Stochastic recursive equations with applications to queues with dependent vacations

Eitan Altman
INRIA, BP93, 06902 Sophia Antipolis, France, and
C.E.S.I.M.O., Universidad de Los Andes, Facultad de Ingeneria, Mérida, Venezuela

November 2001

Abstract. We focus on a special class of nonlinear multidimensional stochastic recursive equations in which the coefficients are stationary ergodic (not necessarily independent). Under appropriate conditions, an explicit ergodic stationary solution for these equations is obtained and the convergence to this stationary regime is established. We use these results to analyze several queueing models with vacations. We obtain explicit solutions for several performance measures for the case of general non-independent vacation processes. We finally extend some of these results to polling systems with general vacations.

1. Introduction

There has been a rich amount of work on stochastic recursive equations under general stationary ergodic setting. One research direction has been to analyze stability properties of such equations, such as the existence of an ergodic stationary regime, as well as the convergence and even coupling to that regime [3, 8, 9, 10, 12, 13]. Another important direction was to obtain an explicit distribution for linear recursive equations [12, 13, 15, 23] (linearity was analyzed not just in the usual algebra but also in the max-plus algebra, see e.g. [15]).

In this paper we follow in particular the second research direction and study a special class of non-linear multidimensional recursive equations of the form $Y_{n+1} = A_n(Y_n) + B_n$. The pair $(A_n(\cdot), B_n)$ need not be i.i.d.: all we require is that it is a stationary ergodic sequence. Although A_n is not linear, we make some assumptions that makes the behavior of the sequence "almost" linear: in most of the paper we assume that A_n is infinitely divisible.

In the second part of the paper we apply our results to vacation models in queueing. We consider Poisson arrivals and general service time distribution which are i.i.d. The vacations sequence, however, are allowed to be correlated: we only assume that they form a stationary ergodic sequence. We focus on gated service discipline and briefly discuss the much simpler exhaustive service discipline. In the gated discipline, upon the completion of a vacation, the server serves only those customers present at that instant and then leaves for another vacation, whereas in the exhaustive service, the server continues service till the queue empties before going for another vacation.

For the case of a single queue with a gated regime, we first show that both the queue length at end of vacations as well as the cycle times (a busy period followed by a vacation) can be described using our stochastic recursive equations formalism, which allows us to obtain explicit expressions for the distribution of both quantities. In particular, we obtain the two first moments of these quantities. This allows us to obtain the expected waiting time and queue length at an arbitrary time.

For completeness, we also mention the exhaustive case, where a degenerate form of our stochastic recursive equations holds, which results in an insensitivity property: the marginal distribution of the queue lengths at the end of a vacation or at an arbitrary time do not depend on the correlation between the vacations, and are the same as those obtained when replacing the vacation sequence by an i.i.d. sequence of vacations with the same marginal distribution.

We then extend some of these results to the multidimensional systems with vacations, i.e. to polling systems, in which several queues are served according to a cyclic order, and vacations are incurred when the server moves from one queue to another. We consider both the exhaustive regime, the gated regime, as well as the globally gated regime [11]. We show again that our formalism for the stochastic recursive equations holds.

We remark that a huge literature exists on queues with vacations and on polling systems. However, except for stability results [14], the only existing analytical results we know of in which vacations are allowed to be correlated are those on expected cycle times and expected number of queued customers at polling instants (see [5]). For an extensive survey on polling, for which vacations are independent, see [19, 20, 22].

2. Convergence to a stationary ergodic regime

Consider the stochastic recursion

$$Y_{n+1} = A_n(Y_n) + B_n, \qquad n \geq n_0. \tag{1}$$

Y_n is a random variable defined on a subset $\mathcal{Y}$ of R^m for some integer m, equipped with some norm $\|\cdot\|$. We assume that $0 \in \mathcal{Y}$ where 0 is the zero vector. $\{A_n(\cdot)\}_{-\infty<n<\infty}$ are a sequence of random processes taking values in R^m.

We present the following assumption:

A1: For each sample path, the vectors $A_n(y)$ and B_n are nonnegative (componentwize) for all n and y. Moreover, $A_n(y)$ are monotone increasing in y for all n and for each sample path.

Theorem 1. Assume that the sequence $\{(A_n(\cdot), B_n), -\infty < n < \infty\}$ is stationary ergodic[1], defined on some probability space $(\Omega, \mathcal{F}, P)$, and assume that A1 holds. Assume that when fixing $Y_0 = 0$,

$$P(\overline{\lim}_{n\to\infty} Y_n \text{ is finite }) > 0. \tag{2}$$

Then
(i) There is a stationary ergodic regime Y_n^* defined on the same probability space and Y_n^* satisfies (1),
(ii) If $Y_0 = 0$ then $\|Y_n - Y_n^*\| \to 0$, $P-$almost surely.

`Proof.` The proof follows the standard Loynes scheme [18]. We define on the same probability space the sequence of processes $\{Y_n^{[\ell]}, n \geq -\ell, \ell = 0, 1, 2, ...\}$ all of whom satisfy (1) and where $Y_{-\ell}^{[\ell]} = 0$. The initial state, taken to be zero, is thus shifted to $n = -\ell$ in the process $Y^{[\ell]}$. It is easily seen that $Y_n^{[\ell]}$ is monotone increasing in ℓ for each n. Let Y_n^* denote its limit; since the sequence $(A_n(\cdot), B_n)$ is stationary ergodic, the random variable Y_n^* is finite with probability one or with probability zero. Condition (2) implies that it is indeed finite with probability 1. Clearly Y_n^* is stationary ergodic. This establishes (i).

Now since Y_n has the same distribution as $Y_0^{[n]}$, and since $Y_0^{[n]}$ monotonely increases to Y_0^* (samplewize), it follows that

$$E[Y_n^*] - E[Y_n] \text{ decreases to zero} \tag{3}$$

(The above difference is a vector and it decreases to zero componentwize). On the other since $Y_0 = 0 \leq Y_0^*$, it follows from the monotonicity assumption that

$$\liminf_{n\to\infty}(Y_n^* - Y_n) \geq 0.$$

Due to (3), the liminf has to be zero and is in fact a limit, which establishes (ii). ∎

Remark 1. An example is given in [2] where A_n is a constant function (does not depend on $\omega \in \Omega$ nor on n) and for which Y_n has different stationary ergodic limit processes depending on the different initial state Y_0.

[1] We mean by that notation that the sequence $\{((A_n(y))_{y\in\mathcal{Y}}, B_n), -\infty < n < \infty\}$ is stationary ergodic, rather than $\{(A_n(Y_n), B_n), -\infty < n < \infty\}$

Define the composition of the random processes A_iA_j as

$$(A_iA_j)(y) = A_i(A_j(y)).$$

Next we present another useful assumption that implies A1:
A2: For each realization, the vectors $A_n(y)$ and B_n are nonnegative (componentwize) for all n and y. Moreover, $A_n(\cdot)$ is infinitely divisible, i.e. for each n and k, there exist $A_n^{(i)}(\cdot)$, $i = 0, ..., k$ such that for any $x_i \in \mathcal{Y}, i = 0, ..., k$ satisfying $\sum_{i=0}^{k} x_i \in \mathcal{Y}$,

$$A_n\left(\sum_{i=0}^{k} x_i\right) = \sum_{i=0}^{k} A_n^{(i)}(x_i) \tag{4}$$

where $\{A_n^{(i)}(\cdot)\}_{i=0,1,2,...,k}$ are i.i.d. with the same distribution as $A_n(\cdot)$. In particular, $A_n(0) = 0$.

Remark 2. A divisibility assumption has already been used in [23] in a context of one dimensional stochastic recursions under the more restrictive framework in which $A_n(\cdot)$ and B_n are i.i.d. sequences.

It follows from Assumption A2 that if $A_n(\cdot)$ are stationary, then there exists a matrix A such that for any y,

$$E[A_n(y)] = \mathsf{A}y. \tag{5}$$

Moreover, in the case that A_n are i.i.d. then

$$E[A_nA_{n-1}...A_{m+1}(y)] = \mathsf{A}^{n-m}y, \qquad \text{for all } n > m.$$

It is due to (5) that under A2, the stochastic recursive equations will have a similar behavior as that of linear stochastic recursive equations.

Iterating (1) and using (4), we get

$$Y_n = \sum_{j=0}^{n-1}\left(\prod_{i=n-j}^{n-1} A_i^{(n-j)}\right)(B_{n-j-1}) + \left(\prod_{i=0}^{n-1} A_i^{(0)}\right)(Y_0), \quad n > 0 \tag{6}$$

(we understand $\prod_{i=n}^{k} A_i(x) = x$ whenever $k < n$, and $\prod_{i=n}^{k} A_i(x) = A_kA_{k-1}...A_n$ whenever $k > n$).

Theorem 2. Assume that the sequence $\{(A_n(\cdot), B_n), -\infty < n < \infty\}$ is stationary ergodic, defined on some probability space $(\Omega, \mathcal{F}, P)$, and let Assumption A2 hold. Assume that either
(i) Eq. (2) holds and P-almost surely

$$\lim_{n\to\infty}\left(\prod_{i=0}^{n-1} A_i^{(0)}\right)(y) = 0, \tag{7}$$

or (ii) the following holds:

$$\inf_{y\in\mathcal{Y}} P(A_0(y)=0) > 0. \tag{8}$$

Then there is a unique stationary solution Y_n^* of (1), distributed like

$$Y_n^* =_d \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} A_i^{(n-j)} \right) (B_{n-j-1}), \qquad n \in Z, \tag{9}$$

where for each integer i, $\{A_i^{(j)}(\cdot)\}_j$ are independent of each other and have the same distribution as $A_i(\cdot)$. The sum on the right side of (9) converges absolutely P-almost surely. Furthermore, for all initial conditions Y_0, $\|Y_n - Y_n^*\| \to 0$, P-almost surely on the same probability space. In particular, the distribution of Y_n converges to that of Y_0^* as $n \to \infty$. Finally under condition (8), Y_n strongly couples (see [8, p. 260-262] and [9]) with Y_n^* P-almost surely.

`Proof.` We first assume that condition (2) holds. In the same way as we did in the proof of Theorem 1, we can shift the origin of (6) to time $(-\ell)$ and consider the resulting process $Y_n^{[\ell]}$ for a fixed n; by choosing $Y_{-\ell} = 0$ and by taking the limit as ℓ goes to infinity, $Y_n^{[\ell]}$ converges to the stationary ergodic regime which is given by (9). Since the conditions of Theorem 1 hold, this limit is indeed finite.

Now let $Y_n(y)$ denote the value of Y_n given that $Y_0 = y$. Then we have from (6)

$$Y_n(y) - Y_n(0) = \left(\prod_{i=0}^{n-1} A_i^{(0)} \right) (y)$$

which by assumption, tends to zero P-almost surely. Since by Theorem 1, $\lim_{n\to\infty} \|Y_n(0) - Y_n^*\| = 0$ P-almost surely, this establishes the required convergence.

Next we consider assumption (8) instead. Due to the ergodicity of A_n it follows that (7) holds. The existence of a finite stationary ergodic regime as well as the convergence results (including the strong coupling convergence) follow from [8, p. 260-262] or [9, p. 21], using the fact that $A_n(Y_n) = 0$ turns out to be a renovating event [8, 9] which has a positive probability. ∎

Next we present simple sufficient conditions for (2) and (7) to hold.

Lemma 1. Assume that $A_n(\cdot)$ is an i.i.d. sequence, independent of the stationary ergodic sequence B_n. Further assume that

$$E[\|A_0(y)\|] \le \alpha \, \|y\|, \qquad \text{and} \qquad E[(\log \|B_0\|)^+] < \infty, \tag{10}$$

for some $\alpha < 1$ and all $y \geq D$ where $D > 0$ is some constant. Then Assumption A1 implies (2).
Now assume further that (10) holds for all y, and thus

$$\|\mathsf{A}\| < 1 \tag{11}$$

where A is given in (5). Then Assumption A2 implies (7).

`Proof.` Define the stochastic recursion:

$$y_{n+1}^{[\ell]} = \alpha y_n^{[\ell]} + \|B_n\|, \qquad n \geq -\ell, y_{-\ell}^{[\ell]} = 0 \tag{12}$$

where y_n takes values in the nonnegative real numbers. One can show recursively using (10) that $E[\left\|Y_n^{[\ell]}\right\|] \leq y_n^{[\ell]} + D$ for all $\ell \geq 0$ and all $n \geq -\ell$. Hence $E[\|Y_n^*\|] \leq y_n^* + D$ where $y_n^* := \lim_{\ell\to\infty} y_n^{[\ell]}$. y_n^* is finite with probability 1 (see e.g. Brandt [12] and [15]). This implies that $E[Y_n^*]$ is finite and hence Y_n^* is finite with probability one.

Next we establish (7). We have

$$E\left[\left(\prod_{i=0}^{n-1} A_i^{(0)}\right)(y)\right] = \mathsf{A}^n y \to 0 \tag{13}$$

where the convergence to zero as $n \to \infty$ is due to condition (11). Since $Y_n(y) \geq Y_n(0)$, (13) implies that

$$\lim_{n\to\infty} \|Y_n(y) - Y_n(0)\| = 0,$$

P-a.s., which combined with the fact that $\lim_{n\to\infty} \|Y_n(0) - Y_n^*\| = 0$ P-a.s., establishes the required convergence. ∎

Remark 3. By Jensen's inequality, a sufficient condition for the second part of (11) is that $E[\|B_0\|] < \infty$.

Next, we relax the i.i.d. assumptions on A_n. We consider the following representation of A_n. Consider a sequence of i.i.d. random functions $\hat{A}_n(y,\xi)$, $-\infty < n < \infty$, taking values in $\mathcal{Y}$, indexed by $y \in \mathcal{Y}$ and by some parameter ξ taking values in some measurable space Ξ. Introduce on the same probability space $(\Omega, \mathcal{F}, P)$ on which $\hat{A}_n, B_n$ are defined a random process ξ_n. Define now the random sequence $A_n(\cdot)$ by

$$A_n(y) = \hat{A}_n(y, \xi_n), \qquad -\infty < n < \infty, \quad \text{for all } y \in \mathcal{Y}$$

Let $\mathcal{F}_n$ be the σ-algebra generated by $(\xi_k, B_k), k \leq n$. Let M be the set of matrices of dimension $m \times m$. Introduce the following assumption.

A3(i): The process $\hat{A}_n(\cdot,\cdot)$ is independent of the processes B_n and ξ_n.
A3(ii): There exists a function $\overline{\mathsf{A}} : \Xi \to M$ such that for all $Y \in \mathcal{Y}$,

$$[\overline{\mathsf{A}}(\xi_n)]y = E[A_n(y)|\xi_n] = E[A_n(y)|\mathcal{F}_n]$$

P-almost surely ($[\overline{\mathsf{A}}(\xi_n)]y$ is the multiplication between the matrix $\overline{\mathsf{A}}(\xi_n)$ and the column vector y).

It follows from assumption A3 that for any $k > n$,

$$E\left[\prod_{i=n}^{k} A_i(y)\middle|\mathcal{F}_n, \xi_{n+1}, ..., \xi_k\right] = \left[\prod_{i=n}^{k} \overline{\mathsf{A}}(\xi_i)\right] y, \quad \text{for all } y \in \mathcal{Y}. \quad (14)$$

We note that $\overline{Y}_n := E[Y_n|\mathcal{F}_n]$ evolves according to a linear stochastic difference equation as (1) implies

$$\overline{Y}_{n+1} = [\overline{\mathsf{A}}(\xi_n)]\overline{Y}_n + B_n. \quad (15)$$

Lemma 2. Assume that (ξ_n, B_n) is a stationary ergodic sequence. Assume A1, and that

$$E[\log \left\|\overline{\mathsf{A}}(\xi_0)\right\|] < 0 \qquad \text{and} \qquad E[(\log \|B_0\|)^+] < \infty. \quad (16)$$

Then A3 implies (2) and (7). Moreover, if A2 holds with respect to $\hat{A}_n(\cdot, \xi)$ for each $\xi \in \Xi$, and A3(i) holds then A3(ii) holds too.

`Proof.` (16) implies that $\overline{\lim}_{n\to\infty} E[\overline{Y}_n] < \infty$ and that P-almost surely, $\lim_{n\to\infty} \left\|\prod_{i=0}^{n-1} \overline{\mathsf{A}}(\xi_i)\right\| = 0$, see [15]. We obtain $\overline{\lim}_{n\to\infty} E[Y_n] < \infty$ since $E[Y_n] = E[\overline{Y}_n]$, from which (2) follows. We also obtain (7) due to (14). A3(ii) follows from the infinitely divisible property. ∎

3. Queueing with correlated vacations

We focus on the gated vacation model. We then extend it to a polling system with the globally gated regime. We finally discuss briefly the case of the (repeated) exhaustive vacation discipline.

Throughout, we consider Poisson arrivals and independent service times. These assumptions are crucial for applying our formalism. In particular, they will always imply that $A_n(\cdot)$ are i.i.d. hence Lemma 1 can be used.[2] In particular, the reason that the A_n will have the infinitely divisible property will be related to the fact that for the

[2] see however Remark 4 for more general arrival and service processes falling into the category of Lemma 2.

Poisson arrival process, the number of arrivals during any given time interval can always be written as the sum of arrivals during any disjoint partition of that interval to subintervals and the arrivals to each subinterval is again Poisson.

We consider an M/G/1 queue with infinite buffer with vacations. Arrivals are Poisson with rate λ. Let $\{D_n\}$ be the sequence of service times assumed to be i.i.d., generally distributed as a random variable D with first and second moments denoted by d and $d^{(2)}$ respectively, and with LST $D^*(s)$.

3.1. The gated regime

A gating regime is considered, so that at the $(n-1)$th end of vacation, a gate is closed; this is called the the nth polling instant. At that instant the server goes on serving the customers present at the queue at that polling instant. We denote by X_n their number. Once they are served, the server leaves on vacation, whose duration is denoted by V_n with finite first and second moments v and $v^{(2)}$, respectively; we assume that V_n is a stationary ergodic sequence. Define the nth cycle to be the period between the end of the $(n-1)$th vacation till the end of the nth vacation. Cycle n is thus composed of a busy period which we denote by S_n (its duration is zero whenever $X_n = 0$) followed by a vacation of duration V_n. We denote the total duration of cycle n by C_n.

Our goal is to study correlated vacations. Introduce the correlation function:

$$r(n) = E[V_0 V_n]$$

We assume throughout that $\rho := \lambda d < 1$. This implies the conditions of Lemma 1 so we can use Theorem 2.

3.2. The buffer occupancy

From standard balance arguments, it follows [5, 16] that in steady state,

$$x := E[X_n] = \frac{\lambda v}{1-\rho}. \tag{17}$$

In order to obtain a recurrence equation of the form (1), we introduce the following objects:

- B_n:= the number of arrivals during the nth vacation.
- $A_n(X_n)$:= the number of arrivals during the service time of X_n customers (i.e. the number of arrivals during the period S_n).

With these notations (1) holds with X_n replacing Y_n. We note that assumption A2 holds and thus A1 holds as well.

Remark 4. We could allow the rate of the Poisson process as well as the service time distributions to vary from one cycle to another. For example, we may assume that each time the server returns from vacation it might be in some new state, and thus service times in the next cycle may have another distribution. This situation can be handled using the framework of Lemma 2 where assumptions A3 are seen to hold. Note also that we could allow a Poisson batch arrival instead of a simple Poisson arrival process for our framework to hold.

Theorem 3. The second moment of X_n is given by

$$E[X_n^2] = \frac{\lambda^3 v d^{(2)}}{(1-\rho)^2(1+\rho)} + \frac{\lambda v}{1-\rho} + \frac{\lambda^2(r(0) + 2\sum_{j=1}^{\infty} \rho^j r(j))}{1-\rho^2} \tag{18}$$

`Proof.` To compute the second moment of X_n, we recall the following relations. If N is a random variable independent of the sequence D_n, and

$$\tau(N) := \sum_{i=1}^{N} D_i$$

then

$$E[\tau(N)^2] = E[N^2]d^2 + E[N](d^{(2)} - d^2). \tag{19}$$

Let $\mathcal{N}(T)$ denote the number of arrivals during a random duration T, where the arrival process is Poisson with rate λ, and is independent of T. Then

$$E[\mathcal{N}(T)^2] = \lambda^2 E[T^2] + \lambda E[T]. \tag{20}$$

In particular, if we now take $T = \tau(N)$, then

$$E[\mathcal{N}(\tau(N))]^2 = \lambda^2 \Big[E[N^2]d^2 + E[N](d^{(2)} - d^2)\Big] + \lambda d E[N]. \tag{21}$$

We also mention that if we take an arbitrary T and choose $N = \mathcal{N}(T)$, then we get from (19)-(20)

$$\begin{aligned} E[\tau(\mathcal{N}(T))^2] &= d^2(\lambda^2 E[T^2] + \lambda E[T]) + \lambda E[T](d^{(2)} - d^2) \\ &= d^2\lambda^2 E[T^2] + \lambda E[T] d^{(2)}. \end{aligned} \tag{22}$$

which will be used later on.

Substituting (1) and using relations (19)-(21), we get

$$\begin{aligned} E[X_{n+1}^2] &= E[A_n(X_n)^2] + E[B_n^2] + 2E[A_n(X_n)B_n] \\ &= \lambda^2 \Big(E[X_n^2]d^2 + E[X_n](d^{(2)} - d^2)\Big) + \lambda d E[X_n] \\ &\quad + (\lambda^2 v^{(2)} + \lambda v) + 2\lambda d E[X_n B_n] \end{aligned}$$

In stationary regime, we have $E[X_n^2] = E[X_{n+1}^2]$. Hence:

$$(1-\rho^2)E[X_0^2] = E[X_0](\lambda^2(d^{(2)}-d^2)+\lambda d)+(\lambda^2 v^{(2)}+\lambda v)+2\lambda d E[X_0 B_0] \tag{23}$$

To complete the above, we compute $E[X_n B_n]$ by substituting (9) to obtain for the stationary regime:

$$\begin{aligned} E[X_0 B_0] &= E\left(\sum_{j=0}^{\infty}\left(\prod_{i=-j}^{-1} A_i^{(-j)}\right)(B_{-j-1})B_0\right) \\ &= \sum_{j=0}^{\infty}\rho^j E[B_{-j-1}B_0] = \lambda^2\sum_{j=0}^{\infty}\rho^j r(j+1). \end{aligned}$$

Substituting this as well as (17) into (23) yields

$$(1-\rho^2)E[X_n^2] = \frac{\lambda v(\lambda^2 d^{(2)} - \rho^2 + 1)}{1-\rho} + \lambda^2\Big(r(0) + 2\sum_{j=1}^{\infty}\rho^j r(j)\Big)$$

■

3.3. The cycle times

Denote by $\hat{\mathcal{A}}_n(C_n) = \tau(\mathcal{N}(C_n))$, i.e. the sum of service times of all the arrivals during C_n. Then we get the following recurrence equations which have again the form of (1):

$$C_{n+1} = \hat{\mathcal{A}}_n(C_n) + V_{n+1}. \tag{24}$$

Here again, the dynamics satisfies assumptions A1 and A2.

Theorem 4. The first and second moments of C_n in stationary regime are given by

$$E[C_n] = dE[X_n] + v = \frac{v}{1-\rho}.$$

$$E[C_n^2] = \frac{1}{(1-\rho^2)}\left(\frac{\lambda v d^{(2)}}{1-\rho} + r(0) + 2\sum_{j=1}^{\infty}\rho^j r(j)\right). \tag{25}$$

`Proof.` The first moment is directly obtained from (17) (see also [5, 16]). Using (24) as well as (22), we get

$$\begin{aligned} E[C_{n+1}^2] &= E[\hat{\mathcal{A}}_n(C_n)^2] + v^{(2)} + 2E[\hat{\mathcal{A}}_n(C_n)V_{n+1}] \\ &= \Big(\rho^2 E[C_n^2] + \lambda E[C_n]d^{(2)}\Big) + v^{(2)} + 2E[\hat{\mathcal{A}}_n(C_n)V_{n+1}]. \end{aligned}$$

Moreover, using again (9) and the fact that we are in stationary regime, we get

$$E[\hat{\mathcal{A}}_n(C_n)V_{n+1}] = E[\hat{\mathcal{A}}_0(C_0)V_1] = E\left[\hat{\mathcal{A}}_0\left(\sum_{j=0}^{\infty}\left(\prod_{i=-j}^{-1}\hat{\mathcal{A}}_i^{(-j)}\right)(V_{-j})\right)V_1\right]$$
$$= \rho\sum_{j=0}^{\infty}\rho^j E[V_{-j}V_1] = \sum_{j=1}^{\infty}\rho^j r(j).$$

Substituting this in the previous equation, we obtain (25). ∎

3.4. The expected waiting time and queue size

Theorem 5. The expected waiting time of an arbitrary customer is given by

$$E[W_n] = (1+\rho)\frac{E[C_0^2]}{2E[C_0]}, \tag{26}$$

where $E[C_0^2]$ and $E[C_0]$ are given in Theorem 4. The expected number of customers in queue in stationary regime (not including service) is $\lambda E[W_n]$.

`Proof.` Consider an arbitrary customer. Upon arrival, it has to wait for

1. The residual cycle time, which we denote by C_{res},
2. The service time of all the customers that arrived during C_{past} which is the past cycle time.

We have from [7]:

$$E[C_{res}] = E[C_{past}] = \frac{E[C_0^2]}{2E[C_0]}.$$

Thus the expected waiting time of an arbitrary customer is given by

$$E[W_n] = (1+\rho)\frac{E[C_0^2]}{2E[C_0]},$$

where $E[C_0^2]$ and $E[C_0]$ are given in Theorem 4. Finally, the expected number of customers is obtained using Little's Theorem. ∎

3.5. The globally gated polling system

We now extend the previous results to the case of a polling system with a globally gated regime (introduced in [11]). There are M queues, and the arrivals at queue i are Poisson with rate λ_i. The service times of customers arriving at queue i are i.i.d. with first and second moments denoted by d_i and $d_i^{(2)}$, respectively. Define $\rho_i = \lambda_i d_i$ and let $\rho := \sum_{i=1}^M \rho_i$. Let $\lambda := \sum_{i=1}^M \lambda_i$.

When the server arrives at queue 1 at the nth time, a gate is simultaneously closed in all queues. Let X_n^i be the number of customers found at queue i at that moment. The queues are now visited according to the cyclic order: 1,2,...,M. In each queue, the X_n^i customers present at the instant that the gate closed are served. When completing serving queue i the server requires some walking time which we denote by V_n^i (during which it idles) to move to the next queue.

Let $\hat{V}_n := \sum_{i=1}^M V_n^i$ and $X_n := \sum_{i=1}^M X_n^i$. Denote as before C_n to be the nth cycle, where a cycle is defined to be the duration between the nth and the $n+1$st arrival of the server at queue 1.

We assume that the arrival sequences, walking times and service times are all independent.

Let us now denote by $\hat{D}_n$ the service time of the nth customer that arrives to the system and let $\hat{d}$ and $\hat{d}^{(2)}$ be its two first moments. With probability $p_i = \lambda_i/\lambda$ it is a customer that goes to the queue i. Thus

$$\hat{d} = \sum_{i=1}^M p_i d_i, \qquad \hat{d}^{(2)} = \sum_{i=1}^M p_i d_i^{(2)}.$$

It is now easy to see that the processes C_n and X_n with $X_0 = 0$ are the same as those of the single queue with gated regime studied in the previous subsections, with the vacation sequence $V_n = \hat{V}_n$, service times $D_n = \hat{D}_n$, and in which we have the same arrivals. Below we shall omit the "ˆ".

We assume that V_n are stationary ergodic, and that $\rho < 1$. The latter is known to be the sufficient condition for a stationary regime to exist [14]. We conclude that

Theorem 6. The first moments in the stationary regime are given by

$$E[C_0] = \frac{v}{1-\rho}, \qquad E[X_0] = \lambda E[C_0], \qquad E[X_0^i] = \lambda_i E[C_0].$$

The second moment of X_n and of C_n in the stationary regime are given by (18) and (25), respectively. Finally,

$$E[(X_n^i)^2] = \lambda_i E[C_n^2] + \lambda_i E[C_n].$$

Proof. We already explained the moments of X_n and C_n. The last equation follows from (20). ∎

As in [11], we can express the waiting time of an arbitrary customer that arrives at queue i as the sum of

1. The residual cycle time C_{res},
2. The service time of all the customers that arrive at queue j for all $j < i$ during both the past and the residual cycle time,
3. The service time of all those that arrive at queue i during the past cycle time C_{past}.
4. The total walking times which we denote by V_+^i, $i = 1, ..., M$, that will be incurred during the next cycle from queue 1 till queue i.

Till now we have not specified any probabilistic assumptions on the specific V_n^i (only on their sum V_n). Yet we need to do so in order to be able to express the last component of the waiting time.

We shall assume henceforth that $(V_n^1, V_n^2, ..., V_n^M)$ are stationary ergodic, $n = ..., -2, -1, 0, 1, 2, ...$, and denote $r(j,i) = E[V_0 V_j^i]$.

We note for the last point that since the cycle time duration as seen by an arbitrary customer is not distributed like C_n (it has a stationary distribution, rather than the Palm distribution), then the walking times within that cycle do not have the same distribution as V_n^i. Since V_{n+1}^i may depend on V_n, then the walking times in the next cycle need not be typical as well and their distribution may be different than that of V_n^i.

Consider the stationary ergodic point process N where the nth point corresponds to the beginning of the nth cycle, defined on a probability space $(\Omega, \mathcal{F}, P^s)$; let θ_t be the corresponding measurable flow,

Denote by E^s the expectation with respect to the time stationary probability measure, i.e. the measure corresponding to the process as seen at an arbitrary moment (or due to PASTA, as seen by an arrival), and denote by E_N^0 the expectation with respect to the Palm probability associated with the point process N.

Using Remark 4.2.1 in [7], we have

$$\begin{aligned} E[V_+^i] &= E^s[V_1] = E^s(E[V_1^i|C_0]) = \frac{E_N^0(C_0 E[V_1^i|C_0])}{E_N^0[C_0]} = \frac{E_N^0(C_0 V_1^i)}{E_N^0[C_0]} \\ &= \frac{E\left(\sum_{j=0}^{\infty}\left(\prod_{i=-j}^{-1}\hat{\mathcal{A}}_i^{(-j)}\right)(V_{-j})V_1^i\right)}{E_N^0[C_0]} = \frac{1-\rho}{v}\sum_{j=0}^{\infty}\rho^j r(j,i) \end{aligned}$$

We thus obtain:

Theorem 7. The expected waiting time of an arbitrary customer that arrives at queue i is given by

$$E[W_n^i] = \frac{E[C_0^2]}{2E[C_0]} \left(1 + 2\sum_{j<i} \rho_j + \rho_i\right) + \frac{1-\rho}{v} \sum_{j<i} \sum_{k=0}^{\infty} \rho^k r(k,j)$$

where $E[C_0]$ and $E[C_0^2]$ are given in Theorem 4.

3.6. Exhaustive service

We consider again the case of a single queue with vacations. At each end of vacation, the server stays at the queue as long as it is nonempty. Once it is empty, it leaves for a vacation. If upon completion of the vacation the queue is empty then immediately another vacation is taken, and so on.

In this section we show that the vacation model under the exhaustive service discipline has an insensitivity property: the marginal distribution of the queue length is insensitive to the correlation between vacations, unlike the case of the gated discipline.

The nth cycle is defined for the case of exhaustive service as a duration of the nth vacation followed by the busy period S_n that follows it. If no arrivals occurred during the vacation then the cycle contains only the vacation. We may express S_n as

$$S_n = \Theta_n(\mathcal{N}_n(V_n)),$$

where $\Theta(N)$ denotes a busy period duration in an M/G/1 queue that starts with N customers present.

The duration of the nth cycle is thus given by

$$C_n = V_n + \Theta_n(\mathcal{N}_n(V_n)).$$

We see from this expression that the distribution of the nth cycle duration depends on the the process of the durations of the vacations only through the duration of the nth vacation. We see that it is unchanged if we alter the vacation process to an i.i.d. vacation sequence having the same marginal distribution as the previous vacation sequence!

This phenomenon can be explained by the fact that the dynamics has a very degenerated form of (1), as the "A_n" there is identically zero. We can identify the right hand side of the previous equation as the B_n that appears in (1).

Mean queue length. Denote by Z_t the queue length at time t. We shall use (eq. 4.3.2) in [7] to compute the stationary distribution of the queue length. More precisely, the expectation of any function g of the queue length in stationary regime is given by:

$$E^s[g(Z_0)] = \frac{E_N^0\left[\int_0^{T_1} g(Z_t)dt\right]}{E_N^0[T_1]}$$

where in the right hand side we consider the Palm probability according to which a cycle starts at time 0 and ends at time T_1. It is now easily seen that this expression does not depend on the joint distribution of walking times between different cycles. We conclude that the expected queue length is the same as the one obtained for the vacation model in which we change the vacation times to i.i.d. vacations having the same marginal distribution!

4. Polling systems

We already discussed the globally gated polling system. Although it contains M queues, we were able to obtain some of the performance measures related to that system using one dimensional recursive dynamics.

In this section we consider multidimensional dynamics that describe the gated and the exhaustive polling systems. We remark that a huge literature exist on polling system. However, except for stability results [14], the only existing analytical results in which vacations are allowed to be correlated are those on expected cycle times and expected number of queued customers at polling instants (see [5]). For extensive survey on polling, for which vacations are independent, see [19, 20, 22].

There are M queues, and the arrivals at queue i are Poisson with rate λ_i. The services of customers arriving at queue i are i.i.d. with first and second moments denoted by d_i and $d_i^{(2)}$, respectively. Define $\rho_i = \lambda_i d_i$ and let $\rho := \sum_{i=1}^M \rho_i$.

When completing serving the nth queue ($n = 1, 2, ..., M, M+1, ...$) that it visits, the server requires some walking time which we denote by V_n (during which it idles) to move to the next queue. We denote by S_n the busy period in the nth visited queue. The queues are visited according to the cyclic order: 1,2,...,M. We define the nth cycle as the time between the arrival of the server at the nth queue that it visits, till the next time it arrives at that queue. We denote the duration of that cycle by C_n.

We consider two possible regimes: the gated and the exhaustive.

4.1. Polling systems with gated service

At the nth time that a server arrives at a queue, it serves the X_n^i customers found there upon arrival. Define $\mathbf{X_n} := (X_n^1, X_n^2, ..., X_n^M)$, $n = 1, 2, ..., M, M+1, M+2, ...$

Buffer occupancy. Let $A_n^i(X_n^j)$ denote the number of arrivals at queue i during the service time of X_n^j customers in queue j. Let $j(n)$ be the queue that is polled at the nth polling instant. Then

$$\mathbf{X}_{n+1} = Q_n(\mathbf{X}_n) + \mathbf{B}_n \tag{27}$$

where

$$Q_n^i(\mathbf{X}_n) = \begin{cases} X_n^i + A_n^i(X_n^{j(n)}) & \text{for } i \neq j(n) \\ A_n^i(X_n^i) & \text{for } i = j(n), \end{cases}$$

$$\mathbf{B}_n^i = \mathcal{N}_n^i(V_n), \qquad \mathbf{B}_n = (\mathbf{B}_n^1, \mathbf{B}_n^2, ...\mathbf{B}_n^M),$$

and where $\mathcal{N}_n^i(V_n)$ denotes the number of arrivals at queue i during the walking time V_n.

We do not know of a norm for which the condition (11) holds; moreover, the sequence $(A_n(\cdot), B_n)$ is in general nonstationary, since there is a periodicity (of period M) in the probability distribution due to the cyclic structure of the polling. However, by iterating M times (27) we get a stationary ergodic structure for the dynamics:

$$\mathbf{X}_{n+M} = \hat{Q}_n(\mathbf{X}_n) + \hat{\mathbf{B}}_n. \tag{28}$$

Here, $\hat{Q}_n(\cdot) = (Q_{n+M-1}Q_{n+M-2}...Q_n)(\cdot)$, whereas

$$\begin{aligned} \hat{\mathbf{B}}_n^i \;=\; & A_{n+M-1}^i\left(\sum_{i=0}^{M-1} V_{n+i}\right) + A_{n+M-1}^i A_{n+M-2}^{i-1}\left(\sum_{i=0}^{M-2} V_{n+i}\right) \\ & +...+ A_{n+M-j}^i A_{n+M-2}^{i-1} \cdots A_n^{i-M}\left(V_n\right) \end{aligned}$$

We may assume without loss of generality that $j(0) = M$; in that case (28) can be rewritten as

$$\mathbf{X}_{(n+1)M} = \hat{Q}_{nM}(\mathbf{X}_{nM}) + \hat{\mathbf{B}}_{nM}. \tag{29}$$

This stochastic recursive equation is now of type of equation (1), where the sequence $\{\hat{Q}_{nM}, \hat{\mathbf{B}}_{nM}\}_{n\in Z}$ is stationary ergodic.

We now consider the following norm:

$$\|y\| := \sum_{i=1}^{M} d_i y_i. \tag{30}$$

Using this norm, it can be shown that for $\rho < 1$, $\|\mathsf{A}\| \leq \alpha$ for some $\alpha < 1$ (this follows from the proof of Prop. 3.5 in [5]; note that in the model there, vacations are independent; however the computation does not rely on that independence, especially since only data concerning a single cycle is involved in the relation $\|\mathsf{A}\| \leq \alpha$. The condition in Remark 3 and 1 hold, so we can indeed apply Theorem 2.

Station times. We briefly discuss an alternative representation of the dynamics in the form of our stochastic recursive equations (1). Let $\hat{\mathcal{A}}_n(C_{n-M})$ denote again the sum of service times of all those who arrive at queue n during cycle $n-M$.

$$S_n = \hat{\mathcal{A}}_n(C_{n-M}) = \hat{\mathcal{A}}_n\left(\sum_{j=1}^{M}(S_{n-j}+V_{n-j})\right)$$

Denote the vector of "station times": $Y_n = (Y_n^1, Y_n^2, ..., Y_n^M)$ where $Y_n^j = S_{n-j} + V_{n-j}$. Then we have:

$$Y_n = \mathcal{Q}_n(Y_{n-1}) + V_n,$$

where $\mathcal{Q}_n(Y_{n-1}) = \left(\mathcal{Q}_n^1(Y_{n-1}), ..., \mathcal{Q}_n^M(Y_{n-1})\right)$ and

$$\mathcal{Q}_n^1(Y_{n-1}) = \hat{\mathcal{A}}_n\left(\sum_{j=1}^{M} Y_{n-1}^j\right), \qquad \mathcal{Q}_n^j(Y_{n-1}) = Y_{n-1}^{j-1}, j = 2, ..., M.$$

Again we need to iterate these equations over a whole period (i.e. over M steps) in order to get stochastic recursive equations as (1) with stationary ergodic data, since without it we have a periodic structure.

To establish condition (11) assuming $\rho < 1$, one may use the computations in [6].

4.2. Polling systems with exhaustive service

At the nth time that a server arrives at a queue, it stays there until that queue empties. It thus serves all the X_n^i customers found there upon arrival as well as all those that arrive there during its service time there.

In the case of a single queue served exhaustively, the dynamics had a very degenerated form of (1), as the A_n function was zero. In the polling system, however, the function corresponding to A_n is trivial.

Let $\Theta_n(N)$ be the duration of a busy period in the nth queue visited, given that N customers are found there at the beginning of service there. This corresponds to the sum of N independent busy periods in the corresponding M/G/1 queue each starting with a single customer.

Let $\mathcal{N}_n^i(T_n)$ denote the number of arrivals at queue i during some disjoint random periods T_n (that will be precised below).

Buffer occupancy. The buffer occupancies at polling instants evolve according to

$$X_{n+1}^i = \begin{cases} X_n^i + \mathcal{N}_n^i(\Theta_n(X_n^{j(n)}) + V_n) & \text{for } i \neq j(n) \\ \mathcal{N}_n^i(V_n) & \text{for } i = j(n) \end{cases}$$

Thus we can write

$$\mathbf{X}_{n+1} = R_n(\mathbf{X}_n) + \beta_n \tag{31}$$

where

$$R_n(\mathbf{X}_n) = (R_n^1(\mathbf{X}_n), ..., R_n^M(\mathbf{X}_n)),$$

$$R_n^i(\mathbf{X}_n) = \begin{cases} X_n^i + \mathcal{N}_n^i(\Theta_n(X_n^{j(n)})) & \text{for } i \neq j(n) \\ 0 & \text{for } i = j(n) \end{cases}$$

and

$$\beta_n^i = \mathcal{N}_n^i(V_n), \qquad i = 1, ..., M.$$

Again, as in the case of the gated regime, the above dynamics has a periodicity in the distribution. In order to apply Theorem 2 we need to iterate (31) M times; this allows to obtain the stationary ergodic framework of equation (1).

The conditions of Lemma 1 hold for $\rho < 1$, so we can apply Theorem 2. This is a direct consequence from the fact that it holds for the gated regime, and that the workload of the gated regime (after one cycle) is stochastically greater than that of the exhaustive regime [17].

Station times. We have

$$S_n = \Theta_n \left(\mathcal{N}_n^{j(n)} \left(V_{n-M} + \sum_{i=1}^{M-1} S_{n-i} + V_{n-i} \right) \right)$$

Thus, if we write $\mathbf{S}_n = (S_n, S_{n-1}, ..., S_{n-M+1})$, then the dynamics can be written as

$$\mathbf{S}_{n+1} = \mathcal{R}_n(\mathbf{S}_n) + \mathcal{B}_n, \tag{32}$$

where $\mathcal{R}_n = (\mathcal{R}_n^1, ..., \mathcal{R}_n^M)$ is given by

$$\mathcal{R}_n^j(\mathbf{S}_n) = \begin{cases} \tilde{\Theta}_{n+1,1} \left(\tilde{\mathcal{N}}_{n+1,1}^{j(n+1)} \left(\sum_{i=0}^{M-1} S_{n-i} \right) \right) & \text{for } j = 1 \\ S_{n-j+1} & \text{for } j > 1. \end{cases}$$

and where $\mathcal{B}_n = (\mathcal{B}_n^1, ..., \mathcal{B}_n^M)$ is given by

$$\mathcal{B}_n^j = \begin{cases} \tilde{\Theta}_{n+1,2} \left(\tilde{\mathcal{N}}_{n+1,2}^{j(n+1)} \left(\sum_{i=0}^{M-1} V_{n-i} \right) \right) & \text{for } j = 1 \\ 0 & \text{for } j > 1. \end{cases}$$

Above, $\tilde{\mathcal{N}}_{n,1}^{j(n)}$ and $\tilde{\mathcal{N}}_{n,2}^{j(n)}$ are two independent stochastic processes both with the same distribution as $\mathcal{N}_n^{j(n)}$ satisfying

$$\begin{aligned}&\mathcal{N}_n^{j(n)}\left(V_{n-M}+\sum_{i=1}^{M-1} S_{n-i}+V_{n-i}\right)\\&=\tilde{\mathcal{N}}_{n,1}^{j(n)}\left(\sum_{i=0}^{M-1} S_{n-i}\right)+\tilde{\mathcal{N}}_{n,2}^{j(n)}\left(\sum_{i=0}^{M-1} V_{n-i}\right)\end{aligned}$$

(This is the infinitely divisible property of the Poisson process; The above equation reflects the fact that the number of arrivals can be written as the sum of the arrivals during the corresponding two disjoint epochs.) Similarly, $\tilde{\Theta}_{n,1}(\cdot)$ and $\tilde{\Theta}_{n,2}(\cdot)$ are two independent discrete time processes with the same distribution as $\Theta_n(\cdot)$ such that

$$\Theta_n(\cdot)=\tilde{\Theta}_{n,1}(\cdot)+\tilde{\Theta}_{n,2}(\cdot).$$

They correspond to the duration of busy periods generated by the arrivals during the time that the server is busy serving, and during vacation periods, respectively.

Again by iterating (32) M times, we are back to the framework of the stochastic recursive equation (1) with stationary ergodic coefficients.

5. Concluding Remarks

We have obtained in this paper some explicit expressions for the distribution of the stationary regime of a special class of stochastic recursive equations. We also established convergence to that stationary regime. We then applied these results to an M/G/1 queue with general correlated vacations and obtained explicit expressions for useful performance measures. We extended some of these results to polling models. Our technique can extend to more complex polling systems (for example, one may consider other polling orders [4], including non-cyclic ones and randomized order of polling [21], and other gating mechanisms [4, 16]). Batch arrivals can also be handled in our framework, and in particular correlated arrivals, i.e. at the nth arrival, a batch of size K_n^i arrives at queue i, where for a given n and different i's, K_n^i need not be independent. We leave these and other issues to future research.

The explicit results that we obtained for vacation models were inspired by previous results that were obtained in applying *linear* stochastic recursive equations to analyze the well known TCP/IP congestion control protocol [1].

References

1. E. Altman, K. Avratchenkov and C. Barakat, "A stochastic model of TCP/IP with stationary random losses", *ACM Sigcomm*, Aug. 28 - Sept. 1, Stockholm, Sweden, 2000. (see also INRIA Research Report RR-3824).
2. E. Altman, K. Avratchenkov, C. Barakat and R. Núñez-Queija, "TCP modeling in the presence of nonlinear window growth", *Proceedings of the 17th International Teletraffic Congress*, Salvador da Bahia, Brazil, Dec., 2001.
3. E. Altman and A. Hordijk, "Applications of Borovkov's Renovation Theory to Non-Stationary Stochastic Recursive Sequences and their Control", *Advances of Applied Probability* **29**, pp. 388-413, 1997.
4. E. Altman, A. Khamisy and U. Yechiali, "On Elevator Polling with Globally Gated Regime", *Queueing Systems* **11** , (special issue on "Polling Models", Eds. H. Takagi and O. Boxma), pp. 85-90, 1992.
5. E. Altman, P. Konstantopoulos, Z. Liu, "Stability, Monotonicity and Invariant Quantities in General Polling Systems", *Queueing Systems* **11** , pp. 35-57, 1992.
6. E. Altman and F. Spieksma, "Polling systems: moment stability of station times and central limit theorems", *Stochastic Models*, **12** No. 2, 307-328, 1996.
7. F. Baccelli and P. Brémaud, *Elements of Queueing theory*, Springer-Verlag, 1994.
8. A. A. Borovkov, *Asymptotic Methods in Queueing Theory*, John Wiley & Sons, 1984 (translated from Russian).
9. A. A. Borovkov and S. G. Foss, "Stochastically recursive sequences and their generalizations", *Siberian Advances in Mathematics*, **2**, No. 1, pp. 16-81, 1992.
10. A. A. Borovkov and S. G. Foss, "Two ergodicity criteria for stochastically recursive sequences", *ACTA Applicande Math.* **34**, pp. 125-134, 1994.
11. O. J. Boxma, H. Levy and U. Yechiali, "Cyclic Reservation Schemes for Efficient Operation of Multiple-Queue Single-Server Systems", *Annals of Operations Research* **35**, pp. 187-208, 1992.
12. A. Brandt, "The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients", *Advances in Applied Probability*, Vol. 18, 1986.
13. A. Brandt, P. Franken and B. Lisek, *Stationary Stochastic Models*, Akademie-Verlag, Berlin, 1992.
14. S. Foss and N. Chernova, "Ergodic Properties of Polling Systems". Research Report 6/1995, Institute of Mathematics, Novosibirsk, 1995.
15. P. Glasserman and D. D. Yao, "Stochastic vector difference equations with stationary coefficients", *J. Appl. Prob.*, Vol 32, pp 851-866, 1995.
16. A. Khamisy, E. Altman and M. Sidi, "Polling Systems with Synchronization Constraints", *Annals of Operations Research* **35**, special issue on Stochastic Modeling of Telecommunication Systems, Eds. P. Nain and K. W. Ross, pp. 231-267, 1992.
17. H. Levy, M. Sidi, O. J. Boxma, "Dominance Relations in Polling Systems", *Queueing Systems*, **6**, pp. 155-172, 1990.
18. R. Loynes, "The stability of a queue with non-independent inter-arrival and service times", *Proc. Cambr. Phil. Soc.* **58**, No. 3, pp. 497-520, 1962.
19. H. Takagi, *Analysis of Polling Systems*, The MIT Press, 1986.
20. H. Takagi, "Queueing Analysis of Polling Models: an Update", in *Stochastic Analysis of Computer and Communications Systems*, (H. Takagi, Ed.), Elsevier Science Pub., pp. 267-318, 1990.
21. Tedijanto, "Exact Results for the Cyclic Service Queue with a Bernoulli Schedule", *Performance Evaluation*, **11**, pp. 107-115, 1990.

22. U. Yechiali, "Analysis and Control of Polling Systems", in *Performance Evaluation of Computer and Communications Systems* (L. Donantiello and R. Nelson, Eds.) pp. 630-650, Springer-Verlag, 1993.
23. W. Vervaat, "On stochastic difference equation and a representation of non-negative infinitely random variables", *Advances of Appl. Prob.*, **11**, 750-783, 1979.