*/\pi_n \rightarrow \delta^k$ as $n \rightarrow \infty$. Hence by dominated convergence

$$u_n(\xi) = \alpha (1 - G(\xi)) \sum_{k=0}^{n-1} \frac{\pi_{n-k}^*}{\pi_n} p_k(\xi) \rightarrow \alpha (1 - G(\xi)) \sum_{k=0}^{\infty} \delta^k p_k(\xi) = u_\infty(\xi).$$

In a similar manner it follows that $v_n(\xi) \rightarrow v_\infty(\xi)$.

Suppose conversely that U_n has a proper limit U_∞ . Then, appealing to (1.5), V_n has a proper limit V_∞ . It can be assumed that the support of G is unbounded (since otherwise (1.2) is automatic) and then passing to the limit in (1.5) shows that V_∞ is not degenerate at zero. Let ξ be some continuity point of V_∞ with $V_\infty(\xi) < 1$. Then integrating (3.5) from 0 to ξ shows that

$$\int_0^\xi v_n(x) dx = \alpha \left\{ \frac{\pi_{n-1}}{\pi_n} \int_0^\xi (1 - V_{n-1}) - \int_0^\xi (1 - V_n) + \int_0^\xi (1 - G) \right\}$$

has a limit (viz. $V_\infty(\xi)$). Since $\int_0^\xi (1 - V_{n-1}) \rightarrow \int_0^\xi (1 - V_\infty) \neq 0$, π_{n-1}/π_n must have a limit, say δ . Let $\tilde{\delta} > \delta$, $\gamma = \alpha(\delta - 1)$, $\tilde{\gamma} = \alpha(\tilde{\delta} - 1)$ and choose K such that $\pi_{n-k}^*/\pi_n \leq K \tilde{\delta}^k$ for all n, k . Then by (3.1), $u_n(\xi)$ is dominated by $K\alpha(1 - G(\xi))e^{\tilde{\gamma}\xi}$ and tends to $\alpha(1 - G(\xi))e^{\gamma\xi}$. Thus for any continuity point x of U_∞ ,

$$U_\infty(x) = \lim_{n \rightarrow \infty} \int_0^x u_n(\xi) = \alpha \int_0^x (1 - G(\xi)) e^{\gamma\xi} d\xi.$$

Letting $x \rightarrow \infty$ shows that Condition (1.2) is satisfied. \square

One might note as a contrast to (1.7), that in general $U_\infty \neq V_\infty$. A marked difference is that the tail $1 - V_\infty(x)$ tend to decrease more rapidly (always as $\alpha(e^{-\gamma x})$) than $1 - U_\infty(x)$. E.g. (1.2), (1.3) suffice for the existence of all moments of V_∞ but only the mean of U_∞ .

As an obvious application of 3.3, consider the length $C_t = A_t + B_t$ of the current service period:

3.4. Corollary. *Conditions (1.2), (1.3) imply the existence of $W_\infty(\xi) = \lim_{n \rightarrow \infty} P_e(C_t \leq \xi | Q_t = n)$. W_∞ is larger than G in the stochastic ordering and is absolutely continuous w.r.t. G with density*

$$\frac{dW_\infty(\xi)}{dG(\xi)} = \frac{\alpha}{\gamma} \{e^{\gamma\xi} - 1\}. \tag{3.9}$$

Proof. Since $\int f dU_n \rightarrow \int f dU_\infty$ if f is bounded and a.e. continuous,

$$\begin{aligned} P_e(C_t > \xi | Q_t = n) &= 1 - U_n(\xi) + \int_0^\xi \frac{1 - G(\xi)}{1 - G(u)} dU_n(u) \\ &\rightarrow 1 - U_\infty(\xi) + \int_0^\xi \frac{1 - G(\xi)}{1 - G(u)} dU_\infty(u) \\ &= \alpha \int_\xi^\infty e^{\gamma u} (1 - G(u)) du + \alpha(1 - G(\xi)) \int_0^\xi e^{\gamma u} du \\ &= \frac{\alpha}{\gamma} \int_\xi^\infty (e^{\gamma u} - 1) dG(u). \end{aligned}$$

Since the r.h.s. of (3.9) has G -integral one according to (1.2), it follows that indeed W_∞ exists and has the form (3.9). The stochastic domination follows from the fact that (3.9) is non-decreasing in ξ . Indeed, if $W_\infty(\xi) > G(\xi)$ for some ξ , then necessarily $\frac{\alpha}{\gamma} \{e^{\gamma\xi} - 1\} > 1$ so that a contradiction results from

$$1 = W_\infty(\xi) + \int_\xi^\infty \frac{\alpha}{\gamma} \{e^{\gamma u} - 1\} dG(u) > G(\xi) + \int_\xi^\infty dG(u) = 1. \quad \square$$

4. The Growth to Large Values

The main result of the present section (and one of the main ones of the whole paper) could informally be described by the statement that (in equilibrium and subject to the limit $n \rightarrow \infty$) prior to the large value $Q_t = n$, the process has behaved as if the arrival intensity were $\tilde{\alpha} = \alpha \delta$ and the service time distribution were the distribution \tilde{G} with density $\frac{1}{\delta} e^{\gamma t}$ w.r.t. G . Note that \tilde{G} is stochastically larger than G , cf. the proof of 3.4, and that the M/G/1 model specified by $\tilde{\alpha}, \tilde{G}$ is transient since

$$\begin{aligned} \tilde{\rho} &= \tilde{\alpha} \int_0^\infty x d\tilde{G}(x) = \alpha \int_0^\infty x e^{\gamma x} dG(x) \\ &= \alpha \int_0^\infty \{e^{\gamma x} + \gamma x e^{\gamma x}\} (1 - G(x)) dx = 1 + \alpha \gamma \kappa > 1. \end{aligned}$$

Various formal statements of this result is possible. We start off in 4.1 with the version readily provided by means of regenerative processes and reformulate two corollaries 4.2, 4.3 in more abstract terms.

In order to be able to describe the whole past prior to t , it will be convenient to take $t=0$ and assume the equilibrium queue length process represented as a stationary process $\{Q_t\}_{-\infty < t < \infty}$ with doubly infinite time scale (cf. Breiman (1968) Prop. 6.5) and left-continuous path with right-hand limits. Then the growth prior to 0 is described by means of the random element $(Q_0 - Q_{-t})_{t \geq 0}$ of $D[0, \infty)$. Let $0 > -Y_0 > -Y_0 - Y_1 > \dots > -Y_0 - \dots - Y_j > \dots$ be the instants in $(-\infty, 0]$ where service is completed, T_0, T_j the number of arrivals in $(-Y_0, 0]$, resp. $(-Y_0 - \dots - Y_j, Y_0 - \dots - Y_{j-1}]$, let the arrival instants be of the form $-Y_0 - \dots - Y_{j-1} - Z_r^j$ with $0 < Z_1^j < \dots < Z_{T_j}^j < Y_j$ and let finally $\Phi_j = (Y_j, Z_1^j, \dots, Z_{T_j}^j)$. Then Φ_j is a random element of $\Omega = \bigcup_0^\infty (0, \infty)^{k+1}$, Φ_j taking its value in the k^{th} component on $\{T_j = k\}$, and equipping Ω with the obvious topology, we have

4.1. Theorem. *Suppose that Conditions (1.2), (1.3) hold. Then for any r , the e.d. of Φ_0, \dots, Φ_r given $Q_0 = n$ has a limit as $n \rightarrow \infty$, which can be described as follows: (i) Φ_0, \dots, Φ_r are independent; (ii) the distribution of Y_j is U_∞ for $j=0$ and \tilde{G} for $j>0$; (iii) given $Y_j = y$, T_j is Poisson distributed with mean $\tilde{\alpha}y$; and (iv) given $Y_j = y$, $T_j = k$, the distribution function of Z_1^j, \dots, Z_k^j is $F_{y,k}$, the k -variate d.f. of the order statistics corresponding to k drawings from a uniform distribution on $(0, y)$.*

Proof. Let $\Phi_j(t) = (Y_j(t), Z_1^j(t), \dots, Z_{T_j}^j(t))$ be defined relative to time t rather than time 0, let $F(t)$ be the event that at time t the server is busy and the r preceding service periods fall within the present busy period and define

$$E^r(t) = I(Y_j(t) \leq y_j, T_j(t) = k_j, Z_i^j \leq z_i^j; j = 0, \dots, r, i = 1, \dots, k_j),$$

$E''(t) = E'(t) I(F(t))$. Then the assertion amounts to

$$\begin{aligned} \lim_{n \rightarrow \infty} E_e(E'(0) | Q_0 = n) &= \int_0^{y_0} p_{k_0}(\delta u_0) F_{k_0, u_0}(z_1^0, \dots, z_{k_0}^0) dU_\infty(u_0) \\ &\cdot \prod_{j=1}^r \int_0^{y_j} p_{k_j}(\delta u_j) F_{k_j, u_j}(z_1^j, \dots, z_{k_j}^j) d\tilde{G}(u_j). \end{aligned} \tag{4.1}$$

Now $E''(t) I(Q_t = n)$ is regenerative and hence $E_e(E''(0); Q_0 = n)$ computable by means of (1.8). We use the imbedded Markov chain in a similar manner as in the proof of 3.1. In order for $E''(t) I(Q_t = n)$ to equal one, $X_k = n - k_0 - \dots - k_r + r$ customers must have been present at the start $\tau(k)$ of the r th among the preceding service periods and we must have all $n - k_0 - \dots - k_j + j \geq 1$ (since otherwise the queue is empty between $\tau(k)$ and t). The latter requirement is satisfied if n is sufficiently large, say $n \geq k_0 + \dots + k_j$ and similar arguments as in the proof of 3.1 then yield the expression

$$\begin{aligned} \alpha(1 - \rho) \frac{\pi_{n-k_0-\dots-k_r+r}^*}{\pi_0} \int_0^{y_0} p_{k_0}(u_0) F_{k_0, u_0}(z_1^0, \dots, z_{k_0}^0) (1 - G(u_0)) du_0 \\ \cdot \prod_{j=1}^r \int_0^{y_j} p_{k_j}(u_j) F_{k_j, u_j}(z_1^j, \dots, z_{k_j}^j) dG(u_j), \end{aligned}$$

for $E_e(E''(0), Q_0 = n)$. Dividing by π_n and using (1.1) shows that

$$\begin{aligned} \lim_{n \rightarrow \infty} E_e(E''(0) | Q_0 = n) &= \int_0^{y_0} p_{k_0}(u_0) \delta^{k_0} F_{k_0, u_0}(z_1^0, \dots, z_{k_0}^0) \alpha(1 - G(u_0)) du_0 \\ &\cdot \prod_{j=1}^r \int_0^{y_j} p_{k_j}(u_j) \delta^{k_j} F_{k_j, u_j}(z_1^j, \dots, z_{k_j}^j) \frac{1}{\delta} dG(u_j) = \text{r.h.s. of (4.1)}, \end{aligned} \tag{4.2}$$

using

$$p_k(u) \delta^k = e^{-\alpha u} \frac{(\alpha u \delta)^k}{k!} = e^{\gamma u} p_k(u \delta).$$

Thus (4.1) will follow if $\lim_{n \rightarrow \infty} P_e(F(0) | Q_0 = n) = 1$. But summing (4.2) shows that

$$\lim_{n \rightarrow \infty} P_e(F(0) | Q_0 = n) \geq \sum_{k_0=0}^{K_0} \dots \sum_{k_r=0}^{K_r} \text{r.h.s. of (4.1)}$$

which can be taken arbitrarily close to 1 upon choosing $y_0, \dots, y_r, K_0, \dots, K_r$ large enough. \square

Let N'_t, N''_t be the number of departures, resp. arrivals in $[-t, 0]$ and $\mathbf{N}', \mathbf{N}''$ the corresponding point processes, i.e. random elements of the space \mathfrak{R} of counting measures on $[0, \infty)$. The vague topology on \mathfrak{R} defines the concept of weak convergence of point processes in the usual manner, cf. e.g. Neveu (1977).

4.2. Corollary. *As $n \rightarrow \infty$, the e.d. of (N', N'') given $Q_0 = n$ converges weakly to the distribution of (K', K'') where: K', K'' are independent; K' is a renewal process with delay distribution U_∞ and interarrival distribution \tilde{G} ; K'' is a stationary Poisson process with intensity $\tilde{\alpha}$.*

Note that (except for special cases like G exponential) $u_\infty(\xi)$ is not proportional to $1 - \tilde{G}(\xi)$ and hence K' not stationary. This irregularity is shown to vanish in the set-up of Sect. 5.

Proof. The statement of 4.2 is almost obvious from 4.1, but a formal proof may proceed along the following lines. The statement of 4.1 may be reformulated that the e.d. of the sequence $\{\Phi_j\}_{j \in \mathbb{N}}$ given $\{Q_0 = n\}$ converges weakly in $\Omega^{\mathbb{N}}$ to the product probability measure μ described in 4.1, weak convergence in $\Omega^{\mathbb{N}}$ meaning just weak convergence of coordinates $0, \dots, r$ for any r . For $\phi_0, \phi_1, \dots \in \Omega$, writes $S = S(\phi_0, \phi_1, \dots) = y_0 + y_1 + \dots$, and consider the mapping $\Delta': \Omega^{\mathbb{N}} \rightarrow \mathfrak{R}$ which takes $\{\phi_j\}$ into the counting measure placing unit weights at the points $y_0 + \dots + y_r$ with (say) $y_0 + \dots + y_r \leq S/2$ (i.e. all $y_0 + \dots + y_r$ if $S = \infty$). It is then a matter of routine to check that Δ' is continuous at every $\{\phi_j\}$ with $S = \infty$ and μ being concentrated on $\{S = \infty\}$, it follows that the departure process $N' = \Delta'(\Phi_0, \Phi_1, \dots)$ indeed converges weakly to K' . A mapping $\Delta'': \Omega^{\mathbb{N}} \rightarrow \mathfrak{R}$ constructed in a similar spirit produces the arrival process and since clearly $(\Delta', \Delta''): \Omega^{\mathbb{N}} \rightarrow \mathfrak{R}^2$ maps μ into the distribution of (K', K'') , the proof is complete. \square

4.3. Corollary. *As $n \rightarrow \infty$, the e.d. of $\{Q_0 - Q_{-t}\}_{t \geq 0} = \{N'_t - N''_t\}_{t \geq 0}$ given $\{Q_0 = n\}$ converges weakly in $D[0, \infty)$ (cf. Lindvall (1973)) to the distribution of $\{K'_t - K''_t\}_{t \geq 0}$.*

The proof is a similar obvious application of the continuous mapping theorem.

It is instructive to review the above results in the $M/M/1$ case, where $1 - G(x) = e^{-\sigma x}$ with $\rho = \alpha/\sigma$. Straightforward calculations then show that $\tilde{\alpha} = \sigma$ and that $1 - U_\infty(x) = 1 - \tilde{G}(x) = e^{-\alpha x}$. Hence by 4.2, 4.3 in the limit, $\{Q_0 - Q_{-t}\}_{t \geq 0}$ is the difference between two independent stationary Poisson processes with intensities α , respectively σ . However, it is well-known (Reich (1957)) that $\{Q_t\}_{-\infty < t < \infty}$ is time-reversible at equilibrium. Thus N', N'' are the arrival, resp. departure, processes of the time-reversed process. In particular N' is stationary Poisson with intensity α . Now conditioning $\{Q_t\}$ on the final value $Q_0 = n$ amounts to starting the time reversed process at n . But the departure process of a $M/M/1$ queue started at n is readily verified to approach a Poisson process with intensity σ as $n \rightarrow \infty$. Hence the results of 4.2, 4.3 are exactly the ones implied by the time reversibility.

5. The Virtual Waiting Time

The notation and main results of Sects. 3-4 are used without further reference. Our first objective is to reformulate the results of Sect. 4 in terms of large

virtual waiting times. That is, rather than $\lim_{n \rightarrow \infty} P_e(\cdot | Q_0 = n)$ we consider $\lim_{x \rightarrow \infty} P_e(\cdot | v_0 > x)$. Let \tilde{v} denote the mean of \tilde{G} .

5.1. Theorem. *Suppose that Conditions (1.2), (1.3) hold. Then:*

(i) *For all ξ ,*

$$\lim_{x \rightarrow \infty} P_e(A_t \leq \xi | v_t > x) = \lim_{x \rightarrow \infty} P_e(B_t \leq \xi | v_t > x) = \frac{1}{\tilde{v}} \int_0^\xi (1 - \tilde{G}(y)) dy. \tag{5.1}$$

(ii) *The e.d. given $v_t > x$ of (N', N'') converges weakly as $x \rightarrow \infty$ to the distribution of (L', L'') where: L', L'' are independent; L' is a stationary renewal process with interarrival distribution \tilde{G} ; L'' is a stationary Poisson process with intensity $\tilde{\alpha}$.*

5.2. Remark. Of course, the r.h.s. of (5.1) represents the stationary wait and delay in a renewal process with interarrival distribution \tilde{G} .

Proof. We first note the estimate

$$P_e(v_t > x | Q_t = n) = o(e^{-\gamma x}) = o(P_e(v_t > x)) \quad \text{as } x \rightarrow \infty \tag{5.2}$$

valid for any fixed n . In view of (1.6) it suffices to show $1 - G^{*(n-1)}(x) = o(e^{-\gamma x})$. But, using induction and dominated convergence,

$$e^{\gamma x} (1 - G^{*n}(x)) = \int_0^x e^{\gamma(x-y)} (1 - G^{*(n-1)}(x-y)) e^{\gamma y} dG(y) + e^{\gamma x} (1 - G(x)) \rightarrow 0.$$

Now let the arrivals prior to 0 take place at times $0 > -D_1 > -D_1 - D_2 > \dots$ and define

$$F = \{Y_i \leq y_i \quad i = 1, \dots, r, D_j > \eta_j \quad j = 1, \dots, s\},$$

$$f = \lim_{n \rightarrow \infty} P_e(F | Q_0 = n) = \tilde{G}(y_1) \dots \tilde{G}(y_r) e^{-\tilde{\alpha}(\eta_1 + \dots + \eta_s)}.$$

Then, in view of (1.1), (2.4) and (5.2),

$$\begin{aligned} & P_e(F, Y_0 \leq y_0, B_0 > b | v_0 > x) \\ &= \frac{\sum_{n=1}^{\infty} \pi_n P_e(F, Y_0 \leq y_0, B_0 > b, v_0 > x | Q_t = n)}{P_e(v_0 > x)} \\ &\cong d^{-1} e^{\gamma x} \sum_{n=1}^{\infty} c \delta^{-n} f \int_0^{y_0} u_{\infty}(z) dz \frac{1}{1 - G(z)} \int_{z+b}^{\infty} (1 - G^{*(n-1)}(x - v + z)) dG(v) \\ &= \gamma f \int_0^{y_0} dz \int_{z+b}^{\infty} e^{\gamma(x-v+z)} \sum_{n=1}^{\infty} \delta^{-n} (1 - G^{*(n-1)}(x - v + z)) e^{\gamma v} dG(v). \end{aligned} \tag{5.3}$$

Now consider the transient renewal function $U = \sum_{n=0}^{\infty} F^{*n}$ where $F = \delta^{-1} G$. Since $\int_0^{\infty} e^{\gamma x} dF(x) = 1$, it follows from Feller (1971) pp. 374–377 that

$$U(\infty) - U(x) \cong \frac{e^{-\gamma x}}{\gamma \int_0^\infty x e^{\gamma x} dF(x)} = \frac{e^{-\gamma x}}{\gamma \tilde{v}}.$$

Thus

$$\text{r.h.s. of (5.3)} \cong \frac{f}{\delta \tilde{v}} \int_0^{y_0} dz \int_{z+b}^\infty e^{\gamma v} dG(v) = \frac{f}{\tilde{v}} \int_0^{y_0} (1 - \tilde{G}(z+b)) dz. \tag{5.4}$$

Taking first $b=0$, it follows that the limiting distribution of $Y_0, \dots, Y_r, D_1, \dots, D_s$ is as asserted in Part (ii) and Part (ii) follows easily. For Part (i), take first $y_0 = y_1 = \dots = y_r = \infty, \eta_1 = \dots = \eta_s = 0$ so that $f=1$ and (5.4) reads

$$P_e(B_0 > b | v_0 > x) \cong \frac{1}{\tilde{v}} \int_b^\infty (1 - \tilde{G}(z)) dz$$

which is equivalent to the assertion on B_t .

For the one on A_t , let again $t=0$, let $F(0)$ be as in the proof of 4.1 with $r=0$ and recall that $P_e(F(0) | Q_0 = n) \rightarrow 1$ as $n \rightarrow \infty$. Hence $P_e(F(0) | v_t > x) \rightarrow 1$ as $x \rightarrow \infty$ in view of (5.2). But $A_0 = Y_0$ on $F(0)$ so that Part (ii) applies. \square

The results of Sect. 4 and 5.1 describe the behaviour of the increments of the queue length process. We next turn to the increments of the virtual waiting time process, which are shown to behave as the difference between a linear function and a compound Poisson process. Let as before the paths of $\{v_t\}_{-\infty < t < \infty}$ be normalized to be left-continuous at the jump points (i.e. times of arrivals).

5.3. Theorem. *Suppose that Conditions (1.2), (1.3) hold. Then as $n \rightarrow \infty$, the e.d. given $Q_0 = n$ of $\{v_0 - v_{-t}\}_{t \geq 0}$ converges weakly in $D[0, \infty)$ to the distribution of*

$\left\{ \sum_1^{M_t} Z_j - t \right\}_{t \geq 0}$, where \mathbf{M} is a stationary Poisson process with intensity $\tilde{\alpha}$ and Z_1, Z_2, \dots are independent of \mathbf{M} and i.i.d. with distribution G . If rather than $\lim_{n \rightarrow \infty} P_e(\cdot | Q_0 = n)$ one considers $\lim_{x \rightarrow \infty} P_e(\cdot | v_0 > x)$, the same conclusion holds except that the common distribution of the Z_j is now \tilde{G} .

Proof. Let Z_j be the service time of the j^{th} customer arriving before 0 and M_t the number of arrivals in $[-t, 0]$. Then the paths of $\{v_0 - v_{-t}\}_{t \geq 0}$ and $\left\{ \sum_1^{M_t} Z_j - t \right\}_{t \geq 0}$ coincide on $[0, \tau]$, with $-\tau$ the last time before 0 where the queue has been empty. It follows from the above results and proofs that, subject to the limits considered, $\tau \rightarrow \infty$ in distribution and that the distribution of \mathbf{M} is as claimed. Hence the theorem follows in a routine manner once the Z_j in the limit are shown to have the distributional properties asserted.

Now let F be any measurable subset of \mathfrak{R} and f the probability assigned to F by the Poisson process with intensity $\tilde{\alpha}$. Fix r, z_1, \dots, z_r and let H be the event that the r^{th} customer arriving before 0 starts his service after 0. It then follows easily from Sect. 4 that $P_e(H | Q_0 = n) \rightarrow 1$ as $n \rightarrow \infty$. Hence

$$\begin{aligned}
 &P_e(\mathbf{M} \in F, Z_j \leq z_j, j=1, \dots, r | Q_0 = n) \\
 &\cong P_e(\mathbf{M} \in F, Z_j \leq z_j, j=1, \dots, r, H | Q_0 = n) \\
 &= G(z_1) \dots G(z_r) P_e(\mathbf{M} \in F, H | Q_0 = n) \\
 &\cong G(z_1) \dots G(z_r) P_e(\mathbf{M} \in F | Q_0 = n) \cong G(z_1) \dots G(z_r) f
 \end{aligned}$$

(conditioning upon the past prior to 0 to obtain the equality sign) and the claim follows subject to the conditioning upon Q_0 . For the one upon v_0 , we get as in the proof of 5.1

$$\begin{aligned}
 &P_e(\mathbf{M} \in F, Z_j \leq z_j, j=1, \dots, r | v_0 > x) \\
 &\cong \frac{\sum_{n=1}^{\infty} \pi_n P_e(\mathbf{M} \in F, Z_j \leq z_j, j=1, \dots, r, v_0 > x, H | Q_0 = n)}{P_e(v_0 > x)} \\
 &\cong d^{-1} e^{\gamma x} \sum_{n=r+1}^{\infty} c \delta^{-n} f \int_0^{\infty} dV_{\infty}(b) \int_0^{z_1} dG(y_1) \dots \\
 &\quad \dots \int_0^{z_r} dG(y_r) (1 - G^{*(n-1-r)}(x - b - y_1 - \dots - y_r)) \\
 &\cong \frac{\gamma f}{\alpha} \int_0^{\infty} dV_{\infty}(b) \int_0^{z_1} dG(y_1) \dots \\
 &\quad \dots \int_0^{z_r} dG(y_r) \delta^{-1-r} e^{\gamma x} \sum_{k=0}^{\infty} \delta^{-k} (1 - G^{*k}(x - b - y_1 - \dots - y_r)) \\
 &\cong \frac{\gamma f}{\alpha} \int_0^{\infty} dV_{\infty}(b) \int_0^{z_1} dG(y_1) \dots \int_0^{z_r} dG(y_r) \delta^{-1-r} \frac{e^{\gamma(b+y_1+\dots+y_r)}}{\gamma \tilde{v}} \\
 &= \frac{f}{\delta \alpha \tilde{v}} \tilde{G}(z_1) \dots \tilde{G}(z_r) \int_0^{\infty} e^{\gamma b} dV_{\infty}(b) \\
 &= \frac{f}{\delta \tilde{v}} \tilde{G}(z_1) \dots \tilde{G}(z_r) \int_0^{\infty} e^{\gamma b} db \int_b^{\infty} e^{\gamma(\xi-b)} dG(\xi) \\
 &= \frac{f}{\tilde{v}} \tilde{G}(z_1) \dots \tilde{G}(z_r) \int_0^{\infty} \xi d\tilde{G}(\xi) = f \tilde{G}(z_1) \dots \tilde{G}(z_r). \quad \square
 \end{aligned}$$

Acknowledgements. I would like to thank my colleagues at the Universities of Gothenburg and Copenhagen for stimulating the present research through their comments on the preliminary versions presented to them. Thanks also to the referee for comments and corrections.

References

1. Asmussen, S.: Conditioned Limit Theorems Relating a Random Walk to its Associate, With Applications to Risk Reserve Processes and the GI/G/1 Queue. *Advances in Appl. Probability* **14** (to appear) (1980)
2. Asmussen, S.: Time-dependent Approximations in Some Queuing Systems With Imbedded Markov Chains Related to Random Walks. Submitted for publication (1981)
3. Billingsley, P.: *Convergence of Probability Measures*. New York: Wiley 1968
4. Breiman, L.: *Probability*. Reading, Mass.: Addison-Wesley 1968

5. Cohen, J.W.: Extreme Value Distributions for the $M/G/1$ and $G/M/1$ Queueing Systems. *Ann. Inst. H. Poincaré Sect. B* **4**, 83–98 (1968)
6. Cohen, J.W.: *The Single Server Queue*. Amsterdam: North-Holland 1969
7. Cohen, J.W.: Some Results on Regular Variation for Distributions in Queueing and Fluctuation Theory. *J. Appl. Probability* **10**, 343–353 (1973a)
8. Cohen, J.W.: Asymptotic Relations in Queueing Theory. *Stochastic Processes Appl.* **1**, 107–124 (1973b)
9. Cohen, J.W.: *On Regenerative Processes in Queueing Theory*. Lecture Notes in Economics and Mathematical Systems **121**. Berlin-Heidelberg-New York: Springer 1976
10. Cox, D.R. and Smith, W.L.: *Queues*. London: Methuen 1961
11. Feller, W.: *An Introduction to Probability Theory and Its Applications* **2**, 2nd Ed. New York: Wiley 1971
12. Gaver, D.P., Jr.: Imbedded Markov Chain Analysis of a Waiting Line Process in Continuous Time. *Ann. Math. Statist.* **30**, 698–720 (1959)
13. Gnedenko, B., Kovalenko, I.N.: *An Introduction to Queueing Theory*. Jerusalem: Israel Program for Scientific Translations 1968
14. Hokstad, P.: A Supplementary Variable Technique Applied to the $M/G/1$ Queue. *Scand. J. Statist.* **2**, 95–98 (1975)
15. Iglehart, D.L.: Extreme Values in the $GI/G/1$ Queue. *Ann. Math. Statist.* **43**, 627–635 (1972)
16. Kingman, J.F.C.: A Martingale Inequality in the Theory of Queues. *Proc. Cambridge Philos. Soc.* **60**, 359–361 (1964)
17. Le Gall, P.: *Les Systemes avec ou sans Attente et les Processus Stochastiques*. Paris: Dunod 1962
18. Lindvall, T.: Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0, \infty)$. *J. Appl. Probability* **10**, 109–121 (1973)
19. Miller, D.R.: Existence of Limits in Regenerative Processes. *Ann. Math. Statist.* **43**, 1275–1282 (1972)
20. Neveu, J.: *Processus Ponctuels*. Lecture Notes in Mathematics **598**, 249–445. Berlin-Heidelberg-New York: Springer 1977
21. Prabhu, N.U.: *Queues and Inventories*. New York: Wiley 1965
22. Reich, E.: Waiting Times when Queues are in Tandem. *Ann. Math. Statist.* **28**, 768–773 (1957)
23. Saaty, T.L.: *Elements of Queueing Theory*. New York: McGraw-Hill 1961
24. Smith, W.L.: Regenerative Stochastic Processes. *Proc. Roy. Soc. London Ser. A* **232**, 6–31 (1955)

Received May 7, 1980; in revised form May 10, 1981