Proceedings of 24th Conference
on Decision and Control
Ft. Lauderdale, FL • December 1985



# UNSTABLE QUEUES WITH ARRIVAL RATE LESS THAN SERVICE RATE

V. Anantharam, S. Parikh and P. Varaiya

Department of Electrical Engineering and Computer Sciences
Electronics Research Laboratory
University of California, Berkeley, CA 94720

## Summary

We give three examples of unstable queues in which the average arrival rate is less than the average service rate. In the first example, service times are iid but the arrival rate increases with queue size; in the second, arrivals are iid but the service rate decreases with queue size; in the third, arrivals are ergodic and independent of iid service times.

## Introduction

Consider a single server, discrete time queue. The service time of each customer is exponential with mean 2, and different customers have independent service times. In period $k \in \mathbb{N}$, $a_k \in \{0,1\}$ customers arrive, so the number of customers in queue at the beginning of period $k+1$ is

$$x_{k+1} = (x_k + a_k - s_k)^+, \tag{1}$$

where the virtual service process $\{s_k\}$ is iid with $P(s_k=1)=P(s_k=0)=1/2$.

Under the restriction that the arrival "rate" depends only on the queue size,

$$P\{a_k=1 \mid a_l, s_m, l \neq k\} = P\{a_k=1 \mid x_k\} =: \lambda_{x_k}, \tag{2}$$

the chain $\{x_k\}$ is Markov. If $\lambda_n \equiv \lambda$, a constant, it is known that the queue is stable, i.e.,

$$\lim n^{-1} \sum x_k < \infty \text{ a.s.}$$

iff $\lambda < 1/2$ (arrival rate less than service rate).

For the first example take

$$\lambda_n := \begin{cases} \lambda_0 & \text{if } n=0 \\ 1/2[n/n+1]^\alpha & \text{if } n>0 \end{cases} \tag{3}$$

with $0<\lambda_0<1/4$, $\alpha>0$. Its asymptotic behavior is given by

Theorem 1.

For the queue (1)-(3) one has
(i) $\{x_k\}$ is recurrent;
(ii) if $\alpha \leq 1/2$, $\{x_k\}$ is null recurrent;
(iii) if $1/2<\alpha\leq 1$, $\{x_k\}$ is positive recurrent and a.s.

$$\lim n^{-1}\sum a_k =: a(\lambda_0) \to 0 \text{ as } \lambda_0 \to 0,$$

$$\lim n^{-1}\sum x_k = \infty$$

so the queue is unstable;
(iv) if $1<\alpha$, $\{x_k\}$ is positive recurrent and the queue is stable.

Observe that for $1/2<\alpha\leq 1$ the queue is unstable even with arbitrarily small arrival rate. Queue size-dependent rates could occur in networks of queues. A more interesting example is given by the "dual" situation in which arrivals are iid but service times increase with queue size. For this second example, in place of (2),(3) take

$$P\{a_k=1 \mid a_l, s_m, l \neq k\} = 1/2, \tag{4}$$

$$P\{s_k=1 \mid a_l, s_m, m\neq k\} =: \mu_{x_k},$$

$$\mu_n := \begin{cases} 1 & \text{if } n=0 \\ 1-1/2[n/n+1]^\alpha & \text{if } n>0 \end{cases} \tag{5}$$

We interpret (5) as follows: the server gives one unit of service per unit time but $1/2[n/n+1]^\alpha$ of this service is absorbed as "overhead" in keeping track of customers in queue.

Theorem 2.

For the queue (1),(4),(5), $\{x_k\}$ is recurrent; it is null recurrent if $\alpha\leq 1/2$; it is positive recurrent and unstable if $1/2<\alpha\leq 1$; and it is positive recurrent and stable if $1<\alpha$.

In the third example the service process is iid as in the first example. The arrival process is independent of the service process and constructed in "blocks" as follows. Fix a rational number $\bar{a}=\beta/\gamma<1/2$, and a distribution $q=\{q_k, k\in\mathbb{N}\}$ with $q_k := q_1 k^{-3}$, with $q_1^{-1} := \sum k^{-3}$, so $\sum q_k = 1$. For any $k$ let $N_k := \gamma k$. Now randomly select $k_1 \in \mathbb{N}$ according to the distribution $q$ and construct the first block of $\gamma k_1$ arrivals as 1..10..0 with $a_k=1$ for $1\leq k\leq \beta k_1$ and $a_k=0$ for $\beta k_1 < k \leq \gamma k_1$. Next, randomly select $k_2$ according to $q$ and independent of $k_1$. Construct the second block of $\gamma k_2$ arrivals as 1..10..0 with $a_k=1$ for $\gamma k_1 < k \leq \gamma k_1 + \beta k_2$ and $a_k=0$ for $\gamma k_1+\beta k_2 < k \leq \gamma(k_1+k_2)$. The third block of length $\gamma k_3$ is constructed similarly with $k_3$ distributed according to $q$ and independent of $k_1$, $k_2$, etc.

Theorem 3.

The arrival process is ergodic with $\lim n^{-1}\sum a_k = \bar{a}$ a.s., and $\lim n^{-1}\sum x_k = \infty$ a.s.

## Proofs

The proofs are based on the following lemma[1].

Lemma.

Consider a Markov chain with state space $\{0,1,..\}$ and with transition probability matrix $P$ whose non-zero elements are of the form

$$P = \begin{vmatrix} r_0 & p_0 & . & . & . \\ q_1 & r_1 & p_1 & . & . \\ . & q_2 & r_2 & p_2 & . \\ . & . & . & . & . \end{vmatrix}. \tag{6}$$

Then $\pi=(\pi_0,\pi_1,..)$ with $\pi_0=1$ is a solution of $\pi P=\pi$ iff

$$\pi_n = [p_0 .. p_{n-1}]/[q_1 .. q_n], \; n \geq 1. \tag{7}$$

Moreover

$$\sum 1/p_n \pi_n = \infty \Rightarrow \text{the chain is recurrent}$$

$$\sum \pi_n = \infty \Rightarrow \text{the chain is null recurrent}$$

$$\sum \pi_n < \infty \Rightarrow \text{the chain is positive recurrent.}$$

Proof of Theorem 1.

The chain (1)-(3) has a transition matrix of the form (6) with

$$r_0 = 1-\lambda_0, \; p_0=\lambda_0, \; r_i = 1/2, \; p_i=\lambda_i/2, \; q_i=(1-\lambda_i)/2, \; i\geq 1.$$

Substitution into (7) gives

$$\pi_n = 2\frac{\lambda_0}{1-\lambda_1} .. \frac{\lambda_{n-1}}{1-\lambda_n}, \; n \geq 1,$$

so,

$$\sum_{n=0}^{\infty} \frac{1}{p_n \pi_n} = \frac{1}{\lambda_0} + \frac{1}{2} \sum_{n=1}^{\infty} \frac{1-\lambda_1}{\lambda_0} \cdot\cdot \frac{1-\lambda_n}{\lambda_{n-1}} \cdot \frac{1}{\lambda_n} = \infty$$

since $(1-\lambda_n)/(\lambda_{n-1}) > 4$ by (3). Hence the chain is recurrent. For future reference note from (3) that

$$\frac{\lambda_{k-1}}{1-\lambda_k} = \frac{[k-1/k]^\alpha}{2-[k/k+1]^\alpha} = \frac{[(k-1)(k+1)]^\alpha}{k^\alpha[2(k+1)^\alpha - k^\alpha]} .$$

Case $\alpha \leq 1/2$.
Since

$$\frac{\lambda_{k-1}}{1-\lambda_k} = \frac{[k-1/k]^\alpha}{2-[k/k+1]^\alpha} \geq \frac{[k-1/k]^{1/2}}{2-[k/k+1]^{1/2}} ,$$

it is enough to prove that $\sum \pi_n = \infty$ for $\alpha = 1/2$. For this case

$$\frac{\pi_{n+1}}{\pi_n} = \frac{\lambda_{n+1}}{1-\lambda_n} = \frac{[n/n+1]^{1/2}}{2-[n+1/n+2]^{1/2}} .$$

Since $x \to x^{-1/2}$ is convex, $n^{-1/2} > 2(n+1)^{-1/2} - (n+2)^{-1/2}$. Using this above shows $\pi_{n+1}/\pi_n > n/n+1$ which implies $\sum \pi_n = \infty$ as required.

Case $1/2 < \alpha \leq 1$.
Since $x \to x^\alpha$ is concave,

$$\frac{\lambda_{k-1}}{1-\lambda_k} = \frac{[(k-1)(k+1)]^\alpha}{k^\alpha[2(k+1)^\alpha - k^\alpha]} \leq [\frac{(k-1)(k+1)}{k(k+2)}]^\alpha ,$$

so,

$$\pi_n \leq 4.\lambda_0.2^\alpha.[\frac{1.3}{2.4}]^\alpha..[\frac{(n-1)(n+1)}{n(n+2)}]^\alpha = \frac{4.\lambda_0.6^\alpha}{[n(n+2)]^\alpha} .$$

Hence $\sum \pi_n < \infty$ and the chain is positive recurrent. Also,

$$a(\lambda_0) = \lim n^{-1} \sum a_k = (\sum \pi_n)^{-1} \sum \lambda_n \pi_n < \sum \lambda_n \pi_n$$
$$\leq \lambda_0.\text{constant}.$$

On the other hand,

$$\frac{\lambda_{k-1}}{1-\lambda_k} = \frac{[k-1/k]^\alpha}{2-[k/k+1]^\alpha} \geq \frac{k-1/k}{2-[k/k+1]} = \frac{(k-1)(k+1)}{k(k+2)} .$$

so,

$$\pi_n \geq 8.\lambda_0.[\frac{1.3}{2.4}]..[\frac{(n-1)(n+1)}{n(n+2)}] = 8.\lambda_0.\frac{(n-1)(n+1)}{n(n+2)} .$$

Hence,

$$\lim n^{-1} \sum x_k = (\sum \pi_n)^{-1} \sum n \pi_n = \infty$$

Case $1 < \alpha$.
Choose $\epsilon > 0$ so that $2\alpha - \epsilon > 2 + \epsilon$. Then

$$\frac{\pi_{n+1}}{\pi_n} = \frac{[n/n+1]^\alpha}{2-[n+1/n+2]^\alpha} < [n/n+1]^{2\alpha-\epsilon} = [n/n+1]^{2+\epsilon}$$

for $n$ large enough. Hence $\pi_n < \text{constant}.n^{2+\epsilon}$ for large $n$, so that $\sum n \pi_n < \infty$ and the chain is stable. This completes the proof.

Proof of Theorem 2.

The chain (1),(4),(5) has a transition matrix of the form (6) with

$$p_0 = 1/2,\ p_i = (1-\mu_i)/2,\ r_i = 1/2,\ q_i = \mu_i/2.$$

Using the specification (5) the proof follows as in the preceding argument.

Proof of Theorem 3.

The arrival process is ergodic since the block lengths are iid. Consider a typical arrival block of length $\gamma k$. It has the form 1..10..0 with $\beta k$ "1's" and $(\gamma-\beta)k$ "0's". If this block starts at time $t$ say, one can check that

$$E \sum_{t}^{t+\gamma k} x_k \geq E \sum_{t}^{t+\beta k} x_k \geq \frac{1}{4} \beta^2 k^2 .$$

Hence the average queue length is

$$\lim n^{-1} \sum x_k \geq \frac{E\, 1/4\beta^2 k^2}{E \gamma k}$$
$$= \frac{\sum 1/4\beta^2 q_k}{\sum \gamma k q_k} = \frac{1/4\beta^2 q_1 \sum k^{-1}}{\gamma q_1 \sum k^{-2}} = \infty$$

On the other hand the average arrival rate $\bar{a}$. This completes the proof.

## Conclusion

The study reported here started as an attempt at determining more general conditions under which the statement "arrival rate less than service rate implies stability" would be valid. The statement is true if arrivals and service times are independent, service times are iid, and arrivals form a renewal process. The examples given here show that it will be difficult to find more general "simple" conditions. Of course, if in the first example sup $\lambda_n < 1/2$, or if in the second example sup $\mu_n > 1/2$, then the queues are stable. It is also worth noting that in each of the examples the sample path behavior of the queue when it is unstable exhibits a "bistable" behavior: the queue size stays small for a long time, it then builds up rapidly to a large value, and then slowly returns to a very small value.


## Acknowledgement

Research supported in part by Office of Naval Research Contract N00014-80-C-0507 and Joint Services Electronics Program Contract AFOSR F49620-79-C-0178.


## References


1. S. Karlin and H.M. Taylor, *A first course in stochastic processes*, 2nd edition, New York, Academic Press, p. 86.