Qatar Univ. Sci. J. (2004), 24 : 45 - 56

Performance Analysis of a Fileserver in an ATM Network

Hamed Nassar* and Ali Meligy †

* Faculty of Petroleum and Mining, Suez Canal University, Suez, Egypt.

† Faculty of Science, Computer Section, Shebin El-Kom, Egypt.

تحليل الأداء لخادم ملفات في شبكة صرف آلي

حامد نصار* و علي المليجي†

* كلية البترول والتعدين - جامعة قناة السويس - السويس - مصر

† كلية العلوم - قسم الحاسب - شبين الكوم - مصر

في هذا البحث نقوم بتحليل أداء خادم ملفات في شبكة صرف آلي بافتراض أن طلبات الملفات تصل في صورة رسائل قصيرة مكونة من حزمة واحدة لكل رسالة، وأن الملفات مكونة من عدة حزم موزعة هندسياً، وأن الطلبات تنتظم في طابور حيث تعامل تبعاً لترتيب الوصول، ويتم إخلاء طلب من الطابور عند الانتهاء من نقل الملف المطلوب إلى المحطة الطالبة.

نقوم بتحليل ثلاثة مقاييس لخادم الملفات هي الإشغال، والعمل غير المكتمل، وزمن الانتظار، ونحصل لكل من هذه المقاييس على دالة توليد الاحتمال، والتوزيع التام، والتوقع.

Keywords: *ATM, Fileserver, Unfinished Work, Busy period, Waiting Time.*

ABSTRACT

ATM continues to gain acceptance as a unifying communications technology for traffic of both digital and analog origins. The main characteristic of an ATM network is its discrete time (slotted) operation and fixed-length packets.

In this paper we analyze the performance of a fileserver operating in an ATM network to serve files to a group of workstations upon requests of the latters. We assume that the file requests arrive into the fileserver as short messages of one packet each. The files themselves are assumed to have a geometrically distributed number of packets. The requests arriving into the fileserver are buffered in a queue and are treated on a first come first serve (FCFS) basis. A request is dismissed from the queue only when its associated file has been fully transmitted to the requesting station.

We analyze three metrics for the fileserver: occupancy, unfinished work, and waiting time. For each of these metrics, we obtain the probability generating function (PGF), the entire distribution, and the expectation.

1. Introduction

Asynchronous Transfer Mode (ATM) [1] technology has recently emerged as the dominant standard in communications networks. Its main characteristic is its discrete time (slotted) operation and fixed-length packets of 53 bytes each. This technology has made possible implementing Integrated Services Data Networks (ISDNs), where traffic originating from data, audio and video sources flows harmoniously [11]. As a consequence, there is currently an upsurge of research on the performance analysis of ATM computer netrowks (see e.g. [4], [7], and [5].)

In this paper we analyze the performanc of a fileserver operating in an ATM network containing also $N \leq \infty$ workstations. The fileserver and the workstations communicate with each other using fixed-length packets. Time is divided into units, called slots, each has the duration of the transmission time of one packet.

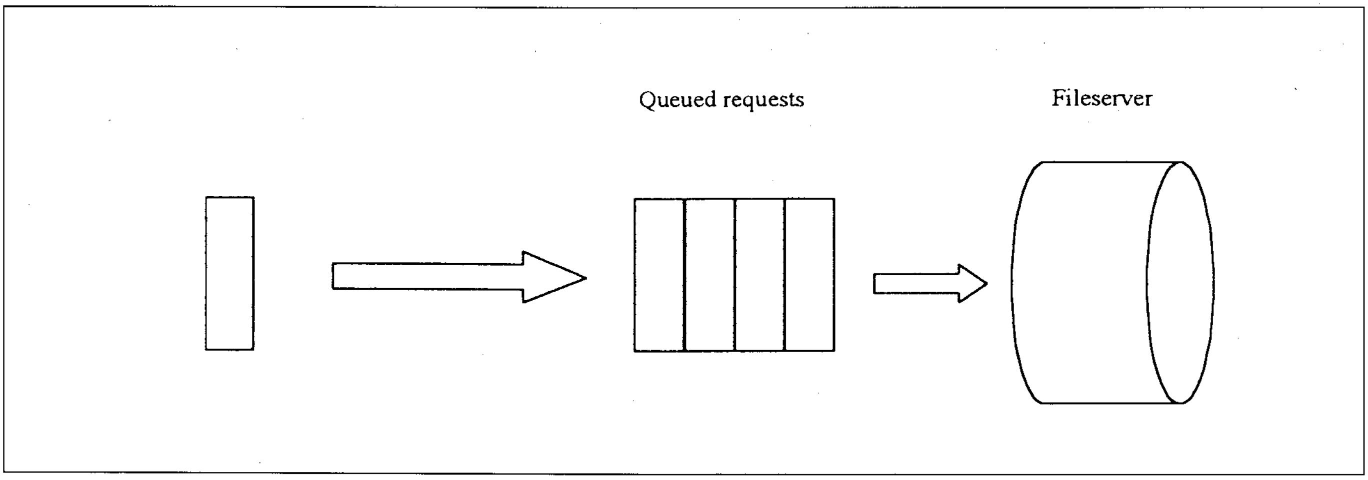

Figure 1: Fileserver model.

When a worksttion requires a file from the fileserver it sends a request, in the form of a packet, to the fileserver. The fileserver buffers the arrivind requests in a queue and handles them on a first come first serve (FCFS) basis. When a request is served, i.e. when the file it is asking for is transmitted to the requesting workstation, the request is dismissed from the queue and the next request in line begins service. The requested files vary in length. A file could be very short, e.e. one packet, or huge, e.g. mega packet, or anything in between. That is, while a request arrived in always 1 slot, its "service time" can be any number X, $1 \leq X \leq \infty$, of slots.

We analyze three work metrics for the fileserver: occupancy, unfinished work, and waiting time. It should be noted that the occupancy was analyzed in a previous work [8]. However, the analysis technique in this paper is different and, more importantly, is more rigorous. Also the occupancy and waiting time were analyzed in [13], but for a constant service time of 1 slot (i.e. $X = 1$).

The analysis tool is queueing theory [6]. In particular, the fileserver is modelled as a discrete time queuing system with Markovina characteristics [8], [9]. For each of the three metrics analyzed, we obtain the probability generating function (PGF), the entire distribution, and the expectation.

This paper is organized as follows. In the next section the assumptions of the fileserver will be introduced. In section 3 the system occupancy is analyzed, in section 4 the unfinished work is analyzed, and finally in section 5 the waiting time is analyzed. In section 6 some concluding remarks are made.

2. Model Assumptions

In this section will introduce the assumptions of the queueing model that will represent the fileserver. First, time is slotted, withe ach slot duration equal to the transmission time of one packet. Requests arrive from the workstations to the fileserver may be more than one packet. A request is always are stored in a buffer in the fileserver. The buffer has infinite capacity. A request is dismissed from the head of the queue only when the file it is asking for has been fully transmitted by the fileserver.

Requests are generated by many sources. However, the fileserver can receive at most one request (from all the workstations) per slot. In particular, we will assume that in any slot, the probability that a request will arrive in r and the probability that it will not arrive is $\bar{r} = 1 - r$. In other words, the request interarrival time is gemoetrically distributed with expectation $1/r$. Also this means that the arrival process is Bernoulli with arrival rate r packets per slot.

We will look on the fileserver system as made up of two distinct parts: the queue (or buffer) and the server. The queue operates in a first-come-first serve (FCFS) manner. A request arriving at the system in a given slot will be stored in the queue, behind the requests that arrived ahead of it, if any, until the server is free, when it begins service. For reasons of synchronization, a request cannot begin service except at the beginning of a slot. Therefore, a request arriving in an empty system will have to waitn in the queue for the remainder of the arrival slot until it begins service at the beginning of the next slot.

The fileserver can serve only one file at a time. That is, our queuing system has a single server. The served files depart the fileserver at the rate of s packets per slot, provided the buffer is not empty. That is, a packet being served in a certain slot will depart the server by the end of that slot with probability s and will not depart with probability $\bar{s} = 1 - s$. This implies that the length of each file, expressed in packets, is geometrically distributed with parameter s, and that the expected length of each file is $1/s$.

Since the packet departure rate is content dependent, it will be denoted by $s_{|n}$, where n is the current number of requests in the file server. Note that $s_{|n} = 0$ for $n = 0$ since the fileserver cannot serve a request when there are not any. Note also that $s_{|n} = 0$ for $n < 0$ since the number n of reuests cannot be negative. Thus, $s_{|n} = s$ only for $n > 0$. That is,

$$s_{|n} = \begin{cases} s & n > 0 \\ 0 & n \leq 0 \end{cases} \tag{1}$$

The above assumptions lead us to a discrete-time counterpart of the continuous-time $M/M/1$ queuing system, which indicates exponentially distributed interarrival time, exponentially distributed service time, and one server [2]. That is because the geometric distribution is the discrete counterpart of the continuous exponential distribution.

In the sequel we will analyze the fileserver under these assumptions. Our analysis will rely principally on random variables (RVs), which are all nonnegative and integral valued.

3. System Occupancy

In this section we will determine the fileserver occupancy, i.e. the numper P of requests in the file server at an arbitary slot in steady state. Clearly, P is a RV. three characterization of P will be obtained: the probability generating function (PGF), the probability distribution and the expectation.

Let P^k denote the fileserver offupancy at the end of an arbitrary slot k, $k = 0, 1, \ldots$. Also, let A^k be the number of requests that arrive into the system during slot k. From the assumptions, we can see that A^k is a Bernoulli RV with the distribution.

$$a^k_n = \begin{cases} r & n=1 \\ \bar{r} & n=0 \end{cases} \tag{2}$$

Let D^k be the number of requests that depart the system during slot k. From the assumptions, we can see that D^{k+1} is a Bernoulli RV dependent on the occupancy P^k, at the end of the preceding slot, with the following conditional distribution

$$d^{k+1}_{i|P^k=j} = \Pr[D^{k+1} = i|P^k = j] = \begin{cases} s & \text{if } i=1, j>0 \\ \bar{s} & \text{if } i=0, j>0 \\ 1 & \text{if } i=0, j=0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

It is now clear that the occupancy at the end of slot $k+1$ is

$$P^{k+1} = P^k - D^{k+1} + A^{k+1}, \tag{4}$$

Note that the RV on the right hand side is never negative due to the way D^{k+1} and P^k are correlated. Let us define the PGF of P^k by

$$P^k(z) = E[z^{P^k}].$$

Using this definition on both sides of (4) we get

$$P^{k+1}(z) = E[z^{P^k - D^{k+1} + A^{k+1}}]$$

Due to the fact that A^{k+1} is independent of both P^k and D^{k+1}, it follows that

$$\begin{aligned} P^{k+1}(z) &= E[z^{P^k - D^{k+1}}]E[z^{A^{k+1}}]. \\ &= A^{k+1}(z)\, E[z^{P^k - D^{k+1}}]. \end{aligned} \tag{5}$$

The factor $E[z^{P^k - D^{k+1}}]$ can be evaluated as follows.

$$\begin{aligned} E[z^{P^k-D^{k+1}}] &= E[z^{P^k-D^{k+1}}|P^k=0, D^{k+1}=0]\Pr[P^k=0, D^{k+1}=0] \\ &+ E[z^{P^k-D^{k+1}}|P^k=0, D^{k+1}=1]\Pr[P^k=0, D^{k+1}=1] \\ &+ E[z^{P^k-D^{k+1}}|P^k>0, D^{k+1}=0]\Pr[P^k>0, D^{k+1}=0] \\ &+ E[z^{P^k-D^{k+1}}|P^k>0, D^{k+1}=1]\Pr[P^k>0, D^{k+1}=1] \\ &= E[z^0]\Pr[D^{k+1}=0|P^k=0]\Pr[P^k=0] \\ &+ E[z^{-1}]\Pr[D^{k+1}=1|P^k=0]\Pr[P^k=0] \\ &+ E[z^{P^k}]\Pr[D^{k+1}=0|P^k>0]\Pr[P^k>0] \\ &+ E[z^{P^k-1}]\Pr[D^{k+1}=1|P^k>0]\Pr[P^k>0] \end{aligned}$$

Substituting from (3) yields

$$\begin{aligned} E[z^{P^k-D^{k+1}}] &= 1.1.p_0^k + 0 + E\left[z^{P^k}\right].\bar{s}.\Pr[P^k>0] \\ &+ E\left[z^{P^{k-1}}\right].s.\Pr[P^k>0] \\ &= p_0^k + \bar{s}(P^k(z) - p_0^k) + \sum_{n=0}^{\infty} z^{n-1} p_n^k \\ &= p_0^k + \bar{s}(P^k(z) - p_0^k) + \frac{s}{z}\sum_{n=0}^{\infty} p_n^k z^n p_0^k \end{aligned}$$

$$= p_0^k + \bar{s}\, P^k(z) - \bar{s}\, p_0^k + \frac{s}{z}\,(P^k(z) - p_0^k)$$

Substituting for $E\,[z^{p^k - D^{k+1}}]$ in (5) we get

$$zP^{k+1}(z) = A^{k+1}(z)\, P^k(z)\,(z\bar{s} + s) + A^{k+1}(z)\, sp_0\,(z-1) \tag{6}$$

If $r < s$, the system reaches steady stake as $k \to \infty$, in which case the PGFs $P^k(z)$ and $A^{k+1}(z)$ converge to the common PGFs $P(z)$ and $A(z)$, respectively. Hence in steady states (6) becomes

$$zP(z) = A(z)\, P(z)\,(z\bar{s} + s) + A(z)\, sp_0\,(z-1).$$

Solving for $P(z)$ we get

$$P(z) = \frac{A(z)\, s\,(z-1)\, p_0}{z - A(z)\,(s + \bar{s}z)}.$$

Substituting for $A(z)$ yields

$$P(z) = \frac{(\bar{r}+rz)\, s\,(z-1)\, p_0}{z - (\bar{r}+rz)\,(s+\bar{s}z)}$$

$$= \frac{(\bar{r} + rz)\, s\,(z-1)\, p_0}{z - s\bar{r} - srz - \bar{r}\bar{s}z - r\bar{s}z^2}$$

$$= \frac{(\bar{r} + rz)\, sp_0}{s\bar{r} - r\bar{s}z} \tag{7}$$

The value of p_0 can be evaluated using the normalization condition

$$P(1) = 1. \tag{8}$$

Substituting by $z = 1$ in (7) and noting that $sr - \bar{r}s = \bar{s} - r$, we find

$$P(1) = \frac{(\bar{r} + r\,.\,1)\, sp_0}{s\bar{r} - r\bar{s}\,.\,1}$$

$$= \frac{sp_0}{s - r} = 1$$

Solving for p_0 so we get

$$p_0 = 1 - \frac{r}{s} = 1 - \rho$$

This is the probability that the fileserver is empty at an arbitrary slot in steady state. Note that for a valid (i.e. nonnegative) value of p_0, we should have $r \le s$. Note, however, that if $r = s$ then $p_0 = 0$, meaning that the fileserver will never be empty. Note lastly that since p_0 is the probability that the fileserver is empty, then its complement ρ is the probability that the fileserver is busy. The higher the ρ the busier the fileserver. Indeed, ρ is called fileserver utilization, offered load, or traffic intensity.

Substituting for p_0 in (7), yields

$$P(z) = \frac{(\bar{r} + rz)\,(s - r)}{s\bar{r} - r\bar{s}z} \tag{9}$$

It is time now to fine the distribution p_n of the occupancy. To do that, $P(z)$ will be rewritten as follows.

$$P(z) = \frac{\rho\,(s-r)}{\bar{r}}\, \frac{z}{1 - \frac{\rho\bar{s}}{\bar{r}}z} + (1-\rho)\, \frac{1}{1 - \frac{\rho\bar{s}}{\bar{r}}z} \tag{10}$$

Each term on the RHS seems to contain a sum of an infinite geometric series. The second sum is straight forward, whereas the first has a problem. To handle this problem, we know that

$$\frac{1}{1-\alpha z} = \sum_{n=0}^{\infty} \alpha^n z^n, \qquad |\alpha z| < 1$$

Multiplying both sides by z gives

$$\frac{z}{1-\alpha z} = \sum_{n=1}^{\infty} \alpha^{n-1} z^n$$

Using this result in (10) and letting $\alpha = \frac{\rho \bar{s}}{\bar{r}}$ we get

$$P(z) = \frac{\rho(s-r)}{\bar{r}} \frac{z}{1-\alpha z} + (1-\rho)\frac{1}{1-\alpha z}$$

$$= r(1-\alpha)\sum_{n=1}^{\infty} \alpha^{n-1} z^n + (1-\rho)\sum_{n=0}^{\infty} \alpha^n z^n$$

$$= \bar{r}(1-\alpha) z^0 + \sum_{n=1}^{\infty} (1-\alpha)(r+\bar{r}\alpha)\,\alpha^{n-1} z^n$$

$$= (1-\rho) z^0 + \sum_{n=1}^{\infty} \rho(1-\alpha)\,\alpha^{n-1} z^n$$

It is now clear that the probability that the fileserver is empty is the coefficient of z^0. That is, $p_0 = (1 - \rho)$, as we obtained before. The probability that the fileserver has $n \geq 1$ requests is the coefficient of z^n. That is,

This result is identical to that obtained in [10] using a different approach.

$$p_n = \rho(1-\alpha)\,\alpha^{n-1}, \qquad n = 1, 2, \ldots$$

4. Unfinished Work

In this section we will determine the fileserver unfinished work U, defined as the number of slots needed to empty the fileserver from the requests present at the end of an arbitrary slot, henceforth called the tagged slot, assuming that no further requests will arrive after that slot. Again, it is clear that U is a RV, for which we will obtain the PGF. Our analysis will be carried out directly in steady state, unlike the case in Section 3, where we started the analysis with the transient states. Here the transient states will have no role to play in the analysis.

Let u_n and U (z) be the distribution and PGF of U. That is, u_n = Pr [U = n], n = 0, 1, ..., and $U(z) \triangleq \sum_{n=0}^{\infty} u_n z^n = E[z^U]$. Since P is the steady fileserver occupancy, it represents the number of requests in the fileserver at the end of the arbitrary slot. Clearly, the unfinished work is but the total service time of these P requests. Let us assume that the P requests presents in the fileserver at the end of the tagged slot have been arranged for service in a certain order, and let $X_{(i)}$ be the service time of the ith request in that order. Clearly, then,

It goes without saying that the $X_{(i)}$ are independent and identically distributed (iid) RVs. Let x_n and $X(z)$

$$U = \sum_{i=1}^{P} X_{(i)} \tag{11}$$

be the common distribution and common PGF of the $X_{(i)}$. That is, $x_n = \Pr[X_{(i)} = n]$ and $X(z) = E[z^{X_{(i)}}]$. Then, from (11) it follows that

$$U(z) = E\left[z^{\Sigma_{i=1}^{P} X_{(i)}}\right]$$

$$= \sum_{n=0}^{\infty} E\left[z^{\Sigma_{i=1}^{P} X_{(i)}} \mid P = n\right] \Pr\left[P = n\right] = \sum_{n=0}^{\infty} E\left[z^{\Sigma_{i=1}^{n} X_{(i)}}\right] p_n$$

Due to the independence of the $X_{(i)}$, we have

$$U(z) = \sum_{n=0}^{\infty} (E[z^{X_{(i)}}])^n p_n$$

$$= \sum_{n=0}^{\infty} (X(z))^n p_n$$

$$= P(X(z)) \tag{12}$$

From the assumptions, x_n is given by

$$x_n = s\bar{s}^{n-1},\ n > 0 \tag{13}$$

And thus $X(z)$ is given by

$$X(z) \triangleq \sum_{n=1}^{\infty} x_n z^n = \sum_{n=1}^{\infty} s\bar{s}^{n-1},\ z^n$$

$$= \frac{s}{\bar{s}}\left(\frac{1}{1-\bar{s}z} - 1\right) = \frac{sz}{1-\bar{s}z} \tag{14}$$

Substituting for $X(z)$ from (14) in (12), it follows that

$$U(z) = P\left(\frac{sz}{1-\bar{s}z}\right) \tag{15}$$

Substituting for $P(z)$ from (9) in (15), we have

$$U(z) = \frac{(s-r)\left(r\frac{sz}{1-\bar{s}z} + \bar{r}\right)}{s\bar{r} - \frac{sz}{1-\bar{s}z} - r\bar{s}}$$

$$= \frac{(s-r)(rsz + \bar{r} - \bar{r}\bar{s}z)}{s\bar{r} - s\bar{r}\bar{s}z - szr\bar{s}}$$

$$= \frac{(s-r)(\bar{r} - z + sz + rz)}{s\bar{r} - s\bar{s}z} \tag{16}$$

Note that if we put $s = 1$ in (160) we get

$$U(z) = \bar{r} + rz,$$

which means that the unfinished work possesses the same distribution as the number of arrivals per slot, and that is intuitively appealing.

5. Waiting Time

As mentioned above, the waiting time W is defined as the interval spent in the fileserver by an arbitrary request, henceforth called the tagged request, arriving at the fileserver in an arbitrary slot, henceforth called the tagged slot. In other words, assuming that the slots are numbered conecutively, and that the tagged request arrives in slot m and departs in slot n, then $W = n - m = 1, 2, \ldots$. As was the case in the unfinished work analysis, W will be analyzed below directly in steady state, since the transient states have no role to play here.

Let w_n, $n = 1, 2, \ldots$ be the distribution of W. That is, $w_n = \Pr[W = n]$. And, let $W(z)$ be the PGF of W. That is, $W(z) \triangleq \sum_{n=0}^{\infty} \Pr[W = n]\, z^n = E[z^W]$. It can be seen that W can be written as

$$W = U - \Theta + X \tag{17}$$

where U is the unfinished work at the end of the slot before the tagged slot, Θ the number of slots elapsing at the end of the tagged slot from U, and X the service time of the tagged request. Clearly, $\Theta = 0, 1$, is an indicator RV dependent on U with the following conditional distribution

$$\theta_{i|U=j} = \Pr[\Theta = i \mid U = j] = \begin{cases} 1 & \text{if } i = 1, j > 0 \\ & \text{or } \ i = 0, j = 0 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

As can be seen, the dependency between Θ and U shows in four cases: first, $U = 0, \Theta = 0$, second, $U = 0, \Theta = 1$, third, $U > 0, \Theta = 0$, and finally, $U > 0, \Theta = 1$. From (17), we get

$$W(z) = E[z^{U - \Theta + X}]$$

Since U and Θ are independent of X, it follows that

$$\begin{aligned} W(z) &= E[z^X]\, E[z^{U-\Theta}] \\ &= E(z)\, E[z^{U-\Theta}] \end{aligned} \tag{19}$$

The factor $E[z^{U-\Theta}]$ is evaluated as follows

$$\begin{aligned} E[z^{U-\Theta}] &= E[z^{U-\Theta} \mid U = 0, \Theta = 0] \Pr[U = 0, \Theta = 0] \\ &\quad + E[z^{U-\Theta} \mid U = 0, \Theta = 1] \Pr[U = 0, \Theta = 1] \\ &\quad + E[z^{U-\Theta} \mid U > 0, \Theta = 0] \Pr[U > 0, \Theta = 0] \\ &\quad + E[z^{U-\Theta} \mid U > 0, \Theta = 1] \Pr[U > 0, \Theta = 1] \\ &= + E[z^0 0] \Pr[\Theta = 0 \mid U = 0] \Pr[U = 0] \\ &\quad + E[z^{-1}] \Pr[\Theta = 1 \mid U = 0] \Pr[U = 0] \\ &\quad + E[z^U] \Pr[\Theta = 0 \mid U > 0] \Pr[U > 0] \\ &\quad + E[z^{U-1}] \Pr[\Theta = 1 \mid U > 0] \Pr[U > 0] \end{aligned}$$

Substituting for the conditional probabilities from (18), we get

$$\begin{aligned} E[z^{U-\Theta}] &= 1.1.u_0 + 0 + 0 + E[z^{U-1} \mid U > 0]\,.\,1\,.\Pr[U > 0] \\ &= u_0 + \sum_{n=1}^{\infty} z^{n-1} u_n \\ &= u_0 + \frac{1}{z}\Big(\sum_{n=0}^{\infty} u_n z^n - u_0\Big) \end{aligned}$$

Substituting for $E[z^{U-\Theta}]$ from (20) in (19) we get

$$= \frac{u_0(z-1) + U(z)}{z} \tag{20}$$

$$W(z) = X(z)\left(\frac{u_0(z-1) + U(z)}{z}\right) \tag{21}$$

But the probability u_0 that the unfinished work is zero is but the probability p_0 that the fileserver is empty. Thus,

$$u_0 = 1 - \frac{r}{s}. \tag{22}$$

Substituting for $X(z)$, $U(z)$ and u_0 from (14), 16 and (22) in (21), it follows that

$$\begin{aligned} W(z) &= \frac{sz}{1-\bar{s}z} \cdot \left(\frac{(z-1)(s-r)}{sz} + \frac{(\bar{r} - z + sz + rz)(s-r)}{z(s\bar{r} - s\bar{s}z)} \right) \\ &= \frac{sz}{1-\bar{s}z} \cdot \frac{(\bar{r} - \bar{s}z)(z-1)(s-r) + (\bar{r} - z + sz + rz)(s-r)}{zs(\bar{r} - \bar{s}z)} \\ &= \frac{z(s-r)}{\bar{r} - \bar{s}z} \end{aligned} \tag{23}$$

To find the distribution w_i, of the waiting time W, the PGF $W(z)$ will be rewritten as follows.

$$\begin{aligned} W(z) &= \frac{s-r}{\bar{r}} \frac{z}{1 - \frac{\bar{s}}{\bar{r}} Z} \\ &= \frac{s-r}{r} \sum_{i=1}^{\infty} \left(\frac{\bar{s}}{\bar{r}} \right)^{i-1} z^i \end{aligned}$$

It is now clear that w_i is

$$w_i = \frac{s-r}{\bar{r}} \left(\frac{\bar{s}}{\bar{r}} \right)^{i-1}, i = 1, 2, \ldots. \tag{24}$$

6. Expectations

In this section we will obtain the expectations for the occupancy and waiting time of the fileserver. First, using the PGF $P(z)$, as given by (9), we can find the expected occupancy of the fileserver as follows.

$$\begin{aligned} E[P] = P'(1) &= \frac{d}{dz} \frac{(s-r)(rz+\bar{r})}{s\bar{r} - zr + rsz} \bigg|_{z=1} \\ = \sum_{n=0}^{\infty} (X(z))^n p_n \quad &= \frac{(s\bar{r} - zr + rsz)\, r(s-r) + (s-r)(rz+\bar{r})\, r\bar{s}}{(s\bar{r} - zr + rsz)^2} \bigg|_{z=1} \\ &= \frac{r\bar{r}}{s-r}. \end{aligned} \tag{25}$$

Second, the expected waiting time $E[W]$ can be evaluated using two methods. The first method is by using Little's Theorem [2] which relates the expected occupancy with the expected time as follows

$$E[P] = rE[W]$$

Substituting for $E[P]$ from (25) we get

$$E[W] = \frac{\bar{r}}{s-r} \tag{26}$$

The second method is by using the PGF $W(z)$, given by (23), as follows

$$E[W] = W'(1)$$
$$= \left. \frac{(s-r)\, r\, (\bar{r} - \bar{s}z) + \bar{s}z\,(s-r)}{(\bar{r} - \bar{s}z)^2} \right|_{z=1}$$
$$= \frac{(s - r + \bar{s})}{s - r}$$
$$= \frac{r}{s - r}, \qquad (27)$$

which is the same result as (26) obtained from the fileserver occupancy using Little's formula.

In Figure 2, we plot the expected fileserver occupancy $E[P]$ as a function of the arrival rate r, for two constant service rates. In Figure 3, we plot $E[W]$ as a function of the arrival rate r, for two constant rates. We present the plots in continuous form for pleasent appearace, bearing in mind that they are actually discontinuous. Notice that the highest value of r should be less than the lowest value for s for fileserver stability.

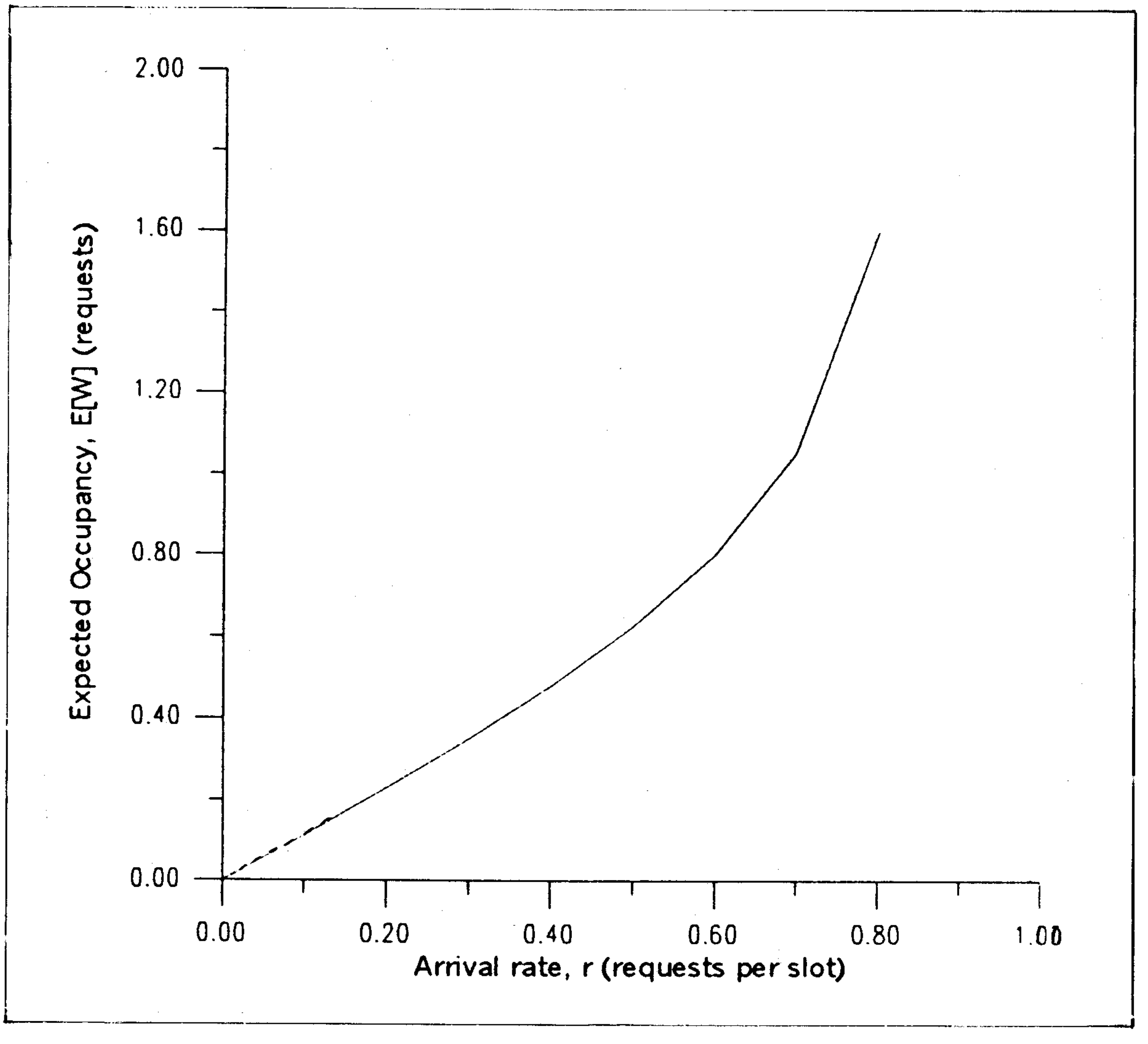

Figure 2: Expected occupancy, E[P], vs. arrival rate, r.

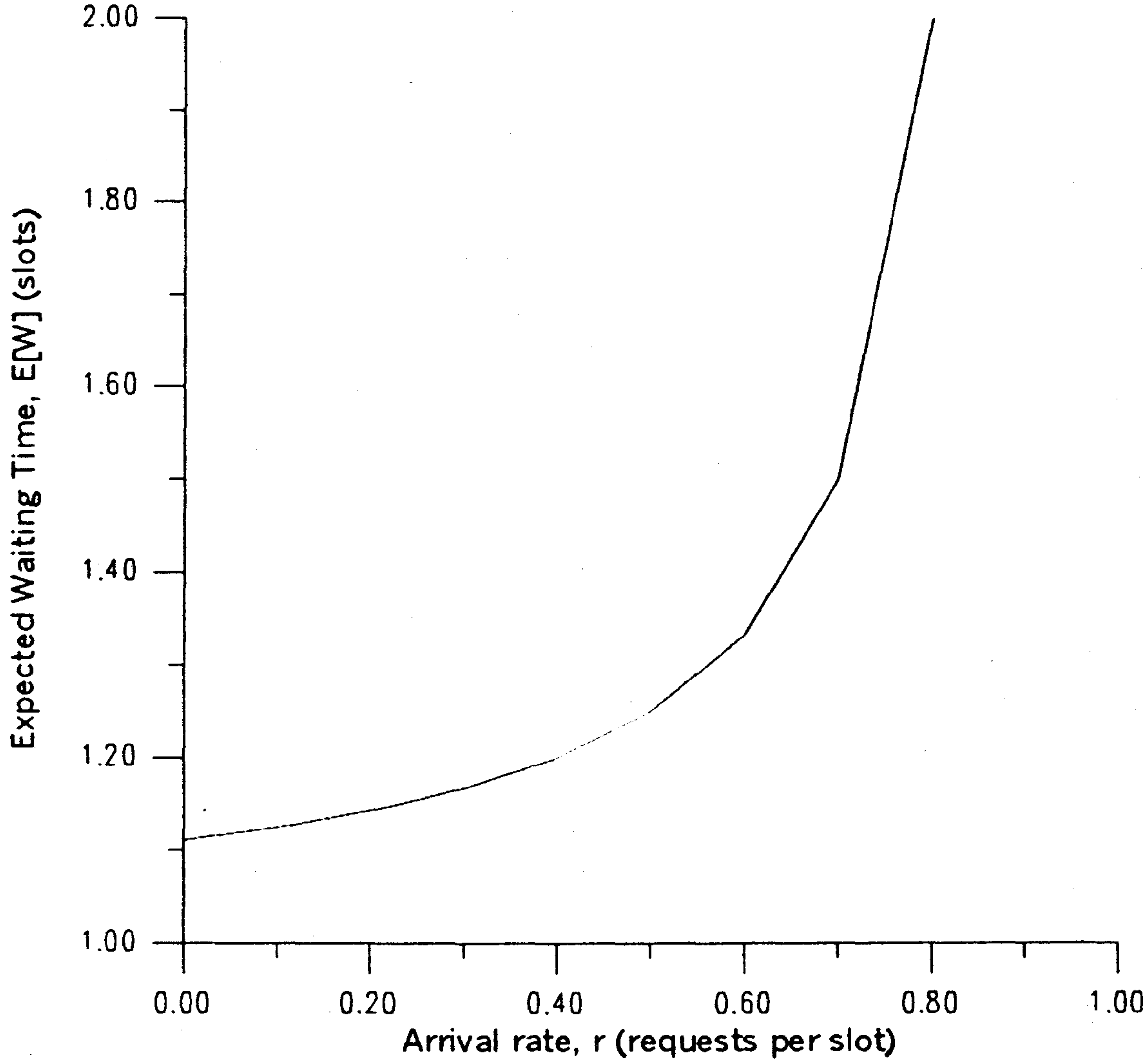

Figure 3: Expected waiting time, E[W], vs. arrival rate, r.

7. Conclusions

In this paper we studied the performance of a fileserver connected to many workstations, in an ATM network. We developed an analytical model for the fileserver, in the form of a discrete time queueing system.

We have analyzed the three main delay measures: unfinished work, busy period, and waiting time. We have found the PGFs for all these measures, and the distribution and expectations for two of them.

This work can be extended in many ways. For example, the server can be made finite. The service time can be assumed general. The server can handle the two classes with two service rates, instead of one. The arriving batches could be of mixed classes. And so on. In all these variations, this work can serve as a basis.

REFERENCES

[1] **J. Coudresuse,** "ATM: State of definition and Major Issues," *Intl. Workshop on Packed Video*, Torino, Italy, Sept. 1988.

[2] **M. E. Woodward,** *Communication and Computer Networks*, IEEE Computer Society Press. 1994.

[3] **H. Bruneel,** "Performance of Discrete-Time Queueing Systems," *Computers Ops Res.*, Vol. 2, pp. 303-320, 1993.

[4] **S. Li,** "A General Solution Technique for Discrete Queueing Analysis of Multimedia Traffic on ATM," *IEEE Trans. Communications*, pp. 1115-1132, 1990.

[5] **H. Bruneel,** "Delay Analysis for ATM Queues with ramdom order of service," *Electronics Letters*, 2, pp. 346-347, 1995.

[6] **H. Bruneel and B. Kim,** *Discrete-Time Models for Communication System Including ATM.* Kluwer Academic, Pub., 1993.

[7] **D. Kouvatsos,** "Performance modelling and evaluation of ATM netrowks," *Comput. Networks*, 34, pp. 1-2, 2000.

[8] **H. Nasser And A. Meligy,** "Fileserver Behaviour in an ATM Computer Netrowk," *Egyptian Computer Science Journal*, 18, pp. 277-294, 1996.

[9] **B. Steyaert and Y. Xiong,** "Analysis of a discrete-time queue with general three-state markovian traffic sources," *Performance Evaluation*, 27, pp. 277-294, 1996.

[10] **H. Nassar,** "A Markov Model for Multiprocessor Systems under Asynchronous Operation," *Information Processing Letters*, 54, pp. 11-16, 1996.

[11] **T. Takine and B. Sengupta,** "An analysis of a discrete-time queue for broadband ISDN with priorities among traffic classes," *IEEE Trans. On Commun.*, 42, pp. 1837-1843, 1994.

[12] **K. Laevens and H. Bruneel,** "Discrete-time multiserver queues with priorities," Per. Eval., vol. 33, pp. 249-275, Mar. 1998. *Proceedings-INFOCOM,* 2, pp. 663-670, 1996.

[13] **H. Bruneel and H. Wuyts,** "Analysis of discrete-time multiserver queueing models with constant service times," *Operations. Research Letters*, pp. 231-236, 1994.