On stochastic recursive equations and infinite server queues

Eitan Altman
INRIA, BP93, 06902 Sophia Antipolis, France

***Abstract*— The purpose of this paper is to investigate some performance measures of the discrete time G/G/∞ queue under a general arrival process. We assume more precisely that at each time unit a batch with a random size may arrive, where the sequence of batch sizes need not be i.i.d. All we request is that it would be stationary ergodic and that the service duration has a phase type distribution. Our goal is to obtain explicit expressions for the first two moments of number of customers in steady state. We obtain this by computing the first two moments of some generic stochastic recursive equations that our system satisfies. We then show that these class of recursive equations allow to solve not only the $G/PH/\infty$ queue but also a network of such queues. We finally investigate the process of residual activity time in a $G/G/\infty$ queue under general stationary ergodic assumptions, obtain the unique stationary solution and establish coupling convergence to it from any initial state.**

Keywords: Stochastic processes/Queueing theory.

I. Introduction

Most explicit expressions for performance measures in queueing networks are known under independence assumptions on the driving processes (service and interarrival times). An interesting challenge is to obtain explicit expressions for the case in which the independence is relaxed and only stationarity and ergodicity of some components of the driving sequences are assumed. One line of research that allows to handle stationary ergodic sequences is based on identifying measures that are insensitive to correlations. For example, the probability of finding a G/G/1 queue nonempty is just the ratio between the expected service time and the expected interarrival time (which follows directly from Little's Law). The expected cycle duration in a polling system (under fairly general condition) too, depends on the interarrival, service and vacation times only through their expectations under general stationary ergodic assumptions (see e.g. [5]). An example of performance measures that depend on the whole distribution of service times but is insensitive to correlations is the growth rate of number of customers or of sojourn time in a (discriminatory) processor sharing queue in overload [4], [13].

In this paper we study a queueing problem under a stationary ergodic arrival process, in which the correlations indeed influence the performance but in which despite the dependence between arrival times, explicit expressions are obtained for the two first moments of the stationary number of customers. More precisely, we study the discrete time G/G/∞ queue in which at each time unit a random a batch with a random size may arrive, where the sequence of batch sizes is stationary ergodic and service durations have a phase type distribution.

We first compute the two moments of some generic stochastic recursive equations that our system satisfies. These are simplified versions of stochastic recursions introduced in [2] which already enabled us to study polling systems [2], [12] and queues with vacations [2] in which vacation times are correlated, and are related to branching process with migration [1]. Yet this is the first time that generic explicit expressions are derived for the first two moments of such equations.

We then show that these class of recursive equations allow to solve not only the $G/PH/\infty$ queue but also a network of such queues.

We finally investigate the process of residual activity time in a $G/G/\infty$ queue under general stationary ergodic assumptions, obtain the unique stationary solution and establish coupling convergence to it from any initial state.

The infinite server queue which is the topic of our paper has had various applications in teletraffic and in networking modeling. The output process of an M/GI/∞ queue has been used to model long range dependent traffic, c.f. in video applicationskruntz. In [17] the connectivity of ad-hoc networks on a line has been considered. The distribution of distance covered by a connected set of mobiles has been shown to correspond to a busy period in the $GI/GI/\infty$ queue and its distribution was computed for various channel conditions. Furthermore the distribution of the number of connected mobiles has been computed using its correspondence to the number of customers served in a busy period of a $GI/GI/\infty$ queue.

Finally, the infinite server queue has also been used in the context of communication networks and distributed computer systems, see e.g. [14].

The structure of the paper is as follows. We introduce in Section II generic stochastic recursive equation corresponding to a branching type process in non Markov random environment with migration. The first and second moments of the corresponding state variables are introduced in Section III. The expressions obtained are shown to further simplify for specific Markovian dynamics that creates the correlation. This allows us to derive in Section IV explicit performance measures for the $G/PH/\infty$ discrete time queue and to provide a numerical example that illustrates the role of correlation. An extension to a whole network of infinite server queues ends this section. Further stability results for the G/G/∞ queue are presented in Section V followed by a concluding section.

II. The model

Consider a column vector Y_n whose entries are Y_n^i, $i = 1, ..., N$ where Y_n^i take values on the nonnegative integers Z^+. Consider the following stochastic recursive equation:

$$Y_{n+1} = A_n(Y_n) + B_n \tag{1}$$

where the i the element of the column vector $A_n(Y_n)$ is given by

$$[A_n(Y_n)]_i = \sum_{j=1}^{N} \sum_{k=1}^{Y_n^j} \xi_{ji}^{(k)}(n) \tag{2}$$

where $\xi^{(k)}(n)$, $k = 1, 2, 3, ...$, $n = 1, 2, 3, ...$ are i.i.d. random matrices of size $N \times N$. Each of its element is a nonnegative integer. Denote $E[\xi_{ij}^{(k)}(n)] = \mathsf{A}_{ji}$. The N-dimensional vector B_n is a stationary ergodic stochastic whose entries $B_n^i, i = 1, ..., N$ are nonnegative integers.

$A_n(y)$ has a divisibility property: if for some k, $y = y^0 + y^1 + ...y^k$ where y^m are integers, then $A_n(y)$ can be represented as

$$A_n(y) = \sum_{i=0}^{k} A_n^{(i)}(y^i)$$

where $\{A_n^{(i)}\}_{i=0,1,2,...,k}$ are i.i.d. with the same distribution as $A_n(\cdot)$. Note also that $A_n(0) = 0$. The divisibility property allows us to use the framework of [2] to characterize the distribution of Y_n and its limiting behavior.

We shall understand below $\prod_{i=n}^{k} A_i(x) = x$ whenever $k < n$, and $\prod_{i=n}^{k} A_i(x) = A_k A_{k-1} ... A_n$ whenever $k > n$.

We note that although (1) is not linear in Y_n, it is linear in expectation; if we let y be a column vector then

$$E[A_n(y)] = \mathsf{A}y. \tag{3}$$

Moreover, we have for $j > 1$ by Wald's equation

$$E\left[\left(\prod_{i=1}^{j} A_j\right)(y)\right] = \mathsf{A}^j y \tag{4}$$

We make the following assumptions throughout the paper:

A1: $\|\mathsf{A}\| < 1$ and $E[\log \|B_0\|] < \infty$ where $\|\mathsf{A}\|$ stands for the largest absolute value of the eigenvalues of A and where $\|B_0\|$ stands for the maximum absolute value of the elements of B_0.

Note that by Jensen's inequality, a sufficient for $E[\log \|B_0\|] < \infty$ is that $E[\|B_0\|] < \infty$.

We recall the following property of our system:

Theorem 1: (i) For $n > 0$, Y_n can be written in the form $Y_n =$

$$\sum_{j=0}^{n-1} \left(\prod_{i=n-j}^{n-1} A_i^{(n-j)}\right)(B_{n-j-1}) + \left(\prod_{i=0}^{n-1} A_i^{(0)}\right)(Y_0) \tag{5}$$

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

$$Y_n^* =_d \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} A_i^{(n-j)}\right)(B_{n-j-1}), \qquad n \in Z, \tag{6}$$

The sum on the right side of (6) 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 $n \to \infty$.

Proof. (5) is obtained by iterating (1). Theorem 2 and Lemma 1 in [2] imply (ii). ∎

III. First and second moments

Denote by y_i and $y_i^{(2)}$ the first and second moment of the ith element of Y_n^*. Denote $cov(Y)_{ij} = E[(Y_0^*)_i (Y_0^*)_j] - y_i y_j$. Let b_i and $b_i^{(2)}$ denote the two first moments of B_n^i. Denote $cov(\xi)_{jk}^i = E(\xi_{ij}^{(0)} \xi_{ik}^{(0)}) - \mathsf{A}_{ji}\mathsf{A}_{ki}$ and define the following $N \times N$ matrices:

$\mathcal{B}(k)$ is the matrix whose ijth entry equals $E[B_0^i B_k^j]$, where k is an integer.

$\hat{B}$ is the matrix whose ijth entry equals $b_i b_j$,

$cov(B)$ is the matrix whose ijth entry equals $E[B_0^i B_0^j] - b_i b_j$.

Define $\hat{\mathcal{B}}(k) := \mathcal{B}(k) - \hat{B}$.

A. General results

Theorem 2: (i) The first moment of Y_n^* is given by

$$E[Y_0^*] = (I - \mathsf{A})^{-1} b, \tag{7}$$

(ii) Assume that the first and second moments b_i and $b_i^{(2)}$'s are finite. Define Q to be the matrix whose (ij)th entry is

$$Q_{ij} = \sum_{k=1}^{N} y_k (cov(\xi)_{ij}^k).$$

Then the matrix $cov(Y^*)$ is the unique solution of the set of linear equations:

$$\begin{aligned} cov(Y) &= cov(B) + \sum_{r=1}^{\infty} \left(\mathsf{A}^r \hat{\mathcal{B}}(r) + \left[\mathsf{A}^r \hat{\mathcal{B}}(r) \right]^T \right) \\ &\quad + \mathsf{A} cov(Y) \mathsf{A}^T + Q. \end{aligned} \tag{8}$$

The second moment matrix $E[YY^T]$ in steady state is the unique solution of the set of linear equations:

$$\begin{aligned} E[YY^T] &= E[B_0 B_0^T] + \sum_{r=1}^{\infty} \left(\mathsf{A}^r \mathcal{B}(r) + \left[\mathsf{A}^r \mathcal{B}(r) \right]^T \right) \\ &+ \mathsf{A} E[YY^T] \mathsf{A} + Q_{ij}. \end{aligned} \tag{9}$$

Remark 1: Note that the sums both in (8) as well as in (9) are finite since the finiteness for all i of the second moments $b_i^{(2)}$ implies that $\mathcal{B}(j)$ are uniformly bounded and since $\|A\| < 1$. Note also that if for some i, $b_i^{(2)}$ is infinite then it follows directly from (1) that $E([Y_n]_i^2)$ is infinite for all $n > 0$ and thus also in the stationary regime.

Proof of Theorem 2. (i) Taking the first moment at stationary regime of (1) we obtain (7).

(ii) To obtain the covariance, we first compute

$$\begin{aligned} & E[(A_0(Y_0))_i (A_0(Y_0))_j] \\ &= E(E[(A_0(Y_0))_i (A_0(Y_0))_j | Y_0]) \\ &= E(\sum_{k=1}^{N} Y_0^k \mathsf{A}_{ki} \sum_{m \neq k} Y_0^m \mathsf{A}_{jm}) \\ &\quad + E \left(\sum_{k=1}^{N} E \left[\sum_{r=1}^{Y_0^k} \sum_{s=1}^{Y_0^k} \xi_{ki}^{(r)} \xi_{kj}^{(s)} \middle| Y_0^k \right] \right) \\ &= \sum_{k=1}^{N} \sum_{m \neq k} \mathsf{A}_{ik} \mathsf{A}_{jm} E[Y_0^k Y_0^m] \\ &\quad + E \left(\sum_{k=1}^{N} E \left[\sum_{r=1}^{Y_0^k} \sum_{s=1, s \neq r}^{Y_0^k} \xi_{ki}^{(r)} \xi_{kj}^{(s)} \middle| Y_0^k \right] \right) \\ &\quad + E \left(\sum_{k=1}^{N} E \left[\sum_{r=1}^{Y_0^k} \xi_{ki}^{(r)} \xi_{kj}^{(r)} \middle| Y_0^k \right] \right) \\ &= \sum_{k=1}^{N} \sum_{m \neq k} \mathsf{A}_{ik} \mathsf{A}_{jm} E[Y_0^k Y_0^m] \\ &\quad + \sum_{k=1}^{N} [(y_k^{(2)} - y_k)] \mathsf{A}_{ik} \mathsf{A}_{jk} + \sum_{k=1}^{N} y_k E[\xi_{ki}^{(0)} \xi_{kj}^{(0)}] \\ &= \sum_{k=1}^{N} \sum_{m=1}^{N} \mathsf{A}_{ik} \mathsf{A}_{jm} E[Y_0^k Y_0^m] + \sum_{k=1}^{N} y_k cov(\xi)_{ij}^k \end{aligned}$$

and, with $\mathbf{B_0^-} := (B_0, B_{-1}, B_{-2}, ...)$ we further compute

$$\begin{aligned} E[(Y_0)_i B_0^r] &= \sum_{j=0}^{\infty} E \left\{ \left[\left(\prod_{i=-j}^{-1} A_i^{(-j)} \right) (B_{-j-1}) \right]_i B_0^r \right\} \\ &= \sum_{j=0}^{\infty} E \left(E \left\{ \left[\left(\prod_{i=-j}^{-1} A_i^{(-j)} \right) (B_{-j-1}) \right]_i B_0^r \right\} \middle| \mathbf{B_0^-} \right) \\ &= \sum_{j=0}^{\infty} E \Big((\mathsf{A}^j B_{-j-1})_i B_0^r \Big) \\ &= \sum_{j=0}^{\infty} \sum_{s=1}^{N} (\mathsf{A}^j)_{is} \mathcal{B}(j+1)_{s,r} \end{aligned}$$

where the last equality follows from (4). Note that the last sum is finite since the finiteness for all i of the second moments $b_i^{(2)}$ implies that $\mathcal{B}(j)$ are uniformly bounded and since $\|A\| < 1$. Next we compute

$$\begin{aligned} E[(A_0(Y_0))_i B_0^r] &= E\Big[((A_0(Y_0))_i B_0^r \Big| Y_0, B_0 \Big] \\ &= \sum_{k=1}^{N} \mathsf{A}_{ik} E[(Y_0)_k B_0^r] \\ &= \sum_{j=1}^{\infty} \left(\mathsf{A}^j \mathcal{B}(j) \right)_{i,r} \end{aligned}$$

We thus obtain

$$\begin{aligned} E[Y_0^i Y_0^j] &= E[B_0^i B_0^j] + E[(A_0(Y_0))_i B_0^j] \\ &\quad + E[(A_0(Y_0))_j B_0^i] \\ &\quad + \sum_{k=1}^{N} \sum_{m=1}^{N} E[Y_0^k Y_0^m] \mathsf{A}_{ik} \mathsf{A}_{jm} + Q_{ij} \\ &= E[B_0^i B_0^j] + \sum_{r=1}^{\infty} \left(\mathsf{A}^r \mathcal{B}(r) + \left[\mathsf{A}^r \mathcal{B}(r) \right]^T \right)_{i,j} \\ &\quad + \sum_{k=1}^{N} \sum_{m=1}^{N} E[Y_0^k Y_0^m] \mathsf{A}_{ik} \mathsf{A}_{jm} + Q_{ij} \end{aligned}$$

which gives in matrix notation (9).

We now rewrite (9) as

$$
\begin{aligned}
&cov(Y) + yy^T \\
&= cov(B) + \hat{B} + \sum_{r=1}^{\infty} \left(\mathsf{A}^r \hat{\mathcal{B}}(r) + \left[\mathsf{A}^r \hat{\mathcal{B}}(r) \right]^T \right) \\
&\quad + \mathsf{A}(I - \mathsf{A})^{-1} \hat{B} + \hat{B}(I - \mathsf{A}^T)^{-1} \mathsf{A}^T \\
&\quad + \mathsf{A} cov(Y) \mathsf{A} + \mathsf{A} yy^T \mathsf{A} + Q_{ij}
\end{aligned}
$$

We now note that

$$
\begin{aligned}
yy^T &= \hat{B} + \mathsf{A}(I - \mathsf{A})^{-1} \hat{B} + \hat{B}(I - \mathsf{A}^T)^{-1} \mathsf{A}^T \\
&\quad + \mathsf{A} yy^T \mathsf{A}
\end{aligned}
$$

which is obtained after some elementary algebra and after substituting $y = (I - \mathsf{A})^{-1} b$. We conclude that $cov(Y)$ is a solution of (8).

Next, we show uniqueness. Let Z_1 and Z_2 be two solutions of (8) and define $Z = Z_1 - Z_2$. Then Z satisfies $Z = \mathsf{A}^T Z \mathsf{A}$. Iterating that we obtain that

$$
Z = \lim_{n \to \infty} \mathsf{A}^n Z (\mathsf{A}^T)^n = 0
$$

where the last equality follows since $\|\mathsf{A}\| < 1$. This implies the uniqueness of the solution for (8). The uniqueness of the solution of (9) is obtained similarly. ∎

B. *Example of a correlated processes*

We assume in this Subsection that B_n are random vectors whose distribution depends on an underlying ergodic Markov chain θ_n taking values in a finite space Θ. We denote its transition probability by $\mathcal{P}$. Let π be the unique steady state probability of the Markov chain. Let $G_r^i(\theta) := P(B_n^i = r | \theta_n = \theta)$, Let $\hat{G}^i$ be a matrix of size $|\Theta| \times \mathrm{Z}^+$ whose lrth component $\hat{G}_r^i(l)$ equals $G_r^i(l)\pi(l)$. Let J be a row vector whose ith entry equals i, $i \in \mathrm{Z}^+$ Then for $j > 0$, Our goal is to compute the quantities that appear in (8) (in particular $\mathcal{B}(k)$).

Lemma 1: In the Markov correlated model described above, we have

$$
[\mathcal{B}(k)]_{ij} = E[B_0^i B_k^j] = J \hat{G}^i \mathcal{P}^{k-1} [G^j]^T J^T. \quad (10)
$$

If we denote by $\mathbf{1}$ the column vector with appropriate size whose entries are all ones, then we further have:

$$
[\hat{\mathcal{B}}(k)]_{ij} = (J - b_i \mathbf{1}^T) \hat{G}^i \mathcal{P}^{k-1} [G^j]^T (J - b_j \mathbf{1})^T. \quad (11)
$$

where $b_i = \sum_{l \in \Theta} E[B_0^i | \theta_0 = l] \pi(l)$.[1] Moreover.

$$
[cov(B)]_{ij} = \sum_{\theta \in \Theta} \pi(\theta) cor[B(\theta)]_{ij} \quad (12)
$$

[1] Note that $(J - b_i \mathbf{1}^T)\hat{G}^i$ is a row vector of dimension $|\Theta|$ whose lth entry equals $E[B_0^i - b_i | \theta_0 = l]\pi(l)$.

where $cor[B(\theta)]_{ij} = E[B_0^i B_0^j | \theta_0 = \theta]$.

Proof: We have

$$
P(B_0^i = r, B_k^j = s | \theta_n = \theta) = G_r^i(\theta) \sum_{\theta' \in \Theta} [\mathcal{P}^{k-1}]_{\theta\theta'} G_s^j(\theta')
$$

which implies

$$
\begin{aligned}
&P(B_0^i = r, B_k^j = s) \\
&= \sum_{\theta \in \Theta} \pi(\theta) G_r^i(\theta) \sum_{\theta' \in \Theta} [\mathcal{P}^{k-1}]_{\theta\theta'} G_s^j(\theta') \\
&= \sum_{\theta \in \Theta} \hat{G}_r^i(\theta) \sum_{\theta' \in \Theta} [\mathcal{P}^{k-1}]_{\theta\theta'} G_s^j(\theta') \\
&= [\hat{G}^i \mathcal{P}^{k-1} (G^j)^T]_{rs}.
\end{aligned}
$$

Hence $[\mathcal{B}(k)]_{ij} = \sum_s \sum_r rs [\hat{G}^i \mathcal{P}^{k-1} G^T]_{rs}$ which gives (10). The rest is direct. ∎

Next, consider the special case that the B_n^i's have only values 0 or 1. Let p and $\hat{p}$ denote the matrices whose (i, θ) entry equal, respectively, to $p_\theta(i) := P(B_n^i = 1 | \theta_n = \theta)$ and $\hat{p}_\theta(i) := P(B_n^i = 1 | \theta_n = \theta) - P(B_n^i = 1)$. Let g denote the matrix whose (i, θ) entry equals $g_\theta(i) = \pi(\theta)\hat{p}_\theta(i)$. Then (11) simplifies to

$$
\hat{\mathcal{B}}(k) = g \mathcal{P}^{k-1} \hat{p}^T \quad (13)
$$

C. *The one dimension case*

We next consider scalar stochastic recursive equations, i.e. $N = 1$. Y_n in (1) is then a scalar instead of a vector and (2) simplifies to

$$
A_n(Y_n) = \sum_{k=1}^{Y_n} \xi^{(k)}(n). \quad (14)
$$

$\xi^{(k)}$ and A are scalar too with $E[\xi^{(k)}(n)] = \mathsf{A}$. Theorem 2 simplifies to:

Theorem 3: (i) The first moment of Y_n^* is given by

$$
E[Y_0^*] = \frac{b}{1 - \mathsf{A}}, \quad (15)
$$

(ii) The variance of Y_n^* is given by

$$
\begin{aligned}
&var[Y^*] = E[(Y^*)^2] - (E[Y^*])^2 \\
&= \frac{var[B] + \sum_{r=1}^{\infty} \left(\mathsf{A}^r \hat{\mathcal{B}}(r) + \left[\mathsf{A}^r \hat{\mathcal{B}}(r) \right]^T \right) + \mathsf{A} b}{1 - \mathsf{A}^2}
\end{aligned}
$$

Next, we shall further restrict to the Markovian setting of Section III-B. We shall provide an explicit expression for $\sum_{r=1}^{\infty} \mathsf{A}^r \mathcal{B}(r)$.

Lemma 2: In the one dimensional state with the Markov model for correlation, we have

$$\sum_{r=1}^{\infty} \mathsf{A}^r \mathcal{B}(r) = \mathsf{A}(J - b\mathbf{1}^T)\hat{G}[I - \mathsf{A}\mathcal{P}]^{-1} G^T (J-b)^T. \quad (16)$$

Proof. We get using (11)

$$\mathsf{A}^r \mathcal{B}(r) = \mathsf{A}(J - b\mathbf{1}^T)\hat{G}[\mathsf{A}\mathcal{P}]^{r-1} G^T (J-b)^T,$$

$\sum_{r=1}^{\infty} [\mathsf{A}\mathcal{P}]^{r-1}$ is well defined since $|\mathsf{A}| < 1$ and since $\mathcal{P}$ is a stochastic matrix. Define J^x to be a row vector whose ith entry equals $\min(i,x)$, $i \in \mathrm{Z}^+$. Then for any $x > 0$, we have by the bounded convergence theorem:

$$\sum_{r=1}^{\infty} \mathsf{A}(J^x - b\mathbf{1}^T)\hat{G}[\mathsf{A}\mathcal{P}]^{r-1} G^T (J^x - b)^T$$

$$= \mathsf{A}(J^x - b\mathbf{1}^T)\hat{G}[I - \mathsf{A}\mathcal{P}]^{-1} G^T (J^x - b)^T.$$

(16) is then obtained by the monotone convergence theorem. ■

We thus obtain the following:

Corollary 1: Consider the scalar case, and consider the Markov model for the correlation process of Section III-B. Then

$$var[Y^*] = \frac{var[B] + 2\mathsf{A}(J - b\mathbf{1}^T)\hat{G}[I - \mathsf{A}\mathcal{P}]^{-1} G^T (J-b)^T + \mathsf{A}b}{1 - \mathsf{A}^2}$$

Moreover, in the special case that the B_n^i's have only values 0 or 1, then we get

$$var[Y^*] = \frac{var[B] + 2\mathsf{A}g(I - \mathsf{A}\mathcal{P})^{-1}\hat{p}^T + \mathsf{A}b}{1 - \mathsf{A}^2} \quad (17)$$

IV. The G/PH/∞ queue

We now consider a discrete time G/PH/∞ queue. We shall apply in this section the general theory of previous sections in order to compute the steady state moments of some performance measures. We shall then strengthen in the following section the stability results (corresponding to Theorem 1) while relaxing further the statistical assumptions.

A. The model

Service times: Service times are considered to be i.i.d. and independent of the arrival process. We represent the service time as the discrete time analogous of a phase type distribution: there are N possible service phases. The initial phase k is chosen at random according to some probability $p(k)$. If at the beginning of slot n a customer is in a service phase i then it will move at the end of the slot to a service phase j with probability P_{ij}. With probability $1 - \sum_{j=1}^{N} P_{ij}$ it ends service and leaves the system at the end of the time slot. Let $\xi^{(k)}(n)$, $k = 1,2,3,...$, $n = 1,2,3,...$ be i.i.d. random matrices of size $N \times N$. Each of its element can take values of 0 or 1, and the elements are all independent. The ijth element of $\xi^{(k)}(n)$ has the interpretation of the indicator that equals one if at time n, the kth customer among those present at service phase i moved to phase j. Obviously, $E[\xi_{ij}^{(k)}(n)] = P_{ij}$. P is a sub-stochastic matrix (it has nonnegative elements and it's largest eigenvalue is strictly smaller than 1), which means that services ends in finite time w.p.1. and that $(I - P)$ is invertible.

Arrivals: Let $B_n = (B_n^1, ..., B_n^N)^T$ be a column vector for each integer n, where B_n^i is the number of arrivals at the nth time slot that start their service at phase i. B_n is assumed to be a stationary ergodic sequence and that they have finite expectation.

The state and the recursive equation: Let Y_n^i denote the number of customers in phase i at time n. Then Y_n satisfies the recursion (1) where A_n is given in (2). In particular, $\mathsf{A} = P$ and indeed we have $\|\mathsf{A}\| < 1$ so that Assumption A1 hold.

We can thus apply the results of the previous sections to get the first two moments as well as the general distribution at stationary regime.

B. Main results

Corollary 2: (i) Theorems 1 and 2 hold for the G/PH/∞ queue.
(ii) The first and second moments of the number of customers at the system in stationary regime are given respectively by $\mathbf{1}^T(I - \mathsf{A})^{-1}b$ and $\mathbf{1}^T cov(Y)\mathbf{1}$, respectively, where $\mathbf{1}$ is a column vector with all entries 1's.

Remark 2: We present a simple interpretation of the first moment of the number of customers at the system. Denote by λ the expected number of arrivals per slot. Clearly $\lambda = |b|$ where $|b|$ is the sum of entries of the vector b. Define ζ to be the expected service time of an arbitrary customer and let $\rho = \lambda\zeta$. We shall first compute ζ. The ijth element of the matrix $(I - \mathsf{A})^{-1}$ has the interpretation of the total expected number of slots that a customer that had arrived at service phase j spent at state i. Thus the jth entry of the vector $\mathbf{1}^T(I - \mathsf{A})^{-1}$ has the interpretation of the total expected number of slots that a customer that had arrived at service phase j spent in the system. and let the vector β be the vector whose $i$the entry is $b/|b|$. Then

$$\zeta = \mathbf{1}^T(I - \mathsf{A})^{-1}\beta$$

and

$$\rho = (\mathbf{1}^T(I-\mathsf{A})^{-1}\beta)|b| = \mathbf{1}^T(I-\mathsf{A})^{-1}b,$$

which is our expression for the first moment of the number of customers at the system. This relation is known to hold in fact for general $G/G/\infty$ queues, see e.g. [6, p. 134].

C. Departure process

One can use the same methodology to describe the departure process. To do that, we can augment the system with a new "phase" which we call "d" (for departure), and update the phase transitions as follows:

$$\begin{aligned}\overline{P}_{ij} &= P_{ij}, \quad i,j \in \{1,...,N\},\\ \overline{P}_{id} &= 1-\sum_{j=1}^{N} P_{ij}, i\in\{1,...,N\}\\ \overline{P}_{di} &= 0, \qquad i\in\{1,...,N,d\}\end{aligned}$$

Quantities corresponding to the new system are denoted by adding a bar. We set $\overline{B}_n^i = B_n^i$ for $i = 1,...,N$ and $\overline{B}_n^d = 0$ for all integers n. Since P is assumed to be substochastic, so is $\overline{P}$. Note that in our new system, only customers in phases 1,...,N correspond to those really present in the original system, whereas customers at phase d are already out of the system.

D. The case of geometric service times

We now study the special case of geometrically distributed service times. In that case the stochastic recursive equation becomes one dimensional. Y_n is a scalar and denotes the number of customers in the system. $\xi_n^{(k)}$ has the interpretation of the indicator that the kth customer present at the beginning of time-slot n will still be there at the end of the time-slot. Thus the probability that a customer in the system finishes its service within a time slot is precisely $p = 1-\mathsf{A}$. We can now apply directly the results of Theorem 3.

E. Numerical results

We consider the following simple scenario. Service times are geometrically distributed, the arrival process depends on a Markov chain as in Subsection III-B, and moreover, there can be either one or no arrival at a time slot.

We consider a Markov chain with two states $\{\gamma,\delta\}$ with transition probabilities given by

$$\mathcal{P} = \begin{pmatrix} 1-\epsilon p & \epsilon p\\ \epsilon q & 1-\epsilon q\end{pmatrix}$$

$\epsilon > 0$ is a parameter that will be varied later in order to vary the correlations. The steady state probabilities of this Markov chain are

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right).$$

Hence

$$b = E[B] = E[B^2] = \frac{qp_\gamma + pp_\delta}{p+q}, \tag{18}$$

$$var[B] = \frac{(qp_\gamma+pp_\delta)(q(1-p_\gamma)+p(1-p_\delta))}{(p+q)^2} \tag{19}$$

Note that $\pi, b, E[B^2]$ and $var[B]$ do not depend on ϵ.

Applying the first part of Theorem 3 we get the following expression for the expected number of customers in the system in stationary regime:

$$E[Y_0^*] = \frac{1}{1-\mathsf{A}} \times \frac{qp_\gamma+pp_\delta}{p+q}.$$

Next, we wish to compute the variance of Y. We have

$$\hat{p} = \Big[\frac{p(p_\gamma-p_\delta)}{p+q}, \frac{q(p_\delta-p_\gamma)}{p+q}\Big],$$

$$g = \frac{pq(p_\gamma-p_\delta)}{(p+q)^2}\Big[1,-1\Big],$$

$$\begin{aligned}(I-\mathsf{A}\mathcal{P})^{-1} &= \frac{1}{(1-\mathsf{A})(1-\mathsf{A}+\epsilon(p+q)\mathsf{A})}\\ &\quad\times\begin{pmatrix}1-\mathsf{A}+\epsilon\mathsf{A}q & \epsilon\mathsf{A}p\\ \epsilon\mathsf{A}q & 1-\mathsf{A}+\epsilon\mathsf{A}p\end{pmatrix}\end{aligned}$$

We thus obtain

$$\begin{aligned}&g(I-\mathsf{A}\mathcal{P})^{-1}\hat{p}^T \\ &= \frac{pq(p_\gamma-p_\delta)^2}{(1-\mathsf{A}+\epsilon(p+q)\mathsf{A})(p+q)^2}\end{aligned} \tag{20}$$

Remark 3: If we consider $var[Y^*]$ in (17), we see that the dependence on ϵ comes only through the term in (20). Moreover, we see that for any value of A, this term, and hence $var[Y^*]$, decrease with ϵ. Large ϵ means that the Markov chain alternates rapidly between its two states, which results in a lower overall effect of correlation. (20) can precisely be used to determine this overall effect as it can be viewed of the total weighted sum of correlations $\hat{\mathcal{B}}(k)$, i.e.

$$g(I-\mathsf{A}\mathcal{P})^{-1}\hat{p}^T = \sum_{k=0}^{\infty} g(\mathsf{A}\mathcal{P})^k\hat{p}^T = \sum_{k=1}^{\infty}\mathsf{A}^{k-1}\hat{\mathcal{B}}(k)$$

where we used (13).

As an example, consider the following parameters: $p = q = 1$, $p_\gamma = 1, p_\delta = 0.5$. Substituting these parameters in (18), (19) and (20) and plugging these expressions into (17), we obtain the following expressions

$$var[Y^*] = \frac{1}{(1-\mathsf{A}^2)}\left(\frac{3}{16} + \frac{2\mathsf{A}}{1-\mathsf{A}+2\epsilon\mathsf{A}} + \frac{3}{4}\mathsf{A}\right).$$

In Fig. 1 we plot the variance of the steady state number of customers, $var[Y^*]$, while varying ϵ and A.

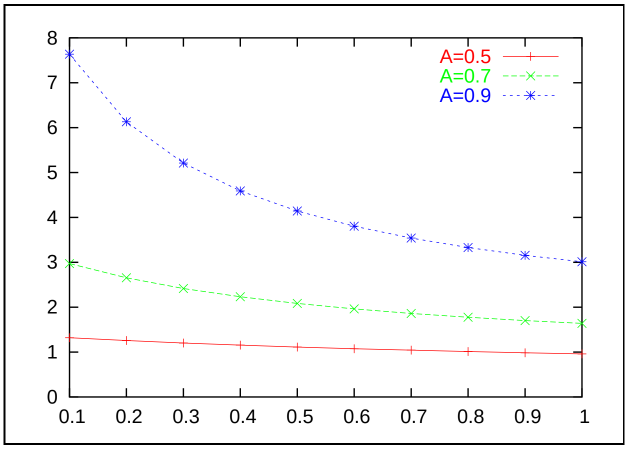

Fig. 1. var[$Y*$] as a function of ϵ and of A

Recall that for a fixed A, the expectation of Y^* does not depend on ϵ. The variance of Y^* on the other hand is seen to be quite sensitive to the correlation between the B_n's as determined by the parameter ϵ. This sensitivity is seen to increase as A increases and sensitivity is largest when A approaches 1. As already mentioned in Remark 3 we see that $\epsilon = 1$ gives the smallest value of $E[Y^*]$ and that $E[Y^*]$ increases as ϵ decrease. For $\mathsf{A} = 0.5$ we get a difference of around 30% between the lowest and the largest value of ϵ, where as form $\mathsf{A} = 0.9$ we obtain a difference of 250%.

F. Extension to a network

Consider now M stations, each with infinite number of servers. The service time at station i has a set $\mathcal{N}_i$ of N_i phases. Let $N = N_1 + ... + N_M$. For any $j = 1, ..., N$ let $s(j)$ denote the station to which j corresponds, i.e. if $j \in \mathcal{N}_i$ then $s(j) = i$.

If at time n a customer was at phase j in station $s(j)$ then it either moves to another phase at the same station or moves to another phase in another station; the next phase k (either at the same station or at another one) is chosen with probability P_{jk}; with probability $1 - \sum_{k=1}^{N} P_{jk}$ the customer leaves the system. Again we assume that the choice of next phase are independent.

Let $B_n = (B_n^1, ..., B_n^N)^T$ be a column vector for each integer n, where B_n^i is the number of arrivals at the nth time slot that start their service at phase i in station $s(i)$. B_n is assumed to be a stationary ergodic sequence.

With this description we see that we can identify the whole network as a single server station problem with infinite number of servers and with N phases. Thus we can apply all previous results.

V. Residual activity time in the $G/G/\infty$ queue

Define the residual activity time at a given instant as the total time till the system empties from that instant onwards if new arrivals do not occur.

We shall analyze in this section the residual activity of a G/G/∞ queue under weaker statistical assumptions than those used so far. We shall obtain the existence of a stationary regime as well as convergence to it in the coupling-convergence sense (see e.g. Borovkov [7], [8]).

a) The model: The nth arrival event occurs at time T_n: a batch of B_n customers arrive. Denote $\tau_n = T_{n+1} - T_n$; they replace the fixed slots we had before. Let σ_n be the largest service time required among the B_n customers that arrive at time T_n. We shall assume that the joint sequence (τ_n, σ_n) is stationary ergodic and that $E[\tau_0]$ and $E[\sigma_0]$ are finite and strictly positive. σ_n in particular, need not have a "phase type distribution" as before. Let V_n be the residual activity time just before T_n. Then V_n can be written recursively as:

$$V_{n+1} = \Big(\max(V_n, \sigma_n) - \tau_n\Big)^+$$

where $(x)^+ := \max(x, 0)$.

Iterating this relation gives:

$$\begin{aligned} &V_{n+2} \\ &= \left(\max\left\{\Big[\max(V_n,\sigma_n) - \tau_n\Big]^+, \sigma_{n+1}\right\} - \tau_{n+1}\right)^+ \\ &= \max\Big(\max(V_n,\sigma_n) - \tau_n - \tau_{n+1}, \sigma_{n+1} - \tau_{n+1}, 0\Big). \end{aligned}$$

Further iterating directly yields:

$$V_{n+k} = \max(Z_n, Z_{n+1}, ..., Z_{n+k-1}, 0)$$

where

$$Z_n = \max(V_n, \sigma_n) - \sum_{i=0}^{k-1} \tau_{n+i},$$

$$Z_{n+j} = \sigma_{n+j} - \sum_{i=j}^{k-1} \tau_{n+i}, \quad j = 1, ..., k-1.$$

b) Stationary solution: We use the Loynes' type scheme [16] to obtain the stationary regime and the convergence to it.

Theorem 4: V_n converges a.s. to a unique stationary regime that is given by

$$V_n^* := \left(\max_{j<n} \left[\sigma_j - \sum_{i=j}^{n-1} \tau_i \right] \right)^+ . \qquad (21)$$

from any initial V_0. Moreover V_n^* is $P - a.s.$ finite.

Proof: Define on the same probability space as the process V_n the shifted processes $V_n^{[m]}$, where m are integers:

$$V_{-m}^{[m]} = 0, \quad V_{n+1}^{[m]} = \left(\max(V_n^{[m]}, \sigma_n) - \tau_n \right)^+, \; n \geq -m.$$

Then as before, we can write for $n > -m$:

$$V_n^{[m]} = \left(\max_{-m \leq j < n} \left[\sigma_j - \sum_{i=j}^{n-1} \tau_i \right] \right)^+$$

which monotonely increases to the sequence V_n^* given in (21). Clearly V_n^* is a stationary ergodic process. We shall show that it is $P - a.s.$ finite. Indeed, since (τ_n, σ_n) is stationary ergodic, the Cesaro sums converge to the expectation $P - a.s.$ and hence there is some R.V. J_0 which is finite $P - a.s.$ such that for all $j > J_0$,

$$\sigma_{-j} < jE[\tau_0]/3 \text{ and } \sum_{i=-j}^{-1} \tau_i > j2E[\tau_0]/3$$

Hence the term in brackets in (21) is negative for all $-j > J_0$ so that V_0^* is finite $P - a.s.$ Due to stationarity this is true for V_n^* for all n.

c) Coupling: We show that for any initial value V_0 there is a time N_0 which is finite $P - a.s.$ such V_n coincides with V_n^* for all $n > N_0$. Indeed, fix V_0 and define

$$N_0 := \inf \left\{ l : \max(V_0, V_0^*) < \sum_{i=0}^{l-1} \tau_i \right\}.$$

N_0 is clearly finite $P - a.s.$ due to the ergodicity of τ_i. Moreover, it is clear from the explicit expressions we have for V_0 and for V_0^* that they coincide for $n > N_0$. Uniqueness of the stationary regime follows from the fact that coupling has been established for arbitrary initial state. ∎

Remark 4: Our construction establishes in fact that we have strong coupling convergence in the sense of [7], [8].

Remark 5: A stability result is already given in [6, p. 133] for a general G/G/∞ queue. Namely, it is shown that V_0 is finite almost surely but the form of the stationary regime and the convergence results are not given.

VI. Concluding comments

In this paper we have studied and used stochastic recursive equations to investigate the discrete infinite server queue with batch arrivals where the size of the batches follow a general stationary ergodic process. We obtained explicit expressions for the first and second moments of the state variables appearing in the stochastic recursive equations and applied them to solve the infinite server queue problem. We proposed then more specific Markov models for correlation that further simplify the expressions for the first two moments. We extended the results of the infinite server queue to a whole network of such queues.

Other stochastic recursive equations have been used to study the stability of the queue under even more general probabilistic assumptions and convergence to a unique stationary regime has been established.

The simple explicit expressions obtained makes our results appealing to various applications of the infinite server queue. For example, they can be used to represent the first and second moments of the number of connected mobiles at an arbitrary location in the one dimensional ad-hoc network of [17], using the equivalence between the ad-hoc network and an infinite server queue given in [17].

Acknowledgement

This work was supported by the EURO NGI network of excellence.

References

[1] S. R. Adke and V. G. Gadag, "A new class of branching processes", *Branching Processes: Proceedings of the First World Congress*, C.C.Heyde (Editor), 1-13, Springer Lecture Notes 99, 1995.

[2] E. Altman, "Stochastic recursive equations with applications to queues with dependent vacations", *Annals of Operations Research*, 112(1): 43-61; Apr, 2002.

[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, T. Jimenez , D. Kofman, "DPS queues with stationary ergodic service times and the performance of TCP in overload", Proceedings of IEEE Infocom, Hong-Kong, March 2004.

[5] E. Altman, P. Konstantopoulos, Z. Liu, "Stability, Monotonicity and Invariant Quantities in General Polling Systems", *Queueing Systems* **11** , pp. 35-57, 1992

[6] F. Baccelli and P. Brémaud, *Elements of Queueing Theory*, Springer, second edition, 2003.

[7] A. A. Borovkov, *Asymptotic Methods in Queueing Theory*, John Wiley & Sons, 1984 (translated from Russian).

[8] A. A. Borovkov and S. G. Foss, "Stochastically recursive sequences and their generalizations", *Siberian Advances in Mathematics*, **2**, No. 1, pp. 16-81, 1992.

[9] A. Brandt, "The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coeffi cients", *Advances in Applied Probability*, Vol. 18, 1986.

[10] A. Brandt, P. Franken and B. Lisek, *Stationary Stochastic Models*, Akademie-Verlag, Berlin, 1992.
[11] P. Glasserman and D. D. Yao, “Stochastic vector difference equations with stationary coefficients”, *J. Appl. Prob.*, Vol 32, pp 851-866, 1995.
[12] R. Groenevelt and E. Altman, “An alternating priority server with correlated and cross correlated vacations”, manuscript.
[13] A. Jean-Marie and P. Robert, *On the transient behavior of the processor sharing queue*, *QUESTA* **17**, No. 1-2, pp. 129-136, 1994.
[14] H. Kameda and Y. Zhang, “Uniqueness of the solution for optimal static routing in open BCMP queueing networks”, *Math. Comput. Modeling* **22**, No. 10-12, 119-130, 1995.
[15] M. Kruntz and A. Makowski, ”Modeling video traffic using M/G/infinity input processes: A compromise between Markovian and LRD models,” *IEEE Journal on Selected Areas in Communications (JSAC)*, pp. 733-748, Vol. 16, No. 5, June 1998.
[16] R. Loynes, “The stability of a queue with non-independent interarrival and service times”, *Proc. Cambr. Phil. Soc.* **58**, No. 3, pp. 497-520, 1962.
[17] D. Miorandi and E. Altman, ”Connectivity in Ad–Hoc Networks: a Queueing Theoretical Approach”, *Proceedings of WiOpt workshop*, Cambridge, UK, March 2004.